isid/ms/2016/08
June 1, 2016
http://www.isid.ac.in/~statmath/index.php?module=Preprint

Estimating the fundamental frequency using modified Newton-Raphson algorithm

Swagata Nandi and Debasis Kundu

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

ESTIMATING THE FUNDAMENTAL FREQUENCY USING MODIFIED NEWTON-RAPHSON ALGORITHM

SWAGATA NANDI[1] AND DEBASIS KUNDU[2]

Abstract. In this paper, we propose a modified Newton-Raphson algorithm to estimate the frequency parameter in the fundamental frequency model in presence of additive stationary error. The proposed estimator is super efficient in nature in the sense that its asymptotic variance is less than that of the least squares estimator having the same rate of convergence as the least squares estimator. With a proper step factor modification, we start Newton-Raphson algorithm with an initial estimator of order $O_p(n^{-1})$ and obtain an estimator with rate $O_p(n^{-\frac{3}{2}})$, the same rate as the least squares estimator. Numerical experiments are performed for different sample sizes, different error variances and different models. Three real datasets are analyzed using fundamental frequency model where the estimators are obtained using the proposed algorithm.

1. Introduction

In this paper, we consider the problem of estimating the frequency present in the following fundamental frequency model;

$$y(t) = \sum_{j=1}^{p} [A_j \cos(j\lambda t) + B_j \sin(j\lambda t)] + e(t), \quad t = 1, \ldots, n \tag{1}$$

Here $y(t)$ is the observed signal at time point t; $A_k, B_k \in \mathbb{R}$ are unknown amplitudes and none of them are not identically equal to zero; $0 < \lambda < \frac{\pi}{p}$ is the fundamental frequency; the number of components p is assumed to be known. The sequence of error random variables $\{e(t)\}$ is from a stationary linear process and satisfies the following assumption.

Assumption 1. *The sequence $\{e(t)\}$ has the following representation:*

$$e(t) = \sum_{k=0}^{\infty} a(k)\epsilon(t-k), \qquad \sum_{k=0}^{\infty} |a(k)| < \infty, \tag{2}$$

2000 *Mathematics Subject Classification.* 62J02; 62E20; 62C05.

Key words and phrases. Fundamental frequency model; Approximate least square; Modified Newton-Raphson algorithm; Super efficient estimator;

where $\{\epsilon(t)\}$ *is a sequence of independent and identically distributed (i.i.d.) random variables with mean zero and finite variance* $\sigma^2 > 0$. *The arbitrary real-valued sequence* $\{a(k)\}$ *is absolutely summable.*

Assumption 1 is a standard assumption of a weakly stationary linear process. Any stationary auto-regressive (AR), moving average (MA) or ARMA process satisfies Assumption 1 and can be expressed as (2).

The fundamental frequency model (1) is a very useful model for periodic signals when harmonics of a fundamental frequency are present. The model has applications in a variety of fields and is a special case of the usual sinusoidal model

$$y(t) = \sum_{j=1}^{p} [A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t)] + e(t). \quad t = 1, \ldots, n. \tag{3}$$

Model (1) is a particular case of model (3) with a restriction in model parameters; the frequency of the jth component of the sinusoidal model $\lambda_j = j\lambda$. When frequencies are at $\lambda, 2\lambda, \ldots, p\lambda$ instead of arbitrary $\lambda_j \in (0, \pi)$, $j = 1, \ldots, p$, these are termed as harmonics of λ. The presence of an exact periodicity is a convenient approximation, but many real life phenomena can be described using model (1). There are many non-stationary signals like speech, human circadian system where the data indicate the presence of harmonics of a fundamental frequency. In such cases, it is conveniently better to use model (1) than (3) because model (1) has one non-linear parameter as compared to p in model (3).

In the literature, many authors considered the following model instead of model (1),

$$y(t) = \sum_{j=1}^{p} \rho_j \cos(tj\lambda - \phi_j) + e(t), \tag{4}$$

where ρ_j's are amplitudes, λ is the fundamental frequency and ϕ_j's are phases and $\rho_j > 0$, $\lambda \in (0, \frac{\pi}{p})$ and $\phi_j \in (0, \pi)$, $j = 1, \ldots, p$. The sequence $\{e(t)\}$ is the error component. Note that model (4) is same as model (1) with a different parametrization. In this case, $A_j = \rho_j \cos(\phi_j)$ and $B_j = -\rho_j \sin(\phi_j)$.

We are mainly interested to estimate the fundamental frequency present in model (1) under assumption 1. The problem was originally proposed by Hannan [6] and then Quinn and Thomson [14] considered model (4) and proposed an weighted least squares approach to estimate the unknown parameters. This is basically an approximate generalized least squares criterion. Later on Nandi and Kundu [10] and Kundu and Nandi [9] studied the asymptotic properties of the least squares estimator of the unknown parameters of model (4) under

Assumption 1. Cristensen et al. [3] proposed joint estimation of fundamental frequency and number of harmonics based on MUSIC criterion. A more general model with presence of multiple fundamental frequencies has been considered by Christensen et al. [4] and Zhou [15]. A further generalized model where fundamental frequencies appear in clusters has been proposed by Nandi and Kundu [11]. It is a well known fact that even for the usual sinusoidal model (3), the Newton-Raphson (NR) algorithm does not work well. In this paper, we have modified Newton-Raphson algorithm by reducing the step factor in NR iterations applied to an equivalent criterion function of the approximate least squares estimator. We have proved that the estimator obtained from modified NR algorithm has the best rate of convergence, the rate of the LSEs and the asymptotic variance of the modified NR (MNR) estimate is four times less that than that of the least squares estimator.

Model (1) is an important model in analyzing periodic data and can be useful in situation where periodic signals are observed with an inherent fundamental frequency. Baldwin and Thomson [1] used model (1) to describe the visual observations of S.Carinae, a variable star in the Southern Hemisphere sky. Greenhouse et al. [5] proposed the use of higher-order harmonic terms of one or more fundamentals and ARMA processes for the errors for fitting biological rhythms (human core body temperature data). We shall use model (1) to analyze two speech data and transformed international airline passenger using the proposed algorithm.

The rest of the paper is organized as follows. In section 2, the least squares and approximate least squares criteria for the fundamental frequency model are described. In section 3, we propose the MNR algorithm and state the main result of the paper. In section 4, we carry out numerical experiments based on simulation. One simulated data and three real data are analyzed for illustrative purposes in section 5, and in final section, we summarize the results and direction for future work..

2. Estimation of Unknown Parameters

In matrix notation, model (1) can be written as

$$\mathbf{Y} = \mathbf{X}(\lambda)\boldsymbol{\theta} + \mathbf{e}, \tag{5}$$

where $\mathbf{Y} = (y(1), \ldots, y(n))^T$, $\mathbf{e} = (e(1), \ldots, e(n))^T$, $\boldsymbol{\theta} = (A_1, B_1, \ldots, A_p, B_p)^T$, $\mathbf{X}(\lambda) = (\mathbf{X_1}, \ldots, \mathbf{X_p})$ and

$$\mathbf{X_j} = \begin{bmatrix} \cos(j\lambda) & \sin(j\lambda) \\ \cos(2j\lambda) & \sin(2j\lambda) \\ \vdots & \vdots \\ \cos(nj\lambda) & \sin(nj\lambda) \end{bmatrix}.$$

The matrix $\mathbf{X_j} = \mathbf{X_j}(\lambda)$, but we do not make it explicit.

The least squares criterion minimizes

$$Q(\boldsymbol{\theta}, \lambda) = (\mathbf{Y} - \mathbf{X}(\lambda)\boldsymbol{\theta})^T(\mathbf{Y} - \mathbf{X}(\lambda)\boldsymbol{\theta}). \tag{6}$$

For a given λ, $Q(\boldsymbol{\theta}, \lambda)$ is minimized at $\widehat{\boldsymbol{\theta}}(\lambda) = (\mathbf{X}(\lambda)^T\mathbf{X}(\lambda))^{-1}\mathbf{X}(\lambda)^T\mathbf{Y}$. Then,

$$\begin{aligned} Q(\widehat{\boldsymbol{\theta}}, \lambda) &= \left(\mathbf{Y} - \mathbf{X}(\lambda)\widehat{\boldsymbol{\theta}}(\lambda)\right)^T \left(\mathbf{Y} - \mathbf{X}(\lambda)\widehat{\boldsymbol{\theta}}(\lambda)\right) \\ &= \mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T \left(\mathbf{X}(\lambda)^T(\mathbf{X}(\lambda)\right)^{-1} \mathbf{X}(\lambda)^T\mathbf{Y}. \end{aligned}$$

Therefore, minimizing $Q(\widehat{\boldsymbol{\theta}}, \lambda)$ with respect to λ is equivalent to maximizing

$$\mathbf{Y}^T \left(\mathbf{X}(\lambda)^T(\mathbf{X}(\lambda)\right)^{-1} \mathbf{X}(\lambda)^T\mathbf{Y}$$

with respect to λ. This quantity is asymptotically equivalent to (see Nandi and Kundu [10])

$$Q_N(\lambda) = \sum_{j=1}^{p} \left| \frac{1}{n} \sum_{t=1}^{n} y(t)e^{itj\lambda} \right|^2. \tag{7}$$

On the other hand, $\mathbf{Y}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y}$ and $\left|\frac{1}{n}\sum_{t=1}^{n} y(t)e^{itj\lambda}\right|^2$ are asymptotically equivalent. Hence our criterion is based on the maximization of

$$g(\lambda) = \sum_{j=1}^{p} \left[\mathbf{Y}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y}\right] \tag{8}$$

with respect to λ. Write $\mathbf{Y}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y} = R_j(\lambda)$, then

$$\widehat{\lambda} = \arg\max_{\lambda} g(\lambda) = \arg\max_{\lambda} \sum_{j=1}^{p} R_j(\lambda). \tag{9}$$

Note that for large n,

$$Q(\widehat{\boldsymbol{\theta}}, \lambda) = \mathbf{Y}^T\mathbf{Y} - \frac{1}{n}\sum_{j=1}^{p} \mathbf{Y}^T\mathbf{X_j}^T(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y}$$

because for large n, $\frac{1}{n}\mathbf{X_j}^T\mathbf{X_k} = \mathbf{0}$, $j \neq k$.

Once $\widehat{\lambda}$ is estimated using (9), the linear parameters are either estimated as least squares estimators,

$$\begin{pmatrix} \widehat{A}_j \\ \widehat{B}_j \end{pmatrix} = (\mathbf{X_j}(\widehat{\lambda})^T \mathbf{X_j}(\widehat{\lambda}))^{-1} \mathbf{X_j}(\widehat{\lambda})^T \mathbf{Y}.$$

or as approximated least squares estimators, given as follows:

$$\widetilde{A}_j = \frac{2}{n} \sum_{t=1}^{n} y(t)\cos(j\lambda t), \quad \widetilde{B}_j = \frac{2}{n} \sum_{t=1}^{n} y(t)\sin(j\lambda t). \tag{10}$$

The estimator of λ defined in (9) is nothing but the approximate least squares estimator (ALSE) of λ which has been studied extensively in the literature.

The asymptotic distribution of the least squares estimators and approximate least squares estimators of the unknown parameters of model (1) under assumption 1 is obtained by Nandi and Kundu [10]. In fact, Nandi and Kundu [10] studied model (4) and observed that the asymptotic distribution of LSEs and ALSEs are same. Under assumption 1, the asymptotic distribution of $\widehat{\lambda}$, the LSE of λ is as follows:

$$n^{3/2}(\widehat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N}\left(0, \frac{24\sigma^2\delta_G}{{\beta^*}^2}\right) \tag{11}$$

where $\beta^* = \sum_{j=1}^{p} j^2(A_j^2 + B_j^2)$, $\delta_G = \sum_{j=1}^{p} j^2(A_j^2 + B_j^2)c(j)$ and $c(j) = \left|\sum_{k=0}^{\infty} a(k)e^{-ijk\lambda}\right|^2$. The notation $\xrightarrow{d}$ means convergence in distribution and $\mathcal{N}(a, b)$, the Gaussian distribution with mean a and variance b.

3. Modified Newton-Raphson Algorithm

Newton-Raphson algorithm is a well-known and widely used algorithm in non-linear optimization. We use a modified version of NR algorithm. The model, considered in this paper, is a highly non-linear model in its frequency parameter λ. For the sum of sinusoidal model, that is, when the effective frequencies are not harmonics of a fundamental frequency, it has been heavily criticized in the literature against the use of NR algorithm for computing LSEs of the unknown frequencies. In many situation the NR algorithm does not converge or converges to a local minima. Therefore, we modify in the aim of obtaining estimator of the fundamental frequency which performs better and is consistent and efficient.

We assume that the number of components, p is known in advance. Also, the amplitudes A_j and B_j, $j = 1, \ldots, p$ satisfy the following constraints.

$$A_1^2 + B_1^2 \geq A_2^2 + B_2^2 \geq \cdots \geq A_p^2 + B_p^2. \tag{12}$$

This restriction is required for the ease of implementation of the proposed algorithm. It is know that the periodogram at a frequency is proportional to the sum of squares of the corresponding amplitudes. Therefore, the constraint (12) ensures that the periodogram maximizer (corresponding to the largest as well as the first peak in periodogram plot) of the observed data is an estimate of the fundamental frequency λ. The constraint (12) can be relaxed, then the first peak in the periodogram can be taken as the initial estimator of the fundamental frequency.

We first describe the NR algorithm in case of $g(\lambda) = \displaystyle\sum_{j=1}^{p} R_j(\lambda)$ in its standard form before proceeding further. Let $\widehat{\lambda}_1$ be the initial estimate of λ and $\widehat{\lambda}_k$ be the estimate at the kth iteration. Then, the NR estimate at the $(k+1)$th iteration is obtained as

$$\widehat{\lambda}_{k+1} = \widehat{\lambda}_k - \frac{g'(\widehat{\lambda}_k)}{g''(\widehat{\lambda}_k)}, \tag{13}$$

where $g'(\widehat{\lambda}_k)$ and $g''(\widehat{\lambda}_k)$ are first and second order derivatives of $g(.)$ evaluated at $\widehat{\lambda}_k$.

The standard NR algorithm (13) is modified by reducing the correction factor as follows:

$$\widehat{\lambda}_{k+1} = \widehat{\lambda}_k - \frac{1}{4}\frac{g'(\widehat{\lambda}_k)}{g''(\widehat{\lambda}_k)}. \tag{14}$$

A smaller correction factor prevents the algorithm to diverge or converge to a local minima. At a particular iteration, if the estimator is close enough to the global minimum, then a comparatively large correction factor may shift the estimate far away from the global minimum. Motivation to take a smaller step factor comes from the following theorem.

THEOREM 3.1. *Let $\widetilde{\lambda}_0$ be a consistent estimator of λ and $\widetilde{\lambda}_0 - \lambda = O_p(n^{-1-\delta})$, $\delta \in (0, \frac{1}{2}]$. Suppose $\widetilde{\lambda}_0$ is updated as $\widetilde{\lambda} = \widetilde{\lambda}_0 - \dfrac{1}{4}\dfrac{g'(\widetilde{\lambda}_0)}{g''(\widetilde{\lambda}_0)}$, then*

(a) $\widetilde{\lambda} - \lambda = O_p(n^{-1-3\delta})$ *if* $\delta \leq \frac{1}{6}$.

(b) $n^{3/2}(\widetilde{\lambda} - \lambda) \rightarrow \mathcal{N}\left(0, \dfrac{6\sigma^2\delta_G}{\beta^{*2}}\right)$, *if* $\delta > \frac{1}{6}$,

where β^* and δ_G are same as defined in the previous section. ∎

This theorem states that if we start from a reasonably good initial estimator, then the MNR algorithm produces estimator with the same convergence rate as that of the LSE of λ. Moreover, the asymptotic variance of the proposed estimator of the fundamental frequency is four times less than the asymptotic variance of the LSE. The argument maximum of the periodogram function over Fourier frequencies $\frac{2\pi k}{n}, k = 1, \ldots, \left[\frac{n}{2}\right]$, as an estimator of the frequency has a convergence rate $O_p(n^{-1})$. We use this estimator as the starting value of the algorithm implemented with a subset of the observed data vector of size n using similar idea of Bai et al. [2], Nandi and Kundu [12], and Kundu et al.[8]. The subset is selected in such a way that the dependence structure present in the data is not destroyed, that is, a subset of predefined size of consecutive points is selected as a starting point. The following algorithm assumes that the amplitudes satisfy (12).

Algorithm:

(1) Obtain the argument maximum of the periodogram function $I(\lambda)$ over Fourier frequencies and denote it as $\widetilde{\lambda}_0$. Then $\widetilde{\lambda}_0 = O_p(n^{-1})$.

(2) At $k = 1$, take $n_1 = n^{6/7}$ and calculate $\widetilde{\lambda}_1$ as

$$\widetilde{\lambda}_1 = \widetilde{\lambda}_0 - \frac{1}{4}\frac{g'_{n_1}(\widetilde{\lambda}_0)}{g''_{n_1}(\widetilde{\lambda}_0)}. \tag{15}$$

where $g'_{n_1}(\widetilde{\lambda}_0)$ and $g''_{n_1}(\widetilde{\lambda}_0)$ are same as $g'(.)$ and $g''(.)$ evaluated at $\widetilde{\lambda}_0$ with a sub-sample of size n_1. Note that $\widetilde{\lambda}_0 - \lambda = O_p(n^{-1})$ and $n_1 = n^{6/7}$, so $n = n_1^{-7/6}$. Therefore, $\widetilde{\lambda}_0 - \lambda = O_p(n^{-1}) = O_p(n_1^{-1-\frac{1}{6}})$ and applying part (a) of theorem 3.1, we have $\widetilde{\lambda}_1 - \lambda = O_p(n_1^{-1-\frac{1}{2}}) = O_p(n^{-\frac{3}{2}\times\frac{6}{7}}) = O_p(n^{-\frac{9}{7}}) = O_p(n^{-1-\frac{2}{7}})$ with $\delta = \frac{2}{7}$.

(3) As $\widetilde{\lambda}_1 - \lambda = O_p(n^{-1-\frac{2}{7}})$, $\delta = \frac{2}{7} > \frac{1}{6}$, we can apply part (b) of theorem 3.1. Take $n_{k+1} = n$, and repeat

$$\widetilde{\lambda}_{k+1} = \widetilde{\lambda}_k - \frac{1}{4}\frac{g'_{n_k+1}(\widetilde{\lambda}_k)}{g''_{n_k+1}(\widetilde{\lambda}_k)}, \quad k = 1, 2, \ldots \tag{16}$$

until a suitable stopping criterion is satisfied. ∎

Using theorem 3.1, the algorithm implies that if at any steps, the estimator of λ is of order $O_p(n^{-1-\delta})$, the updated estimator is of order $O_p(n^{-1-3\delta})$ if $\delta \leq \frac{1}{6}$ and if $\delta > \frac{1}{6}$, the updated estimator is of same order as the LSE. In addition, the asymptotic variance is four times less than the LSE, hence we call it a super efficient estimator. In the proposed algorithm, we have started with a sub-sample of size $n^{6/7}$ of the original sample of size n. The factor $\frac{6}{7}$

is not that important and not unique. There are several other choices of n_1 to initiate the algorithm, for example, $n_1 = n^{\frac{8}{9}}$ and $n_k = n$ for $k \geq 2$.

To obtain an estimator of order $O_p(n^{-1})$ is easy, but an estimator of order $O_p(n^{-1-\delta})$, $\delta \in (0, \frac{1}{2}]$ is required to apply theorem 3.1. We have started the algorithm with a smaller number of observations to overcome this problem. Varying sample size enables us to get estimator of order $O_p(n^{-1-\delta})$, for some $\delta \in (0, \frac{1}{2}]$. With the particular choice of n_1, we can use all the available data points from second step onwards. The proposed algorithm provides super efficient estimator of the fundamental frequency from the relatively poor periodogram maximizer over the Fourier frequencies. It is worth mentioning at this point is that the initial estimator is not the ALSE and is not asymptotically equivalent to the LSE. ALSE of λ in case of fundamental frequency model maximizes the sum of p periodogram functions at the harmonics without the constraints of Fourier frequencies (see Nandi and Kundu [10]).

4. Numerical Experiments

In this section, numerical experimental results are presented based on simulation to observed the performance of the proposed estimator. We consider model (1) with $p = 4$ with two different sets of parameters as follows:

$$
\begin{aligned}
\text{Model 1 :}\quad & A_1 = 5.0, \quad A_2 = 4.0, \quad A_3 = 3.0, \quad A_4 = 2.0, \\
& B_1 = 3.0, \quad B_2 = 2.5, \quad B_3 = 2.25, \quad B_4 = 2.0, \quad \lambda = .25 \\
\text{Model 2 :}\quad & A_1 = 4.0, \quad A_2 = 3.0, \quad A_3 = 2.0, \quad A_4 = 1.0, \\
& B_1 = 2.0, \quad B_2 = 1.5, \quad B_3 = 1.25, \quad B_4 = 1.0, \quad \lambda = .314.
\end{aligned}
$$

The sequence of error random variables $\{e(t)\}$ is a moving average process of order one, $e(t) = .5\epsilon(t-1) + \epsilon(t)$; $\epsilon(t)$ is a sequence of i.i.d. Gaussian random variables with mean zero and variance σ^2. The true values of the amplitudes are chosen in such a way that the constraint (12) is satisfied. The initial estimator used here is the argument maximum of the periodogram function over Fourier frequencies, as is discussed in section 3. We consider different sample sizes, $n = 100, 200, 300, 400$ and 500 and different values of the error variance of $\{\epsilon(t)\}$, $\sigma^2 = .001, .01, .1, .25, .5, .75$ and 1.0. Note that, for the generated MA process $\{e(t)\}$, the variance is $1 + \sigma^2$. For the numerical experiments considered in this section, we assume that p is known.

In each case we generate a sample of size n using the given model parameters and find the initial estimate of the fundamental frequency λ as the argument maximum of the periodogram function $I(\lambda)$. Starting from this initial estimate and known p, we compute the final estimate $\widehat{\lambda}$ of λ using the proposed iterative MNR algorithm. The iterative process is terminated when the absolute difference between two consecutive iterates is less than 10^{-7}. The generation of the data vector and estimation of the fundamental frequency, this whole process is replicated 5000 times and we have computed average estimates (AEs) and mean squared errors (MSEs) based on these 5000 replications. The asymptotic distribution of the proposed estimator as stated in theorem 3.1(b) as well as the asymptotic distribution the LSE provided in (11) are also reported for comparison of the MSEs. The results for Model 1 are reported in Tables 1 and 2 and for Model 2, in Tables 3 and 4.

Some of the salient features of the numerical experiment reported in Table 1-4.

(i) We observe that the average estimators of the fundamental frequency are very close to the true values in all sample sizes and σ^2 considered. The estimator has a positive bias which decreases as the sample size increases and increases as the error variance increases in all the cases except the case of Model 2, when sample size $n = 100$.

(ii) The MSE increases with increase in error variance whereas it decreases with increase in sample size. If verifies the consistency property of the proposed estimator.

(iii) In all the cases considered here, the MSE is close to the asymptotic variance of the MNR estimator. In addition it is smaller than the asymptotic variance of the LSE. Therefore, in line of theorem 3.1, improvement is achievable in practice.

5. Data Analysis

In this section, we have analyzed four datasets using the proposed modified Newton-Raphson algorithm. Out of four datasets, one is a simulated data with eight components and other three are real life data, namely, two sound data “uuu” and “ahh” and airline passenger data.

5.1. **Simulated data.** The data have been generated using the following parameter values

$$A_1 = 5.0, \quad A_2 = 4.5, \quad A_3 = 4.0, \quad A_4 = 3.5, \quad A_5 = 3.0, \quad A_6 = 2.5, \quad A_7 = 2.0, \quad A_8 = 1.5,$$

$$B_1 = 3.0, \quad B_2 = 2.75, \quad B_3 = 2.5, \quad B_4 = 2.0, \quad B_5 = 1.75, \quad B_6 = 1.25, \quad B_7 = 1.0, \quad B_8 = 0.5,$$

TABLE 1. The average estimates, mean squared errors, asymptotic variances of LSEs and MNR estimates of the fundamental frequency in case Model 1 when sample size $n = 100$, 200 and 300.

Sample Size N=100

σ^2	Average	Variance	Asym. Var. (LSE)	Asym. Var. (MNR)
.001	.252129	1.0844e-10	1.2539-10	3.1347e-11
.01	.252129	1.0843e-9	1.2539e-9	3.1347e-10
.1	.252129	1.0848e-8	1.2539e-8	3.1347e-9
.25	.252129	2.7133e-8	3.1347e-8	7.8367e-9
.5	.252129	5.4318e-8	6.2693e-8	1.5673e-8
.75	.252129	8.1566e-8	9.4040e-8	2.3510e-8
1.0	.252129	1.0889e-7	1.2539e-7	3.1347e-8

Sample Size N=200

σ^2	Average	Variance	Asym. Var. (LSE)	Asym. Var. (MNR)
.001	.250441	1.5803e-11	1.5673e-11	3.9183e-12
.01	.250441	1.5802e-10	1.5673e-10	3.9183e-11
.1	.250441	1.5808e-9	1.5673e-9	3.9183e-10
.25	.250441	3.9537e-9	3.9183e-9	9.7959e-10
.5	.250441	7.9124e-9	7.8367e-9	1.9592e-9
.75	.250442	1.1876e-8	1.1755e-8	2.9388e-9
1.0	.250442	1.5846e-8	1.5673e-8	3.9183e-9

Sample Size N=300

σ^2	Average	Variance	Asym. Var. (LSE)	Asym. Var. (MNR)
.001	.250150	4.4782e-12	4.6440e-12	1.1610e-12
.01	.250150	4.4869e-11	4.6440e-11	1.1610e-11
.1	.250150	4.4895e-10	4.6440e-10	1.1610e-10
.25	.250150	1.1225e-9	1.1610e-9	2.9025e-10
.5	.250150	2.2464e-9	2.3220e-9	5.8050e-10
.75	.250150	3.3720e-9	3.4830e-9	8.7074e-10
1.0	.250150	4.4994e-9	4.6440e-9	1.1610e-9

and $\lambda = .25$. The true values of the amplitude parameters are chosen such that the amplitudes satisfy the constraints

$$A_1^2 + B_1^2 > A_2^2 + B_2^2 > \cdots > A_8^2 + B_8^2$$

TABLE 2. The average estimates, mean squared errors, asymptotic variances of LSEs and MNR estimates of the fundamental frequency in case Model 1 when sample size $n = 400$ and 500

Sample Size N=400

σ^2	Average	Variance	Asym. Var. (LSE)	Asym. Var. (MNR)
.001	.250072	1.8527e-12	1.9592e-12	4.8979e-13
.01	.250072	1.8528e-11	1.9592e-11	4.8979e-12
.1	.250072	1.8533e-10	1.9592e-10	4.8979e-11
.25	.250072	4.6352e-10	4.8979e-10	1.2245e-10
.5	.250072	9.2769e-10	9.7959e-10	2.4490e-10
.75	.250072	1.3924e-9	1.4694e-9	3.6734e-10
1.0	.250073	1.8575e-9	1.9592e-9	4.8979e-10

Sample Size N=500

σ^2	Average	Variance	Asym. Var. (LSE)	Asym. Var. (MNR)
.001	.250047	9.7541e-13	1.0031e-12	2.5077e-13
.01	.250047	9.7532e-12	1.0031e-11	2.5077e-12
.1	.250047	9.7511e-11	1.0031e-10	2.5077e-11
.25	.250048	2.4376e-10	2.5077e-10	6.2693e-11
.5	.250048	4.8755e-10	5.0155-10	1.2539e-10
.75	.250048	7.3143e-10	7.5232e-10	1.8808e-10
1.0	.250048	9.7541e-10	1.0031e-9	2.5077e-10

Note that this is not a crucial restriction. This is just for convenience of implementation. The sequence of error random variables $\{e(t)\}$ is same as considered in case of numerical experiments reported in section 4 with $\sigma^2 = 1.0$. The initial estimator of the fundamental frequency λ has been taken as the periodogram maximizer over Fourier frequencies, $\frac{2\pi k}{n}$, $k = 0, \ldots, \left[\frac{n}{2}\right]$ which is equal to 0.251327. The generated data and the corresponding periodogram function are plotted in Fig. 1. Periodogram plot reveals that there are eight harmonics of the fundamental frequency λ. Therefore, we have implemented the MNR algorithm with eight components and $\widehat{\lambda}$ has come out as 0.250232. The linear parameters are estimated using the approximate least squares method . Then the predicted values are obtained using these parameter estimates and are plotted in Fig. 2 along with the generated

Table 3. The average estimates, mean squared errors, asymptotic variances of LSEs and MNR estimates of the fundamental frequency in case Model 2 when sample size $n = 100$, 200 and 300.

Sample Size N=100

σ^2	Average	Variance	Asym. Var. (LSE)	Asym. Var. (MNR)
.001	.316324	2.2248e-10	3.0918e-10	7.7296e-11
.01	.316324	2.2257e-9	3.0918e-9	7.7296e-10
.1	.316324	2.2303e-8	3.0918e-8	7.7296e-9
,25	.316323	5.5922e-8	7.7296e-8	1.9324e-8
.5	.316322	1.1238e-7	1.5459e-7	3.8648e-8
.75	.316321	1.6935e-7	2.3189e-7	5.7972e-8
1.0	.316320	2.2686e-7	3.0918e-7	7.7296e-8

Sample Size N=200

σ^2	Average	Variance	Asym. Var. (LSE)	Asym. Var. (MNR)
.001	.314611	3.5864e-11	3.8648e-11	9.6620e-12
.01	.314611	3.5952e-10	3.8648e-10	9.6620e-11
.1	.314611	3.5970e-9	3.8648e-9	9.6620e-10
.25	.314611	8.9970e-9	9.6620e-9	2.4155e-9
.5	.314611	1.8027e-8	1.9324e-8	4.8310e-9
.75	.314611	2.7098e-8	2.8986e-8	7.2465e-9
1.0	.314611	3.6208e-8	3.8648e-8	9.6620e-9

Sample Size N=300

σ^2	Average	Variance	Asym. Var. (LSE)	Asym. Var. (MNR)
.001	.314272	1.0821e-11	1.1451e-11	2.8628e-12
.01	.314272	1.0646e-10	1.1451e-10	2.8628e-11
.1	.314272	1.0750e-9	1.1451e-9	2.8628e-10
.25	.314272	2.6959e-9	2.8628e-9	7.1570e-10
.5	.314272	5.4035e-9	5.7256e-9	1.4314e-9
.75	.314272	8.1172e-9	8.5884e-9	2.1471e-9
1.0	.314272	1.0836e-8	1.1451e-8	2.8628e-9

dataset. They match reasonably well. For this generated dataset the sequence $\{e(t)\}$ is from an MA process. We have estimated the error as $e(t) = 0.0917 + \epsilon(t) + 0.4799\epsilon(t-1)$.

TABLE 4. The average estimates, mean squared errors, asymptotic variances of LSEs and MNR estimates of the fundamental frequency in case Model 2 when sample size $n = 400$ and 500

Sample Size N=400

σ^2	Average	Variance	Asym. Var. (LSE)	Asym. Var. (MNR)
.001	.314152	4.6585e-12	4.8310e-12	1.2077e-12
.01	.314152	4.6730e-11	4.8310e-11	1.2077e-11
.1	.314152	4.6732e-10	4.8310e-10	1.2077e-10
.25	.314152	1.1671e-9	1.2077e-9	3.0194e-10
.5	.314152	2.3358e-9	2.4155e-9	6.0387e-10
.75	.314152	3.5082e-9	3.6232e-9	9.0581e-10
1.0	.314152	4.6834e-9	4.8310e-9	1.2077e-9

Sample Size N=500

σ^2	Average	Variance	Asym. Var. (LSE)	Asym. Var. (MNR)
.001	.314096	2.4124e-12	2.4735e-12	6.1837e-13
.01	.314096	2.4166e-11	2.4735e-11	6.1837e-12
.1	.314096	2.4243e-10	2.4735e-10	6.1837e-11
.25	.314096	6.0615e-10	6.1837e-10	1.5459e-10
.5	.314096	1.2124e-9	1.2367e-9	3.0918e-10
.75	.314096	1.8188e-9	1.8551e-9	4.6377e-10
1.0	.314096	2.4262e-9	2.4735e-9	6.1837e-10

5.2. **"uuu" data.** This dataset is for vowel sound "uuu". It contains 512 data points sampled at 10 kHz frequency. The data were collected from the Signal Processing lab, IIT Kanpur. The mean corrected data and the periodogram function are presented in Fig. 3. It seems from the periodogram plot that there are four harmonics of the fundamental frequency. Also, note that the fundamental frequency model used to analyze this dataset, does not satisfy the condition (12) because the magnitude of the j-th peak in the periodogram plot is proportional to $A_j^2 + B_j^2$. Therefore, we take the first significance frequency as the fundamental frequency and rest are the harmonics. Initially, we have estimated the fundamental frequency with four components using the proposed MNR algorithm. Then further analysis of the residuals from the four components model indicates that there are two more significant harmonics present in the residuals. Basically, the periodogram of the residuals reveals the

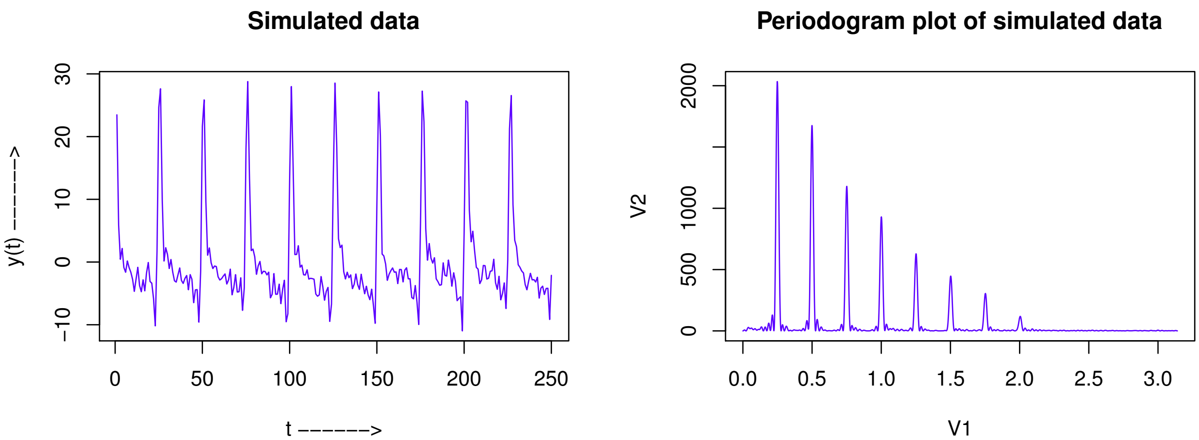

FIGURE 1. Simulated data its periodogram function.

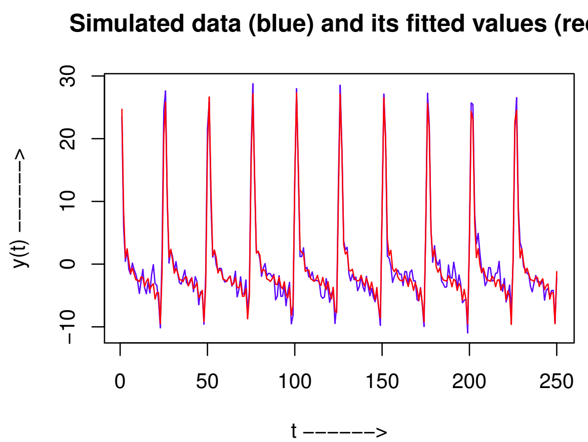

FIGURE 2. Simulated data: the fitted values (red) along with the generated data (blue).

presence of frequencies at 5λ and 6λ. Hence, we consider model (1) with 6 components. The initial estimate of λ is 0.116582 and the MNR estimate is 0.114215. The linear parameters are estimated as above. The fitted values (red)and mean corrected observed "uuu" data (blue) are plotted in Fig. 4. They match very well. Using the parameter estimates, the error sequence is estimated which can be fitted as the following stationary ARMA(2,4) process;

$$\begin{aligned} e(t) &= -4.636 + 1.8793e(t-1) - 0.9308e(t-2) + \epsilon(t) - 0.8657\epsilon(t-1) \\ &\quad -0.4945\epsilon(t-2) + 0.2859\epsilon(t-3) + 0.1415\epsilon(t-4) \end{aligned}$$

This fitting has been done based on minimum AIC.

5.3. **"ahh" data.** The third dataset is again a sound data "ahh". It contains 340 signal values sampled at 10 kHz frequency, The mean corrected data and its periodogram function

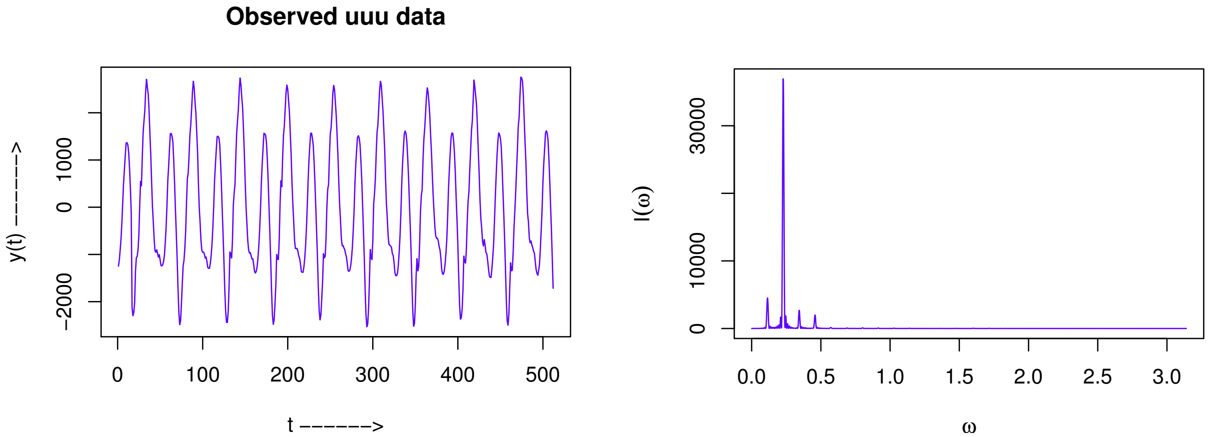

FIGURE 3. Mean corrected "uuu" data and its periodogram function.

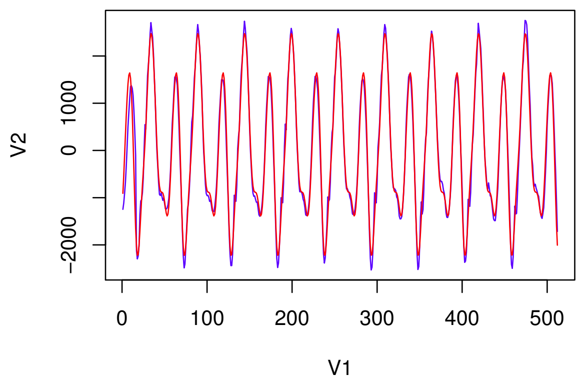

FIGURE 4. The fitted values (red) along with the mean corrected "uuu" data (blue).

are plotted in Fig. 5. Following the same methodology as applied in case of "uuu" data, we assume a fundamental frequency model with six components for "ahh" data. Using the initial estimator of the fundamental frequency λ as .092399, the MNR algorithm has been applied and the final estimate of λ is obtained as .092939. Then, the fitted values are obtained similarly as simulated data and "uuu" data. The fitted values match quite well with the mean corrected "ahh" data. The estimated error in this case is

$$e(t) = 1.8128 + 0.6816e(t-1) + \epsilon(t) + 0.4246\epsilon(t-1) - 0.5315\epsilon(t-2) - 0.6572\epsilon(t-3).$$

This is a stationary ARMA(1,3) process and can be expressed as (2).

5.4. **International airline passenger data.** This dataset is a classical dataset in times series analysis. The data represent the monthly international airline passengers during January 1953 to December 1960 and are collected from the Time Series Data Library of Hyndman [7]. This data has been analyzed by Nandi and Kundu [13] using a linear plus sinusoidal model.

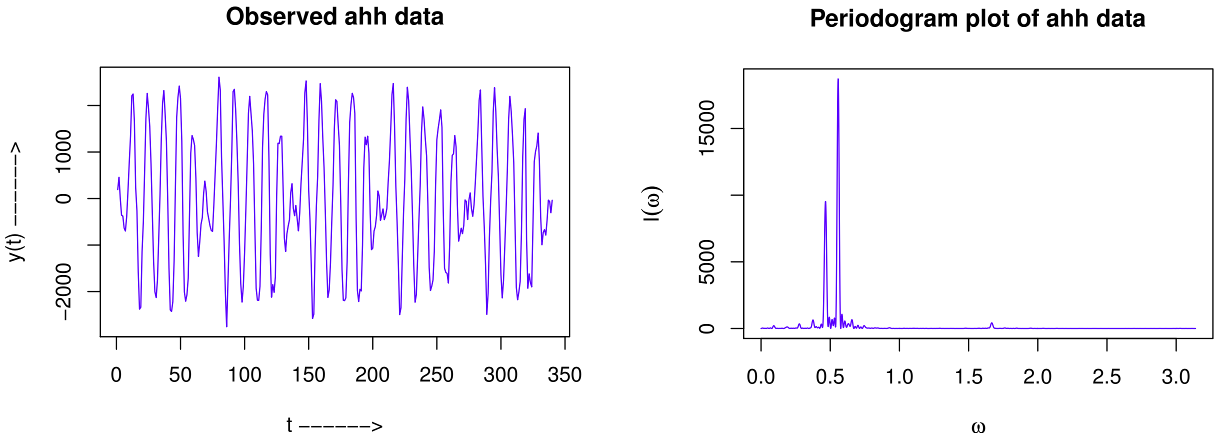

FIGURE 5. Mean corrected "ahh" data and its periodogram function.

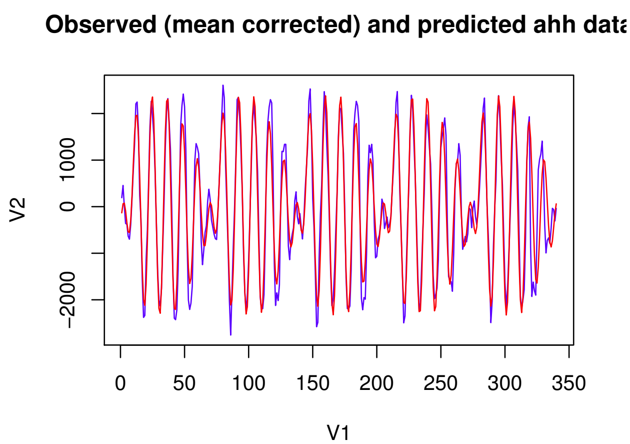

FIGURE 6. The fitted values (red) along with the mean corrected "ahh" data (blue).

The raw data is plotted in Fig. 7 and we observe that the variance is not constant. So, we apply log transform as a variance stabilizing transformation. The log transform data are plotted in Fig. 8. Fig 8 indicates that the variance can be assumed to be constant and there is a linear trend present with superimposed periodic components. Our aim is see whether the periodic part of the log transform data can be analyzed using fundamental frequency model. Hence we take the mean corrected first difference series of the log transform data which is plotted in Fig 9 and the corresponding periodogram in Fig. 10. It appears from the periodogram plot that the model (1) with five harmonics might be a suitable model for the mean corrected first difference log transform data. Now, the periodogram maximizer is 0.529110 and using this as the initial estimate in MNR algorithm, we obtain $\widehat{\lambda} = 0.523273$. The linear parameters are estimated as in section 5.1-5.3. Finally, the fitted values (red) as

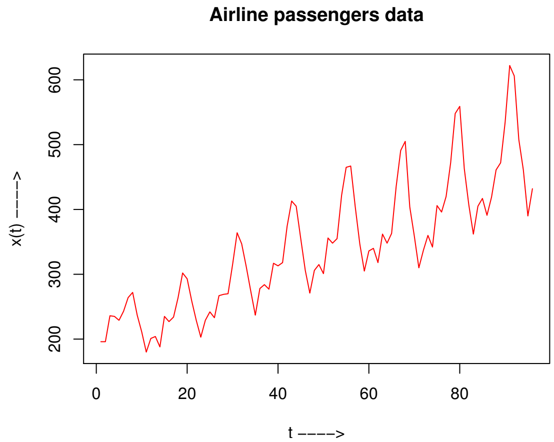

FIGURE 7. The observed airline passengers data.

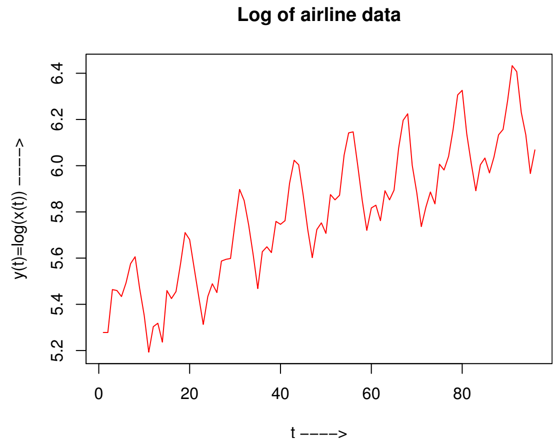

FIGURE 8. The logarithm of the observed data.

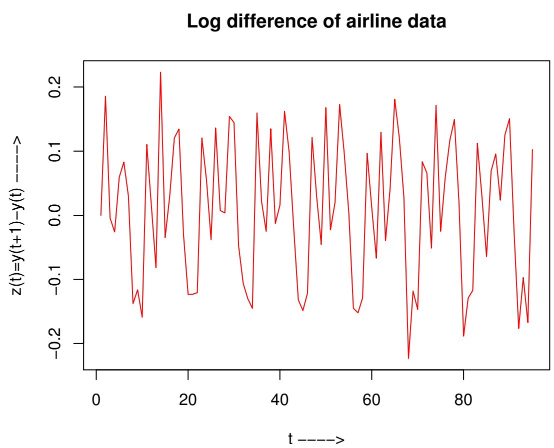

FIGURE 9. The first difference values of the logarithmic data.

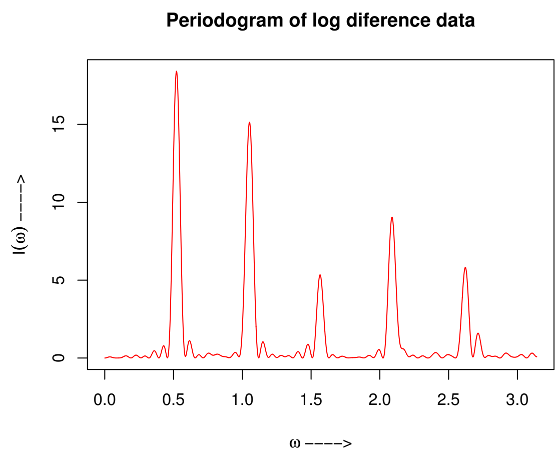

FIGURE 10. The periodogram function of the log-difference data.

well as the mean corrected first difference log transform data (blue) are plotted in Fig 11. The estimated error in this case is an i.i.d. sequence $e(t) = \epsilon(t)$.

6. CONCLUDING REMARKS

In this paper, we have considered the fundamental frequency model. This model is the multiple sinusoidal frequency model, where frequencies are harmonics of a fundamental frequency. We are mainly interested in estimating λ, the fundamental frequency. Once the λ is estimated efficiently, the other linear parameters are easily estimated using LS or approximate LS approach. It is well known that NR algorithm does not work well in case of

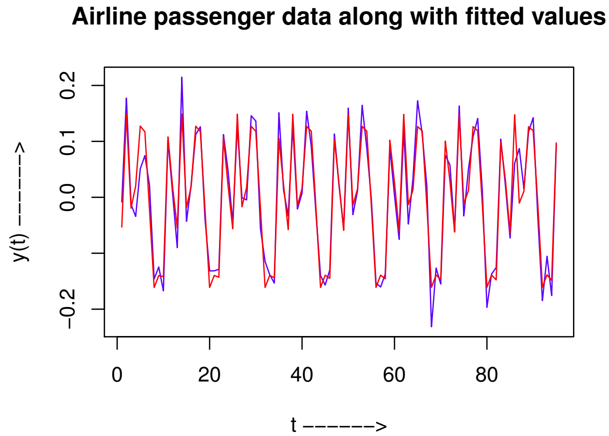

FIGURE 11. The fitted values (red) along with the log difference of airline passenger data (blue).

sinusoidal model. In this paper, we propose to modify the step factor in NR algorithm and observe that it improves the performance of the algorithm quite effectively. The asymptotic variance of the proposed estimator is smaller than that of the LSE. The fundamental frequency as a single nonlinear parameter has a quite complicated form in LS or approximate LS approach. The modified NR algorithm does not require any optimization. The calculation of first and second order derivatives at each step is only required, hence it is very simple to implement.

We think sequential application of the proposed algorithm will be required if higher order harmonic terms are present for more than one fundamental frequency. The proposed algorithm can be extended in case of multiple fundamental frequency model (Chirstensen et al. [4]) and cluster type model (Nandi and Kundu [11]). This a topic of ongoing research and would be reported elsewhere.

Appendix

Proof of Theorem 3.1: In the proof of theorem 3.1, at any iteration we use $\widetilde{\lambda}$ as the initial estimator and $\widehat{\lambda}$ as the updated estimator of λ. Now, define the following matrices to express the first and second order derivatives of $R_j(\lambda)$.

$$\mathbf{D_j} = \text{diag}\{j, 2j, \ldots, nj\}, \quad \mathbf{E} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \dot{\mathbf{X}}_\mathbf{j} = \frac{d}{d\lambda}\mathbf{X} = \mathbf{D_j X_j E}, \quad \ddot{\mathbf{X}}_\mathbf{j} = \frac{d^2}{d\lambda^2}\mathbf{X} = -\mathbf{D_j}^2\mathbf{X_j}.$$

Note that $\mathbf{EE} = -\mathbf{I}$, $\mathbf{EE}^T = \mathbf{I} = \mathbf{E}^T\mathbf{E}$ and

$$\frac{d}{d\lambda}(\mathbf{X_j}^T\mathbf{X_j})^{-1} = -(\mathbf{X_j}^T\mathbf{X_j})^{-1}[\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j} + \mathbf{X_j}^T\dot{\mathbf{X}}_\mathbf{j}](\mathbf{X_j}^T\mathbf{X_j})^{-1}.$$

Write $\frac{d}{d\lambda}R_j(\lambda) = R_j'(\lambda)$ and $\frac{d^2}{d^2\lambda}R_j(\lambda) = R_j''(\lambda)$. Then

$$\frac{1}{2}R_j'(\lambda) = \mathbf{Y}^T\dot{\mathbf{X}}_\mathbf{j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y}, \tag{17}$$

and

$$\begin{aligned}
\frac{1}{2}\,R_j''(\lambda) \;=\;& \mathbf{Y}^T\ddot{\mathbf{X}}_\mathbf{j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y} - \mathbf{Y}^T\dot{\mathbf{X}}_\mathbf{j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}(\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j} + \mathbf{X_j}^T\dot{\mathbf{X}}_\mathbf{j})(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y} \\
+\;& \mathbf{Y}^T\dot{\mathbf{X}}_\mathbf{j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{Y} - \mathbf{Y}^T\dot{\mathbf{X}}_\mathbf{j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y} \\
+\;& \mathbf{Y}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}(\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j} + \mathbf{X_j}^T\dot{\mathbf{X}}_\mathbf{j})(\mathbf{X_j}^T\mathbf{X_j})^{-1}\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y} \\
-\;& \mathbf{Y}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}(\ddot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j})(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y} \\
-\;& \mathbf{Y}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}(\dot{\mathbf{X}}_\mathbf{j}^T\dot{\mathbf{X}}_\mathbf{j})(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y} \\
+\;& \mathbf{Y}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}(\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j} + \mathbf{X_j}^T\dot{\mathbf{X}}_\mathbf{j})(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y} \\
-\;& \mathbf{Y}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{Y}.
\end{aligned} \tag{18}$$

By definition $g(\lambda) = \sum_{j=1}^{p} R_j(\lambda)$, therefore, we have $g'(\lambda) = \sum_{j=1}^{p} R_j'(\lambda)$ and $g''(\lambda) = \sum_{j=1}^{p} R_j''(\lambda)$.

Assume that $\widetilde{\lambda} - \lambda = O_p(n^{-1-\delta})$, $\delta \in (0, \frac{1}{2}]$. Therefore, for large n, at $\lambda = \widetilde{\lambda}$,

$$(\frac{1}{n}\mathbf{X_j}^T\mathbf{X_j})^{-1} = (\frac{1}{n}\mathbf{X_j}(\widetilde{\lambda})^T\mathbf{X_j}(\widetilde{\lambda}))^{-1} = 2\ I + O_p(\frac{1}{n}). \tag{19}$$

Using the large sample approximation (19) in the first term of $\frac{1}{2}R_j'(\lambda)$ in (17), we have at $\lambda = \widetilde{\lambda}$,

$$\begin{aligned}
& \frac{1}{n^3}\mathbf{Y}^T\dot{\mathbf{X}}_{\mathbf{j}}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y} \\
= & \frac{1}{n^3}\mathbf{Y}^T\dot{\mathbf{X}}_{\mathbf{j}}(\widetilde{\lambda})(\mathbf{X_j}(\widetilde{\lambda})^T\mathbf{X_j}(\widetilde{\lambda}))^{-1}\mathbf{X_j}(\widetilde{\lambda})^T\mathbf{Y} \\
= & \frac{2}{n^4}\mathbf{Y}^T\mathbf{D_j}\mathbf{X_j}(\widetilde{\lambda})\mathbf{E}\mathbf{X_j}(\widetilde{\lambda})^T\mathbf{Y} \\
= & \frac{2j}{n^4}\left[\left(\sum_{t=1}^n y(t)t\cos(\widetilde{j\lambda t})\right)\left(\sum_{t=1}^n y(t)\sin(j\widetilde{\lambda}t)\right) - \left(\sum_{t=1}^n y(t)t\sin(j\widetilde{\lambda}t)\right)\left(\sum_{t=1}^n y(t)\cos(j\widetilde{\lambda}t)\right)\right].
\end{aligned}$$

Then along the same line as Bai *et al.*[2] (see also Nandi and Kundu[12]), it can be shown that,

$$\sum_{t=1}^n y(t)\cos(j\widetilde{\lambda}t) = \frac{n}{2}\left(A_j + O_p(n^{-\delta})\right), \quad \sum_{t=1}^n y(t)\sin(j\widetilde{\lambda}t) = \frac{n}{2}\left(B_j + O_p(n^{-\delta})\right). \tag{20}$$

Now consider

$$\begin{aligned}
\sum_{t=1}^n y(t)te^{-ij\widetilde{\lambda}t} &= \sum_{t=1}^n\left(\sum_{k=1}^p\left[A_k\cos(k\lambda t) + B_k\sin(k\lambda t) + e(t)\right]\right)te^{-ij\widetilde{\lambda}t} \\
&= \frac{1}{2}\sum_{k=1}^p(A_k - iB_k)\sum_{t=1}^n t\ e^{i(k\lambda - j\widetilde{\lambda})t} + \\
&\quad \frac{1}{2}\sum_{k=1}^p(A_k + iB_k)\sum_{t=1}^n t\ e^{-i(k\lambda + j\widetilde{\lambda})t} + \sum_{t=1}^n e(t)te^{-ij\widetilde{\lambda}t}
\end{aligned} \tag{21}$$

Similarly as Bai *et al.*[2], the following can be established for harmonics of fundamental frequency;

$$\sum_{t=1}^n t\ e^{-i(k\lambda + j\widetilde{\lambda})t} = O_p(n),\ \forall\ k, j = 1, \ldots p$$
$$\sum_{t=1}^n t\ e^{-i(k\lambda - j\widetilde{\lambda})t} = O_p(n),\ \forall\ k \neq j = 1, \ldots p$$

and for $k = j$,

$$\begin{aligned}
\sum_{t=1}^n t\ e^{i(\lambda - \widetilde{\lambda})jt} &= \sum_{t=1}^n t + i(\lambda - \widetilde{\lambda})j\sum_{t=1}^n t^2 - \frac{1}{2}(\lambda - \widetilde{\lambda})^2j^2\sum_{t=1}^n t^3 \\
&\quad -\frac{1}{6}i(\lambda - \widetilde{\lambda})^3j^3\sum_{t=1}^n t^4 + \frac{1}{24}(\lambda - \widetilde{\lambda})^4j^4\sum_{t=1}^n t^5e^{i(\lambda - \lambda^*)jt}.
\end{aligned} \tag{22}$$

The last term of (22) is approximated as

$$\frac{1}{24}(\lambda-\widetilde{\lambda})^4 j^4 \sum_{t=1}^{n} t^5 e^{i(\lambda-\lambda^*)jt} = O_p(n^{2-4\delta}).$$

For the last term in (21), choose L large enough such that $L\delta > 1$ and using the Taylor series expansion of $e^{-ij\widetilde{\lambda}t}$ we obtain,

$$\begin{aligned}
&\sum_{t=1}^{n} e(t)te^{-ij\widetilde{\lambda}t} \\
= \; & \sum_{m=0}^{\infty} a(m) \sum_{t=1}^{n} e(t-m)te^{-ij\widetilde{\lambda}t} \\
= \; & \sum_{m=0}^{\infty} a(m) \sum_{t=1}^{n} e(t-m)te^{-ij\lambda t} + \sum_{m=0}^{\infty} a(m) \sum_{l=1}^{L-1} \frac{(-i(\widetilde{\lambda}-\lambda)j)^l}{l!} \sum_{t=1}^{n} e(t-m)t^{l+1}e^{-ij\lambda t} \\
& + \sum_{m=0}^{\infty} a(m) \frac{\theta(n(\widetilde{\lambda}-\lambda))^L}{L!} \sum_{t=1}^{n} t|e(t-m)| \text{ (here } |\theta| < 1) \\
= \; & \sum_{m=0}^{\infty} a(m) \sum_{t=1}^{n} e(t-m)te^{-ij\lambda t} + \sum_{l=1}^{L-1} O_p(n^{-(1+\delta)l}) O_p(n^{l+\frac{3}{2}}) + \sum_{m=0}^{\infty} a(m) O_p(n^{\frac{5}{2}-L\delta}) \\
= \; & \sum_{m=0}^{\infty} a(m) \sum_{t=1}^{n} e(t-m)te^{-ij\lambda t} + O_p(n^{\frac{5}{2}-L\delta}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sum_{t=1}^{n} y(t)t\cos(\widetilde{j\lambda}t) \\
= \; & \frac{1}{2}\left[\sum_{k=1}^{p} A_k \left(\sum_{t=1}^{n} t - \frac{1}{2}(\lambda-\widetilde{\lambda})^2 j^2 \sum_{t=1}^{n} t^3\right)\right. \\
& \left. + \sum_{k=1}^{p} B_k \left(\sum_{t=1}^{n} (\lambda-\widetilde{\lambda}) jt^2 - \frac{1}{6}(\lambda-\widetilde{\lambda})^3 j^3 \sum_{t=1}^{n} t^4\right)\right] \\
& + \sum_{m=0}^{\infty} a(m) \sum_{t=1}^{n} e(t-m)t\cos(j\lambda t) + O_p(n^{\frac{5}{2}-L\delta}) + O_p(n) + O_p(n^{2-4\delta}). \qquad (23)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\sum_{t=1}^{n} y(t) t \sin(\widetilde{j\lambda t}) \\
= \;& \frac{1}{2}\left[\sum_{k=1}^{p} B_k \left(\sum_{t=1}^{n} t - \frac{1}{2}(\lambda - \widetilde{\lambda})^2 j^2 \sum_{t=1}^{n} t^3\right)\right. \\
& \left. - \sum_{k=1}^{p} A_k \left(\sum_{t=1}^{n} (\lambda - \widetilde{\lambda}) j t^2 - \frac{1}{6}(\lambda - \widetilde{\lambda})^3 j^3 \sum_{t=1}^{n} t^4\right)\right] \\
& + \sum_{m=0}^{\infty} a(m) \sum_{t=1}^{n} e(t-m) t \sin(j\lambda t) + O_p(n^{\frac{5}{2}-L\delta}) + O_p(n) + O_p(n^{2-4\delta}). \qquad (24)
\end{aligned}$$

Next, the second term of $\frac{1}{2}R'_j(\lambda)$ in (17) is approximated as

$$\begin{aligned}
& \frac{1}{n^3}\mathbf{Y}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\dot{\mathbf{X}}_{\mathbf{j}}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y} \\
= \;& \frac{1}{n^3}\mathbf{Y}^T\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{E}^T\mathbf{X_j}^T\mathbf{D_j}\mathbf{X_j}(\mathbf{X_j}^T\mathbf{X_j})^{-1}\mathbf{X_j}^T\mathbf{Y} \\
= \;& \frac{1}{n^3}\mathbf{Y}^T\mathbf{X_j}\left(2\mathbf{I} + O_p(\frac{1}{n})\right)\mathbf{E}^T \; j \;\left(\frac{1}{4}\mathbf{I} + O_p(\frac{1}{n})\right)\left(2\mathbf{I} + O_p(\frac{1}{n})\right)\mathbf{X_j}^T\mathbf{Y} \\
= \;& \frac{j}{n^3}\mathbf{Y}^T\mathbf{X_j}\mathbf{E}^T\mathbf{X_j}^T\mathbf{Y} + O_p(\frac{1}{n}) = O_p(\frac{1}{n}), \qquad (25)
\end{aligned}$$

for large n and $\lambda = \widetilde{\lambda}$.

Now to simplify $R'_j(\widetilde{\lambda})$ and $R''_j(\widetilde{\lambda})$, we need the following results, for any $\lambda \in (0, \pi)$.

$$\sum_{t=1}^{n} t \;\cos^2(j\lambda t) = \frac{n^2}{4} + O(n), \qquad \sum_{t=1}^{n} t \;\sin^2(j\lambda t) = \frac{n^2}{4} + O(n), \qquad (26)$$

$$\sum_{t=1}^{n} \cos^2(j\lambda t) = \frac{n}{2} + o(n), \qquad \sum_{t=1}^{n} \sin^2(j\lambda t) = \frac{n}{2} + o(n), \qquad (27)$$

$$\sum_{t=1}^{n} t^2 \;\cos^2(j\lambda t) = \frac{n^3}{6} + O(n^2), \qquad \sum_{t=1}^{n} t^2 \;\sin^2(j\lambda t) = \frac{n^3}{6} + O(n^2), \qquad (28)$$

and

$$\frac{1}{n^2}\mathbf{Y}^T\mathbf{D_j}\mathbf{X_j} = \frac{j}{4}(A_j \quad B_j) + O_p(\frac{1}{n}), \qquad \frac{1}{n^3}\mathbf{Y}^T\mathbf{D_j}^2\mathbf{X_j} = \frac{j^2}{6}(A_j \quad B_j) + O_p(\frac{1}{n}), \qquad (29)$$

$$\frac{1}{n^3}\mathbf{X_j}^T\mathbf{D_j}^2\mathbf{X_j} = \frac{j^2}{6}\mathbf{I} + O_p(\frac{1}{n}), \qquad \frac{1}{n}\mathbf{X_j}^T\mathbf{Y} = \frac{1}{2}(A_j \quad B_j)^T + O_p(\frac{1}{n}), \qquad (30)$$

$$\frac{1}{n^2}\mathbf{X_j}^T\mathbf{D_j}\mathbf{X_j} = \frac{j}{4}\mathbf{I} + O_p(\frac{1}{n}). \qquad (31)$$

Next to simplify $\frac{1}{2n^3}\ R_j''(\widetilde{\lambda})$, use (19) at the first step.

$$\begin{aligned}
\frac{1}{2n^3}\ R_j''(\widetilde{\lambda}) \;&=\; \frac{2}{n^4}\mathbf{Y}^T\ddot{\mathbf{X}}_\mathbf{j}\mathbf{X_j}^T\mathbf{Y} - \frac{4}{n^5}\mathbf{Y}^T\dot{\mathbf{X}}_\mathbf{j}(\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j} + \mathbf{X_j}^T\dot{\mathbf{X}}_\mathbf{j})\mathbf{X_j}^T\mathbf{Y_j} + \frac{2}{n^4}\mathbf{Y}^T\dot{\mathbf{X}}_\mathbf{j}\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{Y_j}\\
&-\; \frac{4}{n^5}\mathbf{Y}^T\dot{\mathbf{X}}_\mathbf{j}\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j}\mathbf{X_j}^T\mathbf{Y} + \frac{8}{n^6}\mathbf{Y}^T\mathbf{X_j}(\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j} + \mathbf{X_j}^T\dot{\mathbf{X}}_\mathbf{j})\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j}\mathbf{X_j}^T\mathbf{Y}\\
&-\; \frac{4}{n^5}\mathbf{Y}^T\mathbf{X_j}\ddot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j}\mathbf{X_j}^T\mathbf{Y} - \frac{4}{n^5}\mathbf{Y}^T\mathbf{X_j}\dot{\mathbf{X}}_\mathbf{j}{}^T\dot{\mathbf{X}}_\mathbf{j}\mathbf{X_j}{}^T\mathbf{Y}\\
&+\; \frac{8}{n^6}\mathbf{Y}^T\mathbf{X_j}\dot{\mathbf{X}}_\mathbf{j}{}^T\mathbf{X_j}(\dot{\mathbf{X}}_\mathbf{j}^T\mathbf{X_j} + \mathbf{X_j}^T\dot{\mathbf{X}}_\mathbf{j})\mathbf{X_j}^T\mathbf{Y} - \frac{4}{n^5}\mathbf{Y}^T\mathbf{X_j}\dot{\mathbf{X}}_\mathbf{j}{}^T\mathbf{X_j}\dot{\mathbf{X}}_\mathbf{j}{}^T\mathbf{Y} + O_p(\frac{1}{n}).
\end{aligned}$$

In the second step, use $\ddot{\mathbf{X}}_\mathbf{j} = \mathbf{D_j}\mathbf{X_j}\mathbf{E}$ and $\ddot{\mathbf{X}} = -\mathbf{D_j}^2\mathbf{X_j}$.

$$\begin{aligned}
\frac{1}{2n^3}\ R_j''(\widetilde{\lambda}) \;&=\; -\frac{2}{n^4}\mathbf{Y}^T\mathbf{D_j}^2\mathbf{X_j}\mathbf{X_j}^T\mathbf{Y} - \frac{4}{n^5}\mathbf{Y}^T\mathbf{D_j}\mathbf{X_j}\mathbf{E}(\mathbf{E}^T\mathbf{X_j}^T\mathbf{D_j}\mathbf{X_j} + \mathbf{X_j}^T\mathbf{D_j}\mathbf{X_j}\mathbf{E})\mathbf{X_j}^T\mathbf{Y}\\
&+\; \frac{2}{n^4}\mathbf{Y}^T\mathbf{D_j}\mathbf{X_j}\mathbf{E}\mathbf{E}^T\mathbf{X_j}^T\mathbf{D_j}\mathbf{Y} - \frac{4}{n^5}\mathbf{Y}^T\mathbf{D_j}\mathbf{X_j}\mathbf{E}\mathbf{E}^T\mathbf{X_j}^T\mathbf{D_j}\mathbf{X_j}\mathbf{X_j}^T\mathbf{Y}\\
&+\; \frac{8}{n^6}\mathbf{Y}^T\mathbf{X_j}(\mathbf{E}^T\mathbf{X_j}^T\mathbf{D_j}\mathbf{X_j} + \mathbf{X_j}^T\mathbf{D_j}\mathbf{X_j}\mathbf{E})\mathbf{E}^T\mathbf{X_j}^T\mathbf{D_j}\mathbf{X_j}\mathbf{X_j}^T\mathbf{Y} + \frac{4}{n^5}\mathbf{Y}^T\mathbf{X_j}\mathbf{X_j}^T\mathbf{D_j}^2\mathbf{X_j}\mathbf{X_j}^T\mathbf{Y}\\
&-\; \frac{4}{n^5}\mathbf{Y}^T\mathbf{X_j}\mathbf{E}^T\mathbf{X_j}^T\mathbf{D_j}^2\mathbf{X_j}\mathbf{E}\mathbf{X_j}^T\mathbf{Y} + \frac{8}{n^6}\mathbf{Y}^T\mathbf{X_j}\mathbf{E}^T\mathbf{X_j}^T\mathbf{D_j}\mathbf{X_j}(\mathbf{E}^T\mathbf{X_j}^T\mathbf{D_j}\mathbf{X_j}\\
&+\; \mathbf{X_j}^T\mathbf{D_j}\mathbf{X_j}\mathbf{E})\mathbf{X_j}^T\mathbf{Y} - \frac{4}{n^5}\mathbf{Y}^T\mathbf{X_j}\mathbf{E}^T\mathbf{X_j}^T\mathbf{D_j}\mathbf{X_j}\mathbf{E}^T\mathbf{X_j}^T\mathbf{D_j}\mathbf{Y} + O_p(\frac{1}{n}).
\end{aligned}$$

Next, using (26)-(31), we observe

$$\begin{aligned}
\frac{1}{2n^3}\ R_j''(\widetilde{\lambda}) \;&=\; (A_j^2 + B_j^2)\left[-\frac{j^2}{6} - 0 + \frac{j^2}{8} - \frac{j^2}{8} + 0 + \frac{j^2}{6} - \frac{j^2}{6} + 0 + \frac{j^2}{8}\right] + O_p(\frac{1}{n})\\
&=\; -\frac{j^2}{24}(A_j^2 + B_j^2) + O_p(\frac{1}{n}). \qquad (32)
\end{aligned}$$

The correction factor in Newton-Raphson algorithm can be written as

$$\frac{g'(\widetilde{\lambda})}{g''(\widetilde{\lambda})} = \frac{\dfrac{1}{2n^3}\displaystyle\sum_{j=1}^{p} R_j'(\widetilde{\lambda})}{\dfrac{1}{2n^3}\displaystyle\sum_{j=1}^{p} R_j''(\widetilde{\lambda})} \qquad (33)$$

Using (23), (24) and (25), $\frac{1}{2n^3}\sum_{j1}^{p} R'_j(\widetilde{\lambda})$ is simplified as

$$\begin{aligned}
\frac{1}{2n^3}\sum_{j=1}^{p} R'_j(\widetilde{\lambda}) &= \frac{2}{n^4}\sum_{j=1}^{p} j\left[\frac{n}{2}(B_j+O_p(n^{-\delta}))\left\{\frac{A_j}{2}\left(\sum_{t=1}^{n} t-\frac{1}{2}(\lambda-\widetilde{\lambda})^2 j^2\sum_{t=1}^{n} t^3\right)\right.\right.\\
&\quad+\frac{B_j}{2}\left(\sum_{t=1}^{n}(\lambda-\widetilde{\lambda})jt^2-\frac{1}{6}(\lambda-\widetilde{\lambda})^3 j^3\sum_{t=1}^{n} t^4\right)\\
&\quad\left.+\sum_{k=0}^{\infty} a(k)\sum_{t=1}^{n} e(t-k)t\cos(j\lambda t)+O_p(n^{\frac{5}{2}-L\delta})+O_p(n)+O_p(n^{2-4\delta})\right\}\\
&\quad-\frac{n}{2}(A_j+O_p(n^{-\delta}))\left\{\frac{B_j}{2}\left(\sum_{t=1}^{n} t-\frac{1}{2}(\lambda-\widetilde{\lambda})^2 j^2\sum_{t=1}^{n} t^3\right)\right.\\
&\quad-\frac{A_j}{2}\left(\sum_{t=1}^{n}(\lambda-\widetilde{\lambda})jt^2-\frac{1}{6}(\lambda-\widetilde{\lambda})^3 j^3\sum_{t=1}^{n} t^4\right)\\
&\quad\left.\left.+\sum_{k=-}^{\infty} a(k)\sum_{t=1}^{n} e(t-k)t\sin(j\lambda t)+O_p(n^{\frac{5}{2}-L\delta})+O_p(n)+O_p(n^{2-4\delta})\right\}\right]\\
&= \sum_{j=1}^{p} j\left[\frac{1}{2n^3}(A_j^2+B_j^2)\left\{\sum_{t=1}^{n}(\lambda-\widetilde{\lambda})jt^2-\frac{1}{6}(\lambda-\widetilde{\lambda})^3 j^3\sum_{t=1}^{n} t^4\right\}\right.\\
&\quad+\frac{1}{n^3}\left\{B_j\sum_{k=0}^{\infty} a(k)\sum_{t=1}^{n} e(t-k)t\cos(j\lambda t)\right.\\
&\quad\left.\left.+A_j\sum_{k=0}^{\infty} a(k)\sum_{t=1}^{n} e(t-k)t\sin(j\lambda t)\right\}\right]\\
&\quad+O_p(n^{-\frac{1}{2}-L\delta})+O_p(n^{-2})+O_p(n^{-1-4\delta}),
\end{aligned}$$

and using (32), the denominator of (33) is $\frac{1}{2n^3}\sum_{j=1}^{p} R''_j(\widetilde{\lambda}) = -\frac{1}{24}\sum_{j=1}^{p} j^2(A_j^2+B_j^2)+O_p(\frac{1}{n})$. Therefore,

$$\widehat{\lambda} = \widetilde{\lambda}-\frac{1}{4}\frac{g'(\widetilde{\lambda})}{g''(\widetilde{\lambda})} = \widetilde{\lambda}-\frac{1}{4}\frac{\frac{1}{2n^3}\sum_{j=1}^{p} R'_j(\widetilde{\lambda})}{\frac{1}{2n^3}\sum_{j=1}^{p} R''_j(\widetilde{\lambda})}$$

$$
\begin{aligned}
&= \widetilde{\lambda} - \frac{1}{4}\frac{\frac{1}{2n^3}\sum_{j=1}^{p} R_j'(\widetilde{\lambda})}{-\frac{1}{24}\sum_{j=1}^{p} j^2(A_j^2+B_j^2) + O_p(\frac{1}{n})} \\
&= \widetilde{\lambda} + \frac{6}{(\beta^* + O_p(\frac{1}{n}))}\sum_{j=1}^{p} j\left[\frac{1}{2n^3}(A_j^2+B_j^2)\left\{\sum_{t=1}^{n}(\lambda-\widetilde{\lambda})jt^2 - \frac{1}{6}(\lambda-\widetilde{\lambda})^3 j^3 \sum_{t=1}^{n} t^4\right\}\right] \\
&\quad + \frac{6}{(\beta^* + O_p(\frac{1}{n}))}\sum_{j=1}^{p}\frac{j}{f(j\lambda)}\frac{1}{n^3}\left\{B_j\sum_{k=0}^{\infty}a(k)\sum_{t=1}^{n}e(t-k)t\cos(j\lambda t)\right. \\
&\quad \left. + A_j\sum_{k=0}^{\infty}a(k)\sum_{t=1}^{n}e(t-k)t\sin(j\lambda t)\right\} + O_p(n^{-\frac{1}{2}-L\delta}) + O_p(n^{-2}) + O_p(n^{-1-4\delta}) \quad (34)
\end{aligned}
$$

Here $\beta^* = \sum_{j=1}^{p}\frac{j^2(A_j^2+B_j^2)}{f(j\lambda)}$ is same as defined after (11). When $\delta \le \frac{1}{6}$ in (34), $\widehat{\lambda} - \lambda = O_p(n^{-1-3\delta})$ whereas if $\delta > \frac{1}{6}$, then for large n,

$$
\begin{aligned}
n^{3/2}(\widehat{\lambda} - \lambda) &\stackrel{d}{=} \frac{6n^{-3/2}}{\beta^*}\sum_{j=1}^{p} j\left\{B_j\sum_{k=0}^{\infty}a(k)\sum_{t=1}^{n}e(t-k)t\cos(j\lambda t)\right. \\
&\qquad \left. + A_j\sum_{k=0}^{\infty}a(k)\sum_{t=1}^{n}e(t-k)t\sin(j\lambda t)\right\} \\
&\xrightarrow{d} \mathcal{N}(0,\gamma)
\end{aligned}
$$

where

$$
\begin{aligned}
\gamma &= \frac{36}{{\beta^*}^2}\frac{\sigma^2}{6}\sum_{j=1}^{p} j^2(A_j^2+B_j^2)\left[\left\{\sum_{k=0}^{\infty}a(k)\cos(kj\lambda)\right\}^2 + \left\{\sum_{k=0}^{\infty}a(k)\sin(kj\lambda)\right\}^2\right] \\
&= \frac{6}{{\beta^*}^2}\sigma^2\sum_{j=1}^{p} j^2(A_j^2+B_j^2)\left|\sum_{k=0}^{\infty}a(k)e^{-ikj\lambda}\right|^2 = \frac{6\sigma^2\delta_G}{{\beta_*}^2}.
\end{aligned}
$$

This proves the theorem.

References

[1] Baldwin, A. J., Thomson, P. J. (1978), "Periodogram analysis of S. Carinae.*Royal Astronomy Society New Zealand (Variable Star Section)*, 6,31-38.

[2] Bai, Z.D., Rao, C.R., Chow, M. and Kundu, D. (2003), "An efficient algorithm for estimating the parameters of superimposed exponential signals", *Journal of Statistical Planning and Inference*, 110, 23 - 34.

[3] Christensen, M. G., Jakobsson, A. and Jensen, S. H., (2007), "Joint high-resolution fundamental frequency and order estimation", *IEEE Transactions on Audio, Speech, and Language Processing*, 15(5), 1635-1644.

[4] Christensen, M. G., Hovang, J. L., Jakobsson, A. and Jensen, S. H., (2011), "Joint fundamental frequency and order estimation using optimal filtering", *EURASIP Journal on Advances in Signal Processing*, 2011(13).

[5] Greenhouse, J. B., Kass, R. E. and Tsay, R. S. (1987), "Fitting non linear models with ARMA errors to biological rhythm data" , **Statistics in Medicine**, 6(2),167-183.

[6] Hannan, E. J. (1974), "Time series analysis", *IEEE Trans. Auto. Control*, 19, 706-715.

[7] Hyndman, R.J. (n.d.) Time Series Data Library, http://www.robhyndman.info/TSDL. Accessed on 01.01.2008.

[8] Kundu, D., Bai, Z., Nandi, S. and Bai, L. (2011), "Super efficient frequency estimator", *Journal of Statistical Planning and Inference*, 141, 2576 - 2588.

[9] Kundu, D. and Nandi, S. (2004), "A Note on estimating the frequency of a periodic function", *Signal Processing*, 84(3), 653-661.

[10] Nandi, S. and Kundu, D. (2003), "Estimating the fundamental frequency of a periodic function", *Statistical Methods and Applications*, 12(3), 341-360 .

[11] Nandi, S. and Kundu, D. (2006a), "Analyzing Non-Stationary Signals Using Generalized Multiple Fundamental Frequency Model", *Journal of Statistical Planning and Inference*, Vol. 136, 3871-3903.

[12] Nandi, S. and Kundu, D. (2006b), "A fast and efficient algorithm for estimating the parameters of sum of sinusoidal model", *Sankhya*, 68, 283 - 306.

[13] Nandi, S. and Kundu, D. (2013), "Estimation of parameters of partially sinusoidal frequency model", *Statistics: A Journal of Theoretical and Applied Statistics*, 47(1), 45-60.

[14] Quinn, B. G. and Thomson, P. J. (1991),"Estimating the frequency of a periodic function", *Biometrika*, 78, 65-74.

[15] Zhou, Z. (2013), "Spectral analysis of sinusoidal signals from multiple channels", Ph. D. thesis submitted to Department of Electronic Engineering, City University of Hong Kong.

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

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