J. Japan Statist. Soc.
Vol. 85 No. 1 2005 1–19

ESTIMATING THE NUMBER OF COMPONENTS OF THE FUNDAMENTAL FREQUENCY MODEL

Debasis Kundu* and Swagata Nandi**

We propose a simple estimation procedure of the number of components of the fundamental frequency model when all the adjacent harmonics are present. The proposed method is based on the penalty function approach like other Information Theoretic Criteria. The new method is shown to be consistent. We compute the probability of wrong estimates of a particular penalty function and propose a re-sampling technique to estimate the probability of wrong estimates. It is observed that the probability of wrong estimates can be used to choose the *best* possible penalty function from a particular class of penalty functions. The effectiveness of the proposed method is verified using computer simulations. Two speech data are analyzed using our proposed technique and the performances are quite satisfactory. Finally, we extend our results when all the adjacent harmonics may not be present in the model.

Key words and phrases: Fundamental Frequency, Consistent Estimator, Penalty Function, Information Theoretic Criterion.

1. Introduction

In this paper we consider the estimation of the number of components of the following fundamental frequency model:

$$y(n) = \sum_{j=1}^{p^0} \rho_j^0 \cos(nj\lambda^0 - \phi_j^0) + X(n); \qquad n = 1, \ldots, N, \tag{1.1}$$

where all the $\rho_j^0 > 0$. Here, $0 < \lambda^0 < \frac{\pi}{p^0}$ is the fundamental frequency and $j\lambda^0$ are its harmonics for $j > 1$. The phase components ϕ_j^0's are unknown and $-\pi < \phi_j^0 < \pi$. $X(n)$'s are additive stationary linear processes with mean 0, finite variance σ^2 and satisfy the following assumption.

Assumption 1: $X(n)$ has the following representation:

$$X(n) = \sum_{k=-\infty}^{\infty} a(k)e(n-k), \tag{1.2}$$

where $e(k)$'s are independent and identically distributed (i.i.d.) random variables with mean zero and finite variance τ^2. Moreover,

$$\sum_{k=-\infty}^{\infty} |a(k)| < \infty. \tag{1.3}$$

There are mainly two problems, one is the estimation of p^0 and the other one is the estimation of the ρ_j^0's, ϕ_j^0's and λ^0. In this paper we mainly consider the estimation of p^0.

Received October, 2003. Revised May 15, 2004. Accepted November 2, 2004.

*Department of Mathematics, Indian Institute of Technology Kanpur, Pin 208016, India.
**Department of Economics, University of Mannheim, L 7 3-5, 68131 Mannheim, Germany.

Estimating the parameters of a fundamental frequency model is an important problem in Speech Signal Processing. An additional mean term μ^0 can always be added to the model (1.1) and an efficient estimator of μ^0 can be easily obtained as $\bar{y} = \frac{1}{N}\sum_{n=1}^{N} y(n)$. Therefore, in practice it is recommended to work with the transformed data namely, $y(n) - \bar{y}$ and apply the necessary results. Since the main concern of this paper is to estimate p^0 efficiently, we consider the model (1.1) for brevity. Note that many authors considered the estimation and testing for the related discrete time model;

$$y(n) = \mu^0 + \sum_{j=1}^{p^0} \rho_j^0 \cos(n\lambda_j^0 - \phi_j^0) + X(n), \tag{1.4}$$

where λ_j^0 denotes unknown frequencies $(0 < \lambda_1^0 \cdots < \lambda_{p^0}^0 < \pi)$ and ρ_j^0 and ϕ_j^0 are same as defined before. See for example the work of Fisher (1929), Whittle (1959), Walker (1971), Hannan (1973), Kay and Marple (1981) and Stoica (1993) for an extensive list of references of this problem.

Estimation of p^0 of the model (1.4) has been considered by several authors in the Signal Processing and Time Series literature for the past twenty years. For example, Fuchs (1988), Wang and Kaveh (1986), Kaveh, Wang and Hung (1987), Sakai (1990, 1993) and Kundu (1998) considered the problem when the errors are independent and identically distributed. Quinn (1989) considered the model (1.4) in presence of stationary errors but the frequencies can be only the Fourier frequencies, namely of the form $\frac{2\pi n}{N}$. Wang (1993) considered the general case, *i.e* when the errors are stationary and the frequencies need not be Fourier frequencies only. But the criterion used by Wang (1993) involves a constant whose determination needs the knowledge of the noise spectrum.

Here, we consider a particular case where the unknown frequencies λ_j^0 are harmonics of a fundamental frequency λ^0. Interestingly, although the model (1.4) is a very well studied model, not much work has been done on the fundamental frequency model (1.1). There are many signals like speech, where the data indicate the presence of harmonics with respect to a fundamental frequency. We provide the plot of two speech data. The Figure 1 represents 'uuu' sound and Figure 2 represents 'ahh' sound. The periodogram function of both data are plotted in Figures 3 and 4 respectively. From Figures 3 and 4, it is clear that the harmonics of a particular frequency are present in both cases. In these situations, it is better to use the model (1.1) than (1.4) as the model (1.1) has lesser number of non-linear parameters than (1.4) for fixed $p^0 > 1$. The problem was first proposed by Fisher (1929) and later on Quinn and Thomson (1991) and Nandi and Kundu (2003) proposed different estimation procedures of ρ_j^0, ϕ_j^0 and λ^0 for a fixed p^0. But no where the estimation of p^0 has been considered by any author.

In this paper, we consider the estimation of p^0 of the model (1.1). We mainly use the penalty function approach like AIC, BIC or MDL, but instead of using any fixed penalty function, a class of penalty functions satisfying some special properties has been used. It is observed that any penalty function from that particular class will provide consistent estimates of the unknown parameters. We perform some simulation experiments to observe the behavior of the proposed estimates for small samples. We compute the probability of correct estimates (PCE's) of a particular penalty function. Based on the re-sampling technique as used in Kundu and Mitra (2001), we compute an estimate of the PCE for each penalty function. Once we obtain an estimate of PCE, we use that penalty function for which the estimated PCE is maximum. It is observed as expected that a particular penalty function may not work well for all possible error variances or for all possible parameter values, but our method tries to choose the *best* possible penalty function from the given class of penalty functions. Simulation results

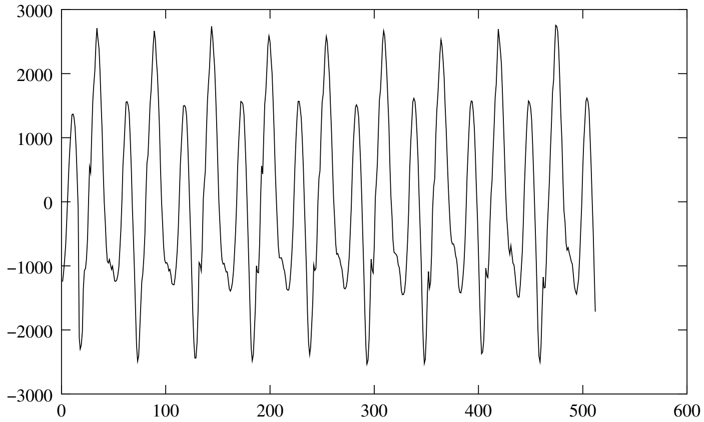

Figure 1. Plot of the mean corrected "uuu" vowel sound.

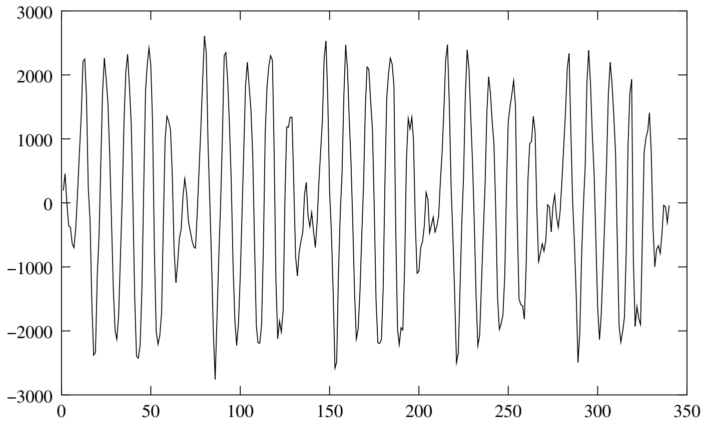

Figure 2. Plot of the mean corrected "ahh" sound.

suggest that our method performs quite well. Two speech data are analyzed using our proposed method and the performances are quite satisfactory.

The rest of the paper is organized as follows. In section 2, we propose the estimation procedure and its implementation in practice. In section 3, we provide the consistency results of the proposed estimates. The practical implementation procedure of the proposed technique is provided in section 4. Some experimental results are provided in section 5 and two speech data are analyzed using our proposed method in section 6. In section 7, we generalize our results to those models where some of the adjacent harmonics might be absent. Finally we conclude the paper in section 8.

2. Estimation procedure

It is assumed that the number of components can be at most K, a fixed number. Suppose, L denotes the possible ranges of p^0, therefore, $L \in \{0, 1, \ldots, K\}$. If $M_0, M_1, \ldots, M_K$ denote the different models of order $0, 1, \ldots, K$ respectively, then the

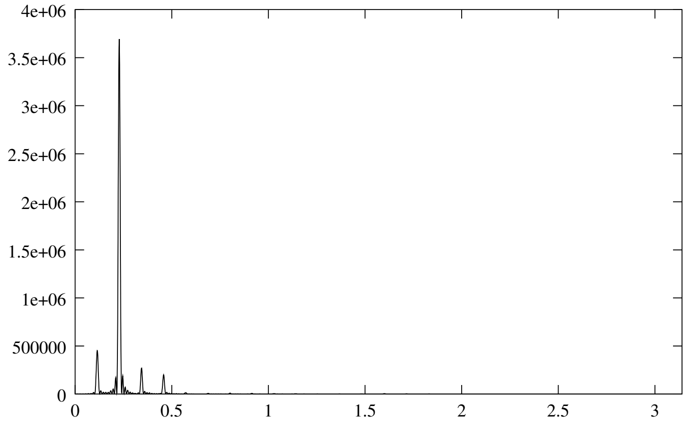

Figure 3. Plot of the periodogram function of the "uuu" sound.

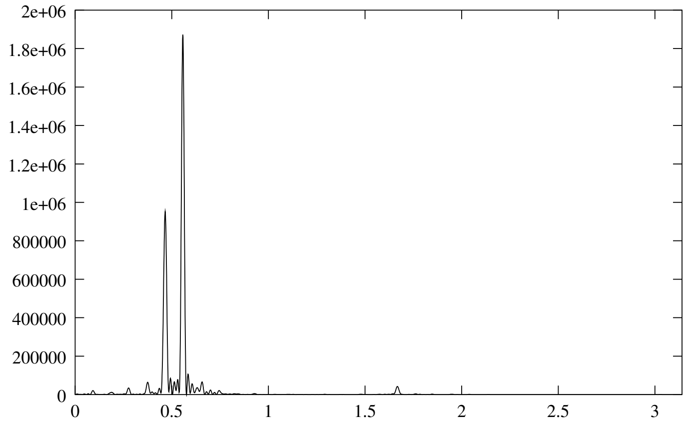

Figure 4. Plot of the periodogram function of the "ahh" sound.

problem is a model selection problem from a class of models. Define

$$R(L) = \min_{\lambda,\rho_j,\phi_j} \frac{1}{N} \sum_{n=1}^{N} \left(y(n) - \sum_{j=1}^{L} \rho_j \cos(nj\lambda - \phi_j) \right)^2. \tag{2.1}$$

Let us denote, $\hat{\lambda}$, $\hat{\rho}_j$ and $\hat{\phi}_j$ as the least squares estimators of λ, ρ_j and ϕ_j respectively if the model order is L. Note that in this case $\hat{\lambda}$, $\hat{\rho}_j$ and $\hat{\phi}_j$ depend on L, but we do not make it explicit for brevity. Consider

$$IC(L) = N \log R(L) + 2LC_N, \tag{2.2}$$

here C_N is a penalty function of N and it satisfies the following conditions;

$$(1) \quad \lim_{N\to\infty} \frac{C_N}{N} = 0 \quad \text{and} \quad (2) \quad \lim_{N\to\infty} \frac{C_N}{\log N} > 1. \tag{2.3}$$

The number of harmonics p^0 is estimated by the smallest value $\hat{p}$ such that;

$$IC(\hat{p}+1) > IC(\hat{p}). \tag{2.4}$$

Note that this criterion is like other Information Theoretic criteria used in model selection. But unlike AIC, BIC or MDL, here we do not have any fixed penalty function. Here the penalty function can be anything provided it satisfies conditions (2.3). Note that for fixed N, $N\log R(L)$ is a decreasing function of L. Therefore, as the model order increases $N\log R(L)$ gradually decreases, whereas the factor $2LC_N$ gradually increases and discourages to add more and more terms in the model. The factor, $2LC_N$ acts as a penalty function and the proposed criterion (2.4) determines the order of the model.

We can state the main result as follows;

THEOREM 2.1. *Let C_N be any function of N satisfying (2.3) and $\hat{p}$ is the smallest value such that $IC(\hat{p}+1) > IC(\hat{p})$, where $IC(j)$ is same as defined in (2.2). If $X(n)$ satisfies assumption 1, then $\hat{p}$ is a strongly consistent estimator of p^0.*

Before proving theorem 1, first we provide how to obtain $R(L)$. Consider

$$\mu_n^L = \sum_{j=1}^{L} \rho_j \cos(j\lambda n - \phi_j) = \sum_{j=1}^{L} \left[\rho_j \cos(\phi_j)\cos(j\lambda n) + \rho_j \sin(\phi_j)\sin(j\lambda n)\right].$$

Therefore,

$$\begin{bmatrix} \mu_1^L \\ \vdots \\ \mu_N^L \end{bmatrix} = \begin{bmatrix} \cos(\lambda) & \sin(\lambda) & \ldots & \cos(L\lambda) & \sin(L\lambda) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cos(N\lambda) & \sin(N\lambda) & \ldots & \cos(NL\lambda) & \sin(NL\lambda) \end{bmatrix} \begin{bmatrix} \rho_1\cos(\phi_1) \\ \rho_1\sin(\phi_1) \\ \vdots \\ \rho_L\cos(\phi_L) \\ \rho_L\sin(\phi_L) \end{bmatrix}$$

$$= \mathbf{A}_L(\lambda)\mathbf{b}_L \qquad \text{(say)}. \tag{2.5}$$

Now,

$$\sum_{n=1}^{N}\left(y(n) - \sum_{j=1}^{L}\rho_j\cos(j\lambda n - \phi_j)\right)^2 = \left(\mathbf{Y} - \mathbf{A}_L(\lambda)\mathbf{b}_L\right)^T\left(\mathbf{Y} - \mathbf{A}_L(\lambda)\mathbf{b}_L\right), \tag{2.6}$$

where $\mathbf{Y} = (y(1),\ldots,y(N))^T$. Using the separable regression technique of Richards (1961), it is observed that for fixed λ, minimization of $[\mathbf{Y} - \mathbf{A}_L(\lambda)\mathbf{b}_L]^T[\mathbf{Y} - \mathbf{A}_L(\lambda)\mathbf{b}_L]$ is obtained when $\hat{\mathbf{b}}_L = \left[\mathbf{A}_L(\lambda)^T\mathbf{A}_L(\lambda)\right]^{-1}\mathbf{A}_L(\lambda)^T\mathbf{Y}$. Now putting $\hat{\mathbf{b}}_L$ back in (2.6), we obtain

$$\left(\mathbf{Y} - \mathbf{A}_L(\lambda)\hat{\mathbf{b}}_L\right)^T\left(\mathbf{Y} - \mathbf{A}_L(\lambda)\hat{\mathbf{b}}_L\right) = \mathbf{Y}^T\left(\mathbf{I} - \mathbf{P}_{\mathbf{A}_L(\lambda)}\right)\mathbf{Y} = Q(\lambda) \quad \text{(say)}, \tag{2.7}$$

where $\mathbf{P}_{\mathbf{A}_L(\lambda)} = \mathbf{A}_L(\lambda)\left(\mathbf{A}_L(\lambda)^T\mathbf{A}_L(\lambda)\right)^{-1}\mathbf{A}_L(\lambda)^T$ is the projection operator on the column space spanned by the columns of $\mathbf{A}_L(\lambda)$. Therefore, the least squares estimator of λ can be obtained by minimizing $Q(\lambda)$ with respect to λ. Let us look at $\frac{1}{N}\mathbf{Y}^T\left(\mathbf{P}_{\mathbf{A}_L(\lambda)}\right)\mathbf{Y}$ for large N.

$$\frac{1}{N}\mathbf{Y}^T\left(\mathbf{P}_{\mathbf{A}_L(\lambda)}\right)\mathbf{Y} = \left(\frac{1}{N}\mathbf{Y}^T\mathbf{A}_L(\lambda)\right)\left(\frac{1}{N}\mathbf{A}_L(\lambda)^T\mathbf{A}_L(\lambda)\right)^{-1}\left(\frac{1}{N}\mathbf{A}_L(\lambda)^T\mathbf{Y}\right).$$

Note that

$$\lim_{N\to\infty}\frac{1}{N}\mathbf{A}_L(\lambda)^T\mathbf{A}_L(\lambda) = \frac{1}{2}\mathbf{I}_{2L}, \tag{2.8}$$

here $\mathbf{I}_{2L}$ denotes the identity matrix of order $2L$. It implies that for large N,

$$\begin{aligned}\frac{1}{N}\mathbf{Y}^T\left(\mathbf{P}_{\mathbf{A}_L(\lambda)}\right)\mathbf{Y} &\approx 2\sum_{j=1}^{L}\left[\left(\frac{1}{N}\sum_{n=1}^{N}y(n)\cos(j\lambda n)\right)^2+\left(\frac{1}{N}\sum_{n=1}^{N}y(n)\sin(j\lambda n)\right)^2\right]\\ &= 2\sum_{j=1}^{L}\left|\frac{1}{N}\sum_{n=1}^{N}y(n)e^{ijn\lambda}\right|^2,\end{aligned}$$

where $i=\sqrt{-1}$. Therefore, for large N, the least squares estimator of λ, say $\hat{\lambda}$, can be obtained by maximizing

$$I_Y(\lambda)=\sum_{j=1}^{L}\left|\frac{1}{N}\sum_{n=1}^{N}y(n)e^{ijn\lambda}\right|^2,$$

with respect to λ and therefore $R(L)$ can be written as

$$R(L)=\frac{1}{N}\mathbf{Y}^T\left(\mathbf{I}-\mathbf{P}_{\mathbf{A}_L(\hat{\lambda})}\right)\mathbf{Y}.$$

It is interesting to observe that to compute $R(L)$ it is not needed to calculate $\hat{\rho}_j$ and $\hat{\phi}_j$ explicitly. Moreover, the minimization of $\frac{1}{N}\mathbf{Y}^T\left(\mathbf{I}-\mathbf{P}_{\mathbf{A}_L(\lambda)}\right)\mathbf{Y}$ and maximization of $\frac{1}{N}I_Y(\lambda)$ are both one dimensional processes and they can be performed easily.

3. Proof of the consistency result

In this section, we provide the proof of theorem 1.

PROOF OF THEOREM 1. Observe that, we need to show

$$IC(0)>IC(1)>\cdots>IC(p^0-1)>IC(p^0)<IC(p^0+1).$$

Consider two cases separately.

Case I: $L<p^0$

$$\begin{aligned}\lim_{N\to\infty}R(L) &= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left(y(n)-\sum_{j=1}^{L}\hat{\rho}_j\cos(j\hat{\lambda}n-\hat{\phi}_j)\right)^2\\ &= \lim_{N\to\infty}\left[\frac{1}{N}\mathbf{Y}^T\mathbf{Y}-2\sum_{j=1}^{L}\left|\frac{1}{N}\sum_{n=1}^{N}y(n)e^{ijn\hat{\lambda}}\right|^2\right]=\sigma^2+\sum_{j=L+1}^{p^0}\rho_j^{0^2}\quad\text{a.s.}\end{aligned}$$

Therefore, for $0\le L<p^0-1$,

$$\lim_{N\to\infty}\frac{1}{N}\left[IC(L)-IC(L+1)\right]=$$

$$\lim_{N\to\infty}\left[\log\left(\sigma^2+\sum_{j=L+1}^{p^0}\rho_j^{0^2}\right)-\log\left(\sigma^2+\sum_{j=L+2}^{p^0}\rho_j^{0^2}\right)-\frac{2C_N}{N}\right]\quad\text{a.s.}\tag{3.1}$$

and for $L=p^0-1$,

$$\lim_{N\to\infty}\frac{1}{N}\left[IC(p^0-1)-IC(p^0)\right]=\lim_{N\to\infty}\left[\log\left(\sigma^2+\rho_{p^0}^{0^2}\right)-\log\sigma^2-\frac{2C_N}{N}\right]\quad\text{a.s.}\tag{3.2}$$

Since $\frac{C_N}{N} \to 0$, therefore as $N \to \infty$ for $0 \le L \le p^0 - 1$,

$$\lim_{N\to\infty} \frac{1}{N}\left[IC(L) - IC(L+1)\right] > 0.$$

It implies that for large N, $IC(L) > IC(L+1)$, when $0 \le L \le p^0 - 1$.

Case II: $L = p^0 + 1$.

To prove this part we need the following lemmas.

Lemma 1. *(An et al.; 1983) Let us define,*

$$I_X(\lambda) = \left|\frac{1}{N}\sum_{n=1}^{N} X(n)e^{in\lambda}\right|^2.$$

If $X(n)$ satisfies assumption 1, then

$$\lim\sup_{N\to\infty} \max_{\lambda} \frac{NI_X(\lambda)}{\sigma^2 \log N} \le 1 \quad a.s. \tag{3.3}$$

Lemma 2. *(Kundu; 1997) If $X(n)$ satisfies assumption 1, then*

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

Now consider

$$R(p^0+1) = \frac{1}{N}\mathbf{Y}^T\mathbf{Y} - 2\sum_{j=1}^{p^0}\left|\frac{1}{N}\sum_{n=1}^{N} y(n)e^{i\hat{\lambda}jn}\right|^2 - 2\left|\frac{1}{N}\sum_{n=1}^{N} y(n)e^{i\hat{\lambda}(p^0+1)n}\right|^2. \tag{3.4}$$

Note that $\hat{\lambda} \to \lambda^0$ *a.s.* as $N \to \infty$ (Nandi; 2002). Therefore, for large N

$$\begin{aligned}
&IC(p^0+1) - IC(p^0) \\
&= N\left(\log R(p^0+1) - \log R(p^0)\right) + 2C_N = N\left[\log\frac{R(p^0+1)}{R(p^0)}\right] + 2C_N \\
&\approx N\left[\log\left(1 - \frac{2\left|\frac{1}{N}\sum_{n=1}^{N} y(n)e^{i\lambda^0(p^0+1)n}\right|^2}{\sigma^2}\right)\right] + 2C_N \ \text{(using lemma 2)} \\
&\approx 2\log N\left[\frac{C_N}{\log N} - \frac{N\left|\frac{1}{N}\sum_{n=1}^{N} X(n)e^{i\lambda^0(p^0+1)n}\right|^2}{\sigma^2\log N}\right] \\
&= 2\log N\left[\frac{C_N}{\log N} - \frac{NI_X(\lambda^0(p^0+1))}{\sigma^2\log N}\right] > 0 \quad \text{a.s.}
\end{aligned}$$

Note that the last inequality follows because of the property of C_N and due to lemma 1.

4. Practical implementation

In this section, we present how to apply our proposed method in practice. Consider the following data $y(1),\dots,y(N)$ from the model (1.1). It is known that $p^0 \le K$ some fixed integer, although p^0 is not known. Compute $\hat{\lambda}$ by maximizing

$$I_Y(\lambda) = \sum_{j=1}^{K} \left| \frac{1}{N} \sum_{n=1}^{N} y(n) e^{ijn\lambda} \right|^2$$

with respect to λ and obtain an estimate of σ^2 by

$$\hat{\sigma}^2 = \frac{1}{N} \mathbf{Y}^T \left(\mathbf{I} - \mathbf{P}_{\mathbf{A}_K(\hat{\lambda})} \right) \mathbf{Y}.$$

For each data set, we calculated the sample variance and normalize the data, so that the error variance becomes 0.5. From now on we mainly talk about the normalized data unless otherwise mentioned. For brevity, we denote the normalized data also as $y(1),\dots,y(N)$ only. For a given choice of the penalty function C_N, we compute $IC(L)$, for different values of $L = 1,\dots K$ and choose $\hat{p}$ an estimate of p as suggested in section 2. Note that we have a wide choice of C_N, but we would like to choose that C_N for which $P[\hat{p} \ne p^0]$ is minimum. Now consider

$$P[\hat{p} \ne p^0] = P[\hat{p} < p^0] + P[\hat{p} > p^0] = \sum_{q=0}^{p^0-1} P[\hat{p} = q] + \sum_{q=p^0+1}^{K} P[\hat{p} = q].$$

Case I: $q < p^0$

$$P(\hat{p} = q) = P[IC(0) > IC(1) > \cdots > IC(q) < IC(q+1)] \le P\left[\log \frac{R(q)}{R(q+1)} < 2\frac{C_N}{N}\right].$$

Note that there exists a $\delta > 0$ such that for large N,

$$\log \frac{R(q)}{R(q+1)} > \delta \quad \text{a.s.}$$

It implies that for large N,

$$\log \frac{R(q)}{R(q+1)} > 2\frac{C_N}{N} \quad \text{a.s.}$$

Therefore, when N is large, for $q < p^0$,

$$P[\hat{p} = q] = 0. \tag{4.1}$$

Case II: $q > p^0$

$$P(\hat{p} = q) = P[IC(0) > IC(1) > \cdots > IC(q) < IC(q+1)].$$

Note that for large N,

$$P(\hat{p} = q) = P[IC(p^0) > \cdots > IC(q) < IC(q+1)],$$

as for large N

$$P[IC(0) > \cdots > IC(p^0)] = 1.$$

Therefore,

$$P[\hat{p} = q] = P\left[\log\frac{R(q+1)}{R(q)} + 2\frac{C_N}{N} > 0, \log\frac{R(j)}{R(j-1)} + 2\frac{C_N}{N} < 0, j = p^0+1,\ldots,q\right]. \tag{4.2}$$

From (4.1) it is immediate that for large N, the probability of under estimation is zero and to compute (4.2) we need to compute the joint distribution of $R(p^0), \ldots, R(K)$, which is not easy to obtain and it depends on the unknown parameters. Without knowing the actual parameter values, it is not possible to estimate the probability of over estimates or the probability of wrong detection. We use re-sampling or bootstrap technique similarly as Kundu and Mitra (2001) to estimate the probability of wrong detection and it will be used to estimate the number of components of the model (1.1). The idea is as follows. From a given realization of the data, first using the penalty function C_N, we estimate the order of the model as $M(C_N)$, using the method proposed in section 2. Now we normalize the data (for brevity we denote them as $y(1), \ldots, y(N)$ only) as described at the beginning of this section, so that the error variance of the data becomes $\frac{1}{2}$. We generate N independent identically distributed Gaussian random variables with mean zero and variance $\frac{1}{2}$, say $\epsilon(1), \ldots, \epsilon(N)$. We now obtain the bootstrap sample as

$$y(t)^B = y(t) + \epsilon(t); \qquad \text{for} \quad t = 1, \ldots, N.$$

Assuming $M(C_N)$ is the correct order model, we check for $q < M(C_N)$, whether

$$\log\frac{R(q)}{R(q+1)} < 2\frac{C_N}{N}$$

and for $q > M(C_N)$, check whether

$$\log\frac{R(q+1)}{R(q)} + 2\frac{C_N}{N} > 0, \quad \log\frac{R(j)}{R(j-1)} + 2\frac{C_N}{N} < 0, \quad j = p^0+1, \ldots q.$$

Repeating the process, say B times, we estimate $P(\hat{p} \neq p^0)$. Finally, we choose that C_N for which the estimated $P(\hat{p} \neq p^0)$ is minimum.

Some justifications regarding the proposed bootstrap estimates of $P(\hat{p} \neq p^0)$ can be given. Note that the realization of $y(t)^B$ can be thought of coming from a model equivalent to model (1.1). The proposed method works quite well with the simulated data and it can be observed in the next section.

5. Numerical results

In this section we perform some numerical experiments to present both the effectiveness of our method and usefulness of the analysis. We consider the following two models;

Model 1:

$$y(n) = \sum_{j=1}^{3} \rho_j^0 \cos(nj\lambda^0 - \phi_j^0) + X(n); \quad n = 1, \ldots, 50.$$

Here $\rho_1^0 = 2.0$, $\rho_2^0 = 2.5$, $\rho_3^0 = 2.5$, $\lambda^0 = 0.8796$, $\phi_1^0 = 1.2$, $\phi_2^0 = 0.95$ and $\phi_3^0 = 0.75$. $X(n) = 0.5e(n-1) + e(n)$. Similarly

Model 2:

$$y(n) = \sum_{j=1}^{4} \rho_j^0 \cos(nj\lambda^0 - \phi_j^0) + X(n); \quad n = 1, \ldots, 50.$$

Here $\rho_1^0 = 2.5$, $\rho_2^0 = 2.0$, $\rho_3^0 = 3.5$ and $\rho_4^0 = 1.0$, $\lambda^0 = 0.75398$, $\phi_1^0 = 0.5$, $\phi_2^0 = 0.9$, $\phi_3^0 = 0.75$ and $\phi_4^0 = 0.5$. $X(n) = 0.5e(n-1) + e(n)$.

In both the cases $e(n)$'s are i.i.d. Gaussian random variables with mean zero and variance τ^2. To assess the sensitivity of the method to different noise levels we consider three different τ^2, namely $\tau^2 = 0.5$, 0.75 and 1.0. For illustration purposes we plot two data sets generated using Models 1 and 2 in Figures 5 and 6 respectively. The corresponding periodogram functions are plotted in Figures 7 and 8 respectively. It is well-known that the number of peaks in the periodogram function plot roughly gives an estimate of the number of components p^0. But it depends on the magnitude of the amplitude associated with each effective frequency and the error variance. If a particular amplitude is relatively small as compared to others, that component may not be significant in the periodogram plot. Considering this fact we have included Model 2 in our simulation study. From Figure 7 it is quite clear that there are three peaks. Figure 8 also exhibits three peaks properly and it is not clear that the data are generated using model (1.1) with four harmonics. Therefore, it seems that estimating the number of components in Model 1 is an easier problem than to detect the number of components for Model 2.

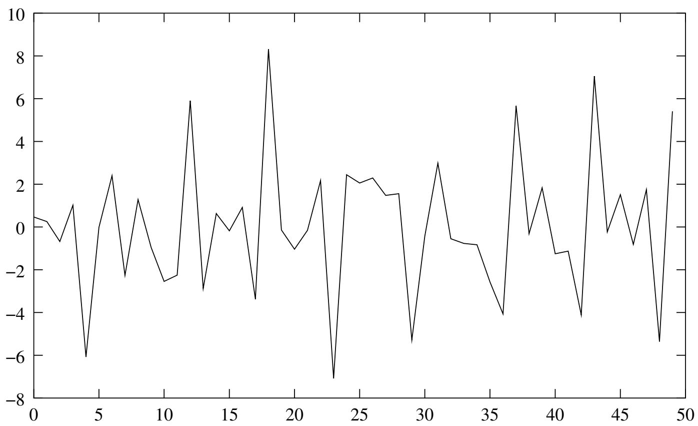

Figure 5. Plot of the data generated by Model 1 with $N = 50$ and error variance $= 1.0$.

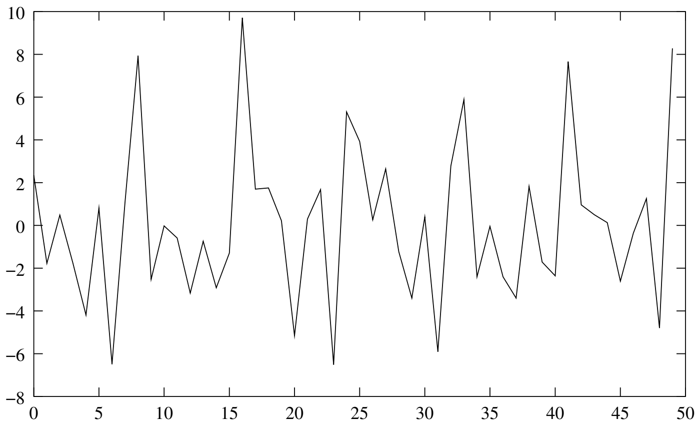

Figure 6. Plot of the data generated by Model 2 with $N = 50$ and error variance $= 1.0$.

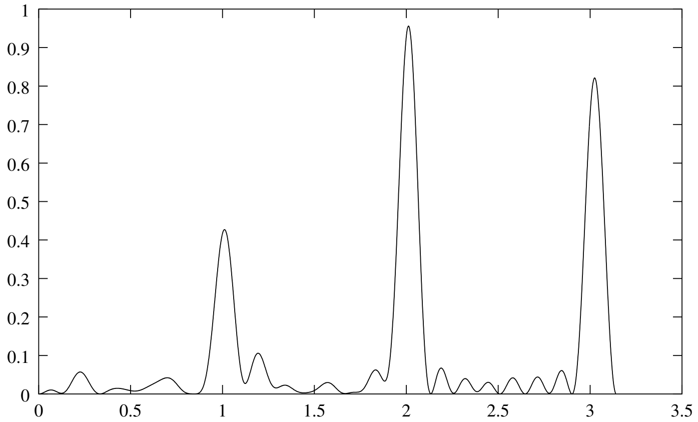

Figure 7. Plot of the periodogram function of the data plotted in figure 5.

Now we estimate the model order p^0 using the method proposed in section 2. It is assumed in both cases that the number of components can be at most 6, *i.e.* $K = 6$. We take 12 different C_N's, all of them satisfy (2.3). We denote them as $C_N(1), \ldots, C_N(12)$ and they are as follows; $C_N(1) = N^{.2} \log\log N$, $C_N(2) = N^{.3}$, $C_N(3) = \frac{(\log N)^3}{\log\log N}$, $C_N(4) = N^{.4}$, $C_N(5) = (\log N)^{1.1}$, $C_N(6) = (\log N)^{1.2}$, $C_N(7) = (\log N)^{1.3}$, $C_N(8) = \frac{N^{.5}}{(\log N)^{.9}}$, $C_N(9) = (\log N)^{1.4}$, $C_N(10) = \frac{N^{.4}}{\log N}$, $C_N(11) = \frac{N^{.5}}{\log N}$ and $C_N(12) = \frac{N^{.6}}{\log N}$. The main idea is to choose a wide variety of C_N's. For each simulated data vector we estimate p for all C_N's. Then this procedure is replicated 1000 times to obtain the probability of correct estimates (PCE's) and the probability of wrong estimates (PWE's). We mainly report the PCE's. The results are reported in Tables 1 and 2 for different penalty functions and for different error variances for Model 1 and Model 2 respectively.

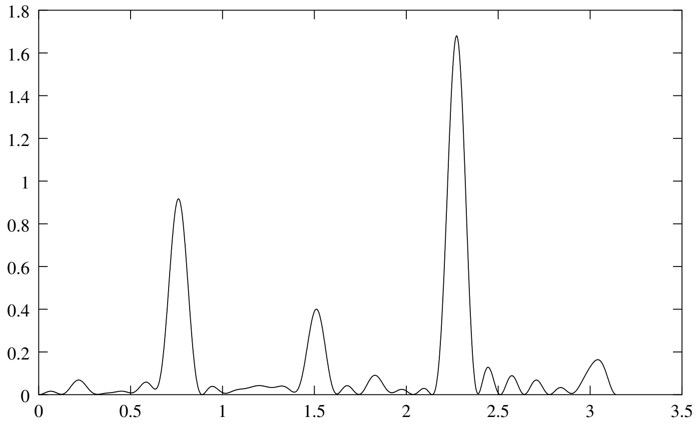

Figure 8. Plot of the periodogram function of the data plotted in figure 6.

Some of the points are quite clear from the Tables. First of all the performances of all $C_N(j)$'s improve as the error variance decreases, which is not surprising. The performances of the different $C_N(j)$'s vary from one extreme to the other. Some of the $C_N(j)$'s can detect the correct order model all the times considered here, whereas some of the $C_N(j)$'s can not detect the correct order model at all, although all of them satisfy (2.3) . It is observed that the correct detection for Model 1 is much easier than Model

Table 1. The PCE's for different penalty functions for Model 1.

Penalty	$\tau^2 = 0.5$	$\tau^2 = 0.75$	$\tau^2 = 1.0$
$C_N(1)$	1.000	1.000	1.000
$C_N(2)$	1.000	1.000	1.000
$C_N(3)$	0.000	0.000	0.000
$C_N(4)$	1.000	1.000	1.000
$C_N(5)$	1.000	1.000	1.000
$C_N(6)$	1.000	1.000	1.000
$C_N(7)$	1.000	1.000	1.000
$C_N(8)$	0.996	0.996	0.996
$C_N(9)$	1.000	1.000	1.000
$C_N(10)$	0.967	0.967	0.967
$C_N(11)$	0.991	0.991	0.990
$C_N(12)$	1.000	1.000	1.000

2. Typically, a particular C_N may not work for a particular model but may work for some others. For example, using $C_N(9)$ the proposed method always detects the correct order for all error variances for Model 1 whereas the same is not true in case of Model 2. Now using the bootstrap method (based on 100 replications) proposed in the previous section we estimate the number of components for both the models and for different error variances. The PCE's for both the models for different error variances are reported in Table 3. For comparison purposes we consider AIC and BIC type estimators also, as they have been considered as standard criteria in the literature. The results are reported in Table 3.

From Table 3, it appears that the proposed method works very well and performs better than both AIC and BIC, in both the cases for different error variances. Although any particular penalty function satisfying condition (2.3) may not work well for all the

Table 2. The PCE's for different penalty functions for Model 2.

Penalty	$\tau^2 = 0.5$	$\tau^2 = 0.75$	$\tau^2 = 1.0$
$C_N(1)$	0.998	0.998	0.996
$C_N(2)$	0.999	0.999	0.994
$C_N(3)$	0.000	0.000	0.000
$C_N(4)$	0.941	0.869	0.799
$C_N(5)$	0.978	0.923	0.865
$C_N(6)$	0.873	0.780	0.702
$C_N(7)$	0.614	0.525	0.459
$C_N(8)$	0.988	0.988	0.988
$C_N(9)$	0.286	0.264	0.235
$C_N(10)$	0.928	0.926	0.926
$C_N(11)$	0.979	0.979	0.979
$C_N(12)$	0.995	0.995	0.995

Table 3. The PCE's for Model 1 and Model 2 using the proposed method, AIC and BIC.

Method	Model	$\tau^2 = 0.5$	$\tau^2 = 0.75$	$\tau^2 = 1.0$
Proposed	Model 1	1.000	1.000	1.000
	Model 2	1.000	1.000	0.998
AIC	Model 1	0.884	0.883	0.883
	Model 2	0.942	0.942	0.944
BIC	Model 1	0.881	0.881	0.879
	Model 2	0.937	0.935	0.935

models and for different error variances but the present method works quite well for different models and for different error variances. It may not be surprising, because here we have a class of penalty functions and from there we are going to choose the *best* possible one.

Now to see how our proposed method works when the error variance is quite high, we generate a sample from Model 2 with error variance $\tau^2 = 5.0$. The periodogram plot of the generated data set is provided in Figure 9. It is not obvious from the periodogram

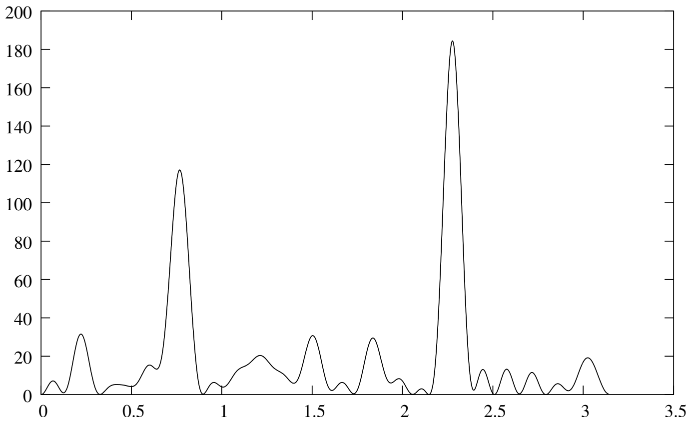

Figure 9. Periodogram plot of the data generated using Model 2 when $\tau^2 = 5.0$.

plot that $p^0 = 4$ in this case. Interestingly, our proposed method detects correctly the number of components in this case also. It should be mentioned here that although using our method it is possible to estimate the number of components correctly, but because of such high variance any existing frequency estimation technique can not estimate the frequencies very accurately in this case. The analysis of any data vector depends on the estimation of p^0, as well as the estimation of λ^0, ρ_i^0's and ϕ_i^0's.

6. Data analysis

In this section we estimate the number of harmonics present in two data sets. The periodogram functions of the datasets motivate us to use model (1.1) and thus the proposed method is used in estimating the model order. The first data set is the vowel sound 'uuu'. It contains 512 signal values sampled at 10kHz frequency. The time series plot is presented in Figure 1 after mean correction and the corresponding sample periodogram plot is in Figure 3. From Figure 3, it is quite apparent that the number of harmonics should be 4. Using the penalty functions mentioned in the previous section and $K = 9$, we obtain the following results using the first step of our method. The penalty functions $C_N(4)$, $C_N(5)$, $C_N(6)$, $C_N(7)$, $C_N(9)$ and $C_N(12)$ estimate the number of harmonics to be 4, the penalty function $C_N(1)$, $C_N(2)$, $C_N(8)$, $C_N(10)$ and $C_N(11)$ estimate it to be 5, whereas $C_N(3)$ estimates it to be 2 only. Therefore, it is not clear which one is to be taken. Now using the second step (based on maximum PCE, obtained using resampling technique discussed in section 4) of our method we obtain the estimate of the number of harmonics to be 4. The residual sums of squares for different orders are as follows: 0.454586, 0.081761, 0.061992, 0.051479, 0.050173, 0.050033, 0.049870, 0.049866 and 0.049606 for the model orders 1, 2, ... and 9 respectively. The residual sum of squares is a decreasing function of p and it almost stabilizes at $\hat{p}$. So it also indicates that the number of harmonics should be 4. So for "uuu" data set the periodogram function and residual sum of squares also give reasonable estimate of the order. Now using $\hat{p} = 4$, we estimate the frequency, amplitudes and phases as $\lambda = 0.113275$, $\rho_1 = 0.327591$, $\rho_2 = 0.861467$, $\rho_3 = 0.192446$, $\rho_4 = 0.144400$, $\phi_1 = -2.565801$, $\phi_2 = 1.142973$, $\phi_3 = 2.972859$, and $\phi_4 = 2.848901$. We further calculate the predicted values using the estimated parameters and they are plotted in Figure 10 along with the observed values. The estimated plot matches quite well with the observed one.

The second data set is the sound 'ahh'. It contains 340 signal values sampled at 10kHz frequency. The data set and the corresponding periodogram function are plotted in Figures 2 and 4 respectively. From the periodogram plot it is not clear exactly how many harmonics are present. As before using the penalty functions mentioned in the previous section we estimate the number of harmonics present in the data and the result is as follows. The penalty functions $C_N(1)$, $C_N(2)$, $C_N(4)$, $C_N(5)$, $C_N(6)$, $C_N(7)$, $C_N(9)$ and $C_N(12)$ estimate it to be 6, $C_N(8)$ and $C_N(11)$ estimate it 7 and $C_N(3)$ and $C_N(10)$ estimate it to be 1 and 8 respectively. The second step of our method suggests that the number of harmonics should be 6. The residual sums of squares for the different orders are 0.49690, 0.49527, 0.49014, 0.48345, 0.36156, 0.11701, 0.11361, 0.11169 and 0.11132 for the model order 1, 2, ..., 9 respectively. Therefore, the residual sums of squares suggest that the number of harmonics should be 6. Similarly as "uuu" data set, we estimate the other parameters i.e. the frequency, amplitudes and phases as follows: $\lambda = 0.0922486$, $\rho_1 = 0.078656$, $\rho_2 = 0.056695$, $\rho_3 = 0.101506$, $\rho_4 = 0.111626$, $\rho_5 = 0.491332$, $\rho_6 = 0.700800$, $\phi_1 = 1.623784$, $\phi_2 = 1.998672$, $\phi_3 = 2.554295$, $\phi_4 = -2.920414$, $\phi_5 = -2.219181$ and $\phi_6 = 0.770404$. The observed and estimated "ahh" are plotted in Figure 11. Thus, for this data set although the periodogram plot does not provide the exact indication of the number of harmonics present but the proposed method provides a reasonable estimate of the model order.

To implement the resampling technique in obtaining the probability of correct estimates for different penalty functions, the observed values are scaled such that the variance is equal is .5. Thus the parameter estimates correspond to the data vectors with mean zero and variance .5 whereas the predicted values are obtained with original scaling to compare the observed values.

7. Some generalizations

So far we have assumed that all the adjacent harmonics of the fundamental frequency are present in the model. In this section we generalize our result to those models where some of the adjacent harmonics might be absent. It is observed that our method can be extended for those models but it involves much heavier computations. Using the idea of

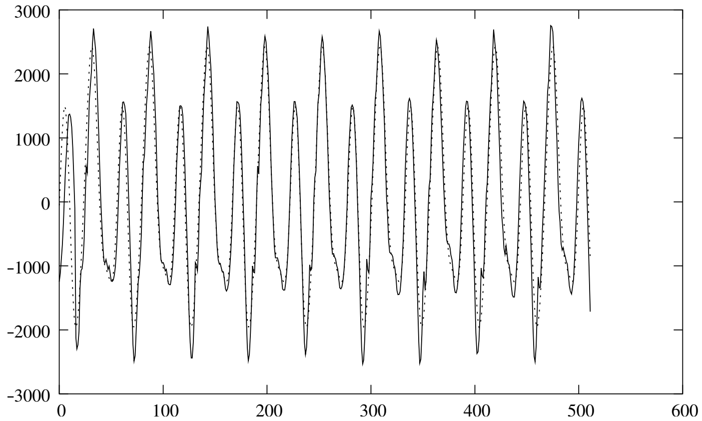

Figure 10. Plot of the observed (continuous line) and estimated (dotted) “uuu” sound.

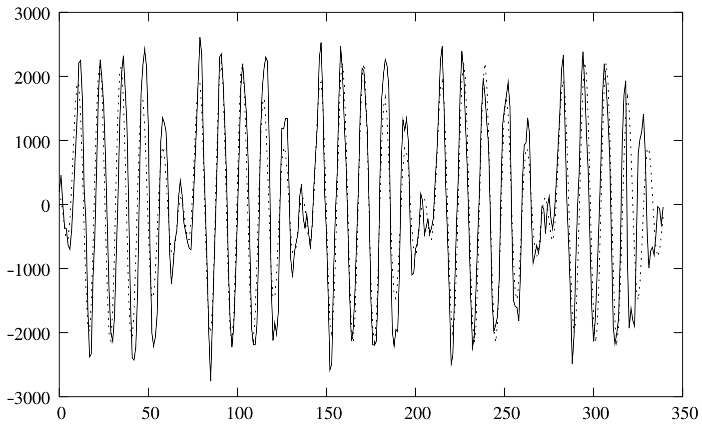

Figure 11. Plot of the observed (continuous line) and estimated (dotted) “ahh” sound.

Sakai (1993), we write the model as follows;

$$y(n) = \sum_{j=1}^{p^0} v_j^0 \rho_j^0 \cos(nj\lambda^0 - \phi_j^0) + X(n); \qquad n = 1, \ldots N. \tag{7.1}$$

Here $\rho_j^0, \lambda^0, \phi_j^0$ and $X(n)$ are same as defined before. The variable v_j^0 is an indicator function as follows;

$$v_j^0 = \begin{cases} 1 & \text{if the harmonics } j\lambda^0 \text{ is present} \\ 0 & \text{otherwise.} \end{cases}$$

In this case it is assumed that the maximum number of harmonics (including the missing ones) can be at most K. Consider a vector $\mathbf{v} = (v_1, \ldots, v_K)$ of length K, where each v_i is either 0 or 1. Therefore, there are total 2^K such vector $\mathbf{v}$ and for each model there is a one to one correspondence to a particular vector $\mathbf{v}$. For each $\mathbf{v}$, compute

$$R(\mathbf{v}) = \min_{\lambda, \rho_j, \phi_j} \frac{1}{N} \sum_{n=1}^{N} \left(y(n) - \sum_{j=1}^{K} \rho_j v_j \cos(nj\lambda - \phi_j) \right)^2. \tag{7.2}$$

Consider

$$IC(\mathbf{v}) = N \log R(\mathbf{v}) + 2C_N \times (\# \text{ of } 1's \text{ in } \mathbf{v}), \tag{7.3}$$

here C_N is same as defined before. Among the 2^K choices of $\mathbf{v}$, choose that particular $\mathbf{v}$ for which $IC(\mathbf{v})$ is minimum. Using the same steps as in section 2, it can be seen that

$$R(\mathbf{v}) \approx \frac{1}{N}\mathbf{Y}^T\mathbf{Y} - 2\max_{\lambda} \sum_{j\in\{k;v_k=1\}} \left| \frac{1}{N}\sum_{n=1}^{N} y(n)e^{inj\lambda} \right|^2 .$$

Using the same idea as in section 3, it can be shown that the proposed method provides consistent estimator of $\mathbf{v}^0$, the true value of $\mathbf{v}$. The re-sampling procedure can be performed along the same manner. Therefore, our method can be used in this situation. But one point should be mentioned that for model (1.1) we need to search among K possible models whereas for the model (7.1), the search space involves 2^K possible models.

8. Conclusions

In this paper we propose a new technique to estimate the number of harmonics of a fundamental frequency model using the penalty function approach. We prove the strong consistency of the proposed method under the assumption of stationary errors. We propose a re-sampling technique to estimate the probability of wrong detection and that was used to compute the number of components of the fundamental frequency model. The performances of the proposed method are much better than the classical methods like AIC or BIC.

One final point should be mentioned that in many practical situations it is possible that more than one fundamental frequencies and their harmonics are present in the data. If this kind of prior knowledge is available and the clusters are well separated, then the parameters of one cluster may be estimated by using the data set where all the other clusters have been suppressed by a suitable band-pass filter. In each step of estimation, the whole data set after being filtered may be used to estimate only the parameters within one cluster. Further work is needed in that direction.

Acknowledgements

The authors would like to thank two referees for some helpful suggestions and the past editor Professor Tatsuya Kubokawa for his encouragements.

References

[1] An, H.Z., Chen, Z.G. and Hannan, E.J. (1983). The maximum of the periodogram. *J. Multivariate Anal.*, **13**, 383–400.

[2] Fisher, R.A. (1929). Tests of significance in Harmonic analysis. *Proc. Roy. Soc. London Ser. A*, **125**, 54–59.

[3] Fuchs, J.J. (1988). Estimating the number of sinusoids in additive white noise. *IEEE Trans. Acoustics, Speech and Signal Processing*, **41**, 1846–1853.

[4] Hannan, E.J. (1973). Estimation of frequencies. *J. Appl. Probab.*, **10**, 510–519.

[5] Kaveh, M., Wang, H. and Hung, H. (1987). On the theoretical performance of a class of estimators of the number of narrow band sources. *IEEE Trans. Acoustics, Speech and Signal Processing*, **35**, 1350–1352.

[6] Kay, S. M. and Marple, S. L. (1981). Spectral analysis - a modern perspective. *Proc. IEEE*, **69**, 1380–1419.

[7] Kundu, D. (1997). Asymptotic properties of the least squares estimators of sinusoidal signals. *Statistics*, **30**, 221–238.

[8] Kundu, D.(1998). Estimating the number of sinusoids and its performance analysis. *J. Statist. Comput. Simulation*, **60**, 347–362.

[9] Kundu, D. and Mitra, A. (2001). Estimating the number of signals of the damped exponential models. *Comput. Statist. Data Anal.*, **36**, 245–256.

[10] Nandi, S. (2002). Analyzing some non-stationary signal processing models. *Ph.D. Thesis*, Indian Institute of Technology Kanpur.

[11] Nandi, S. and Kundu, D. (2003). Estimating the parameter of the fundamental frequency model. *Statistical Methods and Applications*, **12**, 341–360.

[12] Quinn, B.G. (1989). Estimating the number of terms in a sinusoidal regression. *J. Time Ser. Anal.*, **10**, 71–75.

[13] Quinn, B.G. and Thomson, P.J. (1991). Estimating the frequency of a periodic function. *Biometrika*, **78**, 1, 65–74.

[14] Richards, F.S.G. (1961). A method of maximum likelihood estimation. *J. Roy. Statist. Soc. Ser. B*, **23**, 469–475.

[15] Sakai, H. (1990). An application of a BIC-type method to harmonic analysis and a new criterion for order determination of an AR process. *IEEE Trans. Acoustics, Speech and Signal Processing*, **38**, 999–1004.

[16] Sakai, H. (1993). The determination of the number of terms in a multichannel sinusoidal regression. *J. Time Ser. Anal.*, **14**, 621–628.

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

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

[19] Wang, H. (1993). An AIC type estimator for the number of cosinusoids. *J. Time Ser. Anal.*, **14**, 433–440.

[20] Wang, H. and Kaveh, M. (1986). On the performance of signal-subspace processing Part I: narrow band systems. *IEEE Trans. Acoustics, Speech and Signal Processing*, **34**, 1201–1209.

[21] Whittle, P. (1953). The simultaneous estimation of a time series harmonics components and covariance structure. *Trab. Estad.*, **3**, 43–57.