WEIGHTED ESTIMATION OF HARMONIC COMPONENTS IN A MUSICAL SOUND SIGNAL

By Rafael A. Irizarry

Johns Hopkins University

First Version received November 1998

Abstract. The study of musical sound has become a popular research field. Harmonic regression signal plus noise statistical models have been used to analyze sound signals. However, it is common to give estimates of harmonic parameters without indications of their uncertainties. Least squares estimates for harmonic models have been studied and asymptotic variance expression have been developed. In practice, window-based estimates are used. This paper studies the statistical properties of such estimates; in particular, we use asymptotic variance expressions to develop standard errors and construct confidence intervals. We present applications and examples of the statistical techniques to musical sound signal analysis.

Keywords. Harmonic regression; asymptotic variance; weighted least squares estimates; musical sound signals.

1. Introduction

Time series analysis has been applied to music in various ways (Brillinger and Irizarry, 1998; Irizarry, 1998). In this paper, the particular application that will be examined is the analysis of sound signals produced by musical instruments. Researchers in this field are interested in, for example, the problem of determining what particular characteristics of the sound produced by musical instruments permit humans to distinguish one instrument from another, what musicians call *timbre* (Grey, 1977). With today's technology, we are able to process sounds in a data analytic fashion. Risset and Mathews (1969) were the first to successfully make use of the computer to analyze the sound produced by musical instruments, by using discrete samples of the continuous sound signal as data. Brillinger and Irizarry (1998) provide more details on the quantification of sound signals.

When fluctuations of air are approximately periodic, with period in the audible range, we perceive what musicians have defined as a *pitch* (Pierce, 1992, ch. 2).

Figure 1 shows a segment of the sound signal produced by a clarinet playing concert pitch A (441 Hz).

Physical modeling (Fletcher and Rossing, 1991) suggests that, within short segments, we model musical sound signals as summations of sinusoidal components, as done in the additive synthesis model proposed by Risset and Mathews (1969). Serra and Smith (1991) incorporated a non-sinusoidal residual

0143-9782/01/03 000–000 JOURNAL OF TIME SERIES ANALYSIS Vol. 22, No. 3

B	0	8	8	1	5	1	5	B	Dispatch: 21.8.01	Journal: JTSA
Journal No.				Manuscript No.					Author Received:	No. of pages: 20

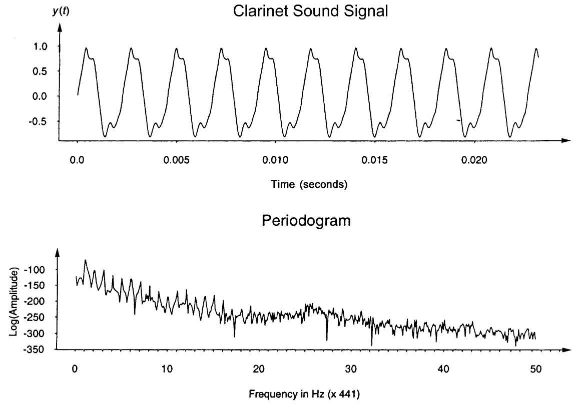

FIGURE 1. 23 millisecond stretch and periodogram of a clarinet sound.

part to the additive synthesis and modelled it as an additive stochastic signal. Since, many have used the so-called additive synthesis plus residual model (Rodet, 1997) in which short segments, called *time-frames* (durations of between 5 and 100 milliseconds), of a signal are modelled with

$$y_t = \sum_{k=1}^{K} \rho_k \cos(\omega_k t + \phi_k) + \epsilon_t \qquad t = 1, \ldots, T \tag{1}$$

An implicit convention is that $\omega_1 < \omega_2 < \cdots < \omega_k$, with ω_1 usually associated with the frequency related to the pitch or note being played and called the *fundamental frequency*. The component related to the frequency ω_k is called the *kth partial*. The behaviour of such partials is believed to be essential in determining timbre (Grey, 1977), thus estimating the parameters of model (1) is of interest. However, in the sound analysis and synthesis literature, it is common to give estimates of sinusoidal parameters without indications of their uncertainties. The variation of the estimated parameters for different segments of the signal are sometimes explained with deterministic arguments. Under the assumption that the signals contain a stochastic element, the possibility exists that such variations are due to chance alone. In this paper, this possibility is explored by defining estimates for which statistical properties can be studied.

The estimation procedures presented in current sound analysis research are based on the assumption that within appropriately chosen time-frames, the model given by (1) holds; in which case, it is equivalent to a harmonic regression signal plus noise model like the one presented in, for example, Walker (1971).

In Walker (1971), Hannan (1973), Brown (1990), Quinn and Thomson (1991), and Hassan (1982), among others, least squares estimates are presented for models with harmonic regression signal plus noise. Consistency is shown for these estimates and asymptotic variance expressions are developed. Since, for this particular application, models are fit to obtain estimates of parameters that are thought to change from time-frame to time-frame, it is only natural to consider window-based estimates. In this paper, the results obtained by Walker (1971) and Hannan (1973) will be generalized to window-based estimates equivalent to weighted least-squares. This will permit one to develop standard errors and confidence intervals for estimates obtained for the music signals.

The remainder of this paper is organized as follows. Section 2 presents the harmonic plus noise model and the weighted least squares estimates for its parameters. Section 3 summarizes key asymptotic theory developed for least squares estimates, with an extension to the weighted case. In Section 4, we present examples of how the estimates and asymptotic theory developed in Sections 2 and 3 can be useful as a data exploration tool in the study of sound signals. Some final remarks are given in Section 5.

2. HARMONIC REGRESSION MODEL

Many signals in nature have been statistically analyzed via sinusoidal regression models (Brillinger, 1977). The harmonic regression signal plus noise model is defined by

$$y_t = s(t; \beta_0) + \epsilon_t \qquad t = 1, \ldots, T$$

where

$$s(t; \beta_0) = \sum_{k=1}^{K} \{A_{k,0} \cos(\omega_{k,0} t) + B_{k,0} \sin(\omega_{k,0} t)\} \tag{2}$$

and $\{\epsilon_t\}$ stationary stochastic process.

This model has been studied by various authors. Under the assumption that $\{\epsilon_t\}$ is white noise with finite variance, Walker (1971) presents estimates that are asymptotically equivalent to least squares estimates. Consistency is shown for these estimates and asymptotic variance expressions are developed. Hannan (1973) does the same under the assumption that $\{\epsilon_t\}$ is ergodic and purely non-deterministic. For a more general model, with modulating amplitudes and under the assumption that the noise is a linear processes satisfying a mixing condition, Hassan (1982) finds estimates that are consistent and asymptotically normal as well. Brown (1990) and Quinn and Thomson (1991) develop similar results when adding the constraint that $\omega_{k,0} = k\lambda_0$, for some fundamental frequency λ_0, to model (2).

For weighted least squares estimates, the result of consistency follows in a similar fashion to the unweighted case. However, some work is needed to obtain asymptotic variance expressions. In the work that follows, we will be presenting the results obtained by Walker (1971) and Hannan (1973), for estimates that are asymptotically equivalent to weighted least squares, under the assumption that the stationary $\{\epsilon_t\}$ has autocovariance function $c_{\epsilon\epsilon}(u) = \mathrm{cov}\{\epsilon_{t+u}, \epsilon_t\}$, satisfies Assumption 1 below, and has power spectrum

$$f_{\epsilon\epsilon}(\lambda) = \frac{1}{2\pi}\sum_{u} c_{\epsilon\epsilon}\exp(-\mathrm{i}\lambda u) \qquad \infty < \lambda < \infty$$

ASSUMPTION 1. The error series $\{\epsilon_t\}$ is a strictly stationary real valued random process all of whose moments exist, with zero mean, and with $c_{\epsilon\ldots\epsilon}(u_1, \ldots, u_{L-1})$ the joint cumulant function of order L of the random process $\{\epsilon_t\}$ for $L=2, 3, \ldots$. Furthermore, the

$$C_L = \sum_{u_1=-\infty}^{\infty} \cdots \sum_{u_{L-1}=-\infty}^{\infty} |c_{\epsilon\ldots\epsilon}(u_1, \ldots, u_{L-1})|$$

satisfy

$$\sum_{k} \frac{C_k z^k}{k!} < \infty$$

for z in a neighbourhood of 0.

This assumption requires that the stochastic process $\{\epsilon_t\}$ have a short span of dependence, that is that the random variables ϵ_t and ϵ_s are less statistically dependent on each other as they become more distant, i.e. as $|\, t - s \,| \to \infty$.

2.1. *Weighted least squares estimates*

The weighted least squares method consists of choosing $\tilde{\beta}$ to minimize the criterion

$$S_T(\beta) = \sum_{t=1}^{T} w\left(\frac{t}{T}\right)\{y_t - s(t, \beta)\}^2 \tag{3}$$

Here $w(s)$ is a weight function. Some of the results regarding the asymptotic behaviour of these estimates require that the weight function satisfy Assumption 2.

ASSUMPTION 2. The function $w(s)$ is non-negative, bounded, of bounded variation, has support [0, 1], $W_0 > 0$ *and,* $W_1^2 - W_0W_2 \neq 0$. *Here*

$$W_n = \int_0^1 s^n w(s)\mathrm{d}s \tag{4}$$

3. ASYMPTOTIC THEORY

Throughout we are going to need the following simple result, proved in the Appendix. Set

$$\Delta_n^T(\lambda) = \sum_{t=1}^{T} w\left(\frac{t}{T}\right) t^n \exp(\mathrm{i}\lambda t)$$

LEMMA 1. *If $w(t)$ satisfies Assumption 2 then we have for $n = 0, 1, \ldots$*

$$\lim_{T\to\infty} T^{-(n+1)} \Delta_n^T(\lambda) = W_n, \qquad \text{for } \lambda = 0, 2\pi \tag{5}$$

$$\Delta_n^T(\lambda) = O(T^n) \quad \text{for } 0 < \lambda < 2\pi \tag{6}$$

To prove consistency and asymptotic variance of the weighted least squares estimates, we first define the estimates $\hat{\beta}$ composed of $\hat{A}_{k,T}$, $\hat{B}_{k,T}$ and $\hat{\omega}_{k,T}$ for $k = 1, \ldots, K$

$$\hat{A}_{k,T} = \frac{2\sum_{t=1}^{T} w\left(\frac{t}{T}\right) y_t \cos(\hat{\omega}_{k,T} t)}{\sum_{t=1}^{T} w\left(\frac{t}{T}\right)} \tag{7}$$

$$\hat{B}_{k,T} = \frac{2\sum_{t=1}^{T} w\left(\frac{t}{T}\right) y_t \sin(\hat{\omega}_{k,T} t)}{\sum_{t=1}^{T} w\left(\frac{t}{T}\right)} \tag{8}$$

where if we write $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K)$ and $\hat{\boldsymbol{\omega}}_T = (\hat{\omega}_1, T, \ldots, \hat{\omega}_{K,T})$, $\hat{\boldsymbol{\omega}}_T$ is such that

$$q_T(\hat{\boldsymbol{\omega}}_T) = \max_{0 \le \boldsymbol{\omega} \le \pi} q_T(\boldsymbol{\omega}) \tag{9}$$

where q_T is defined by

$$q_T(\boldsymbol{\omega}) = \sum_{k=1}^{K} \left| T^{-1} \sum_{t=1}^{T} w\left(\frac{t}{T}\right) y_t \exp(\mathrm{i}t\omega_k) \right|^2$$

Notice that these estimates are the same ones presented in Walker (1971) and Hannan (1973) in the unweighted case, but now using tapered data $w(t/T)y_t$. Similar to the unweighted case, we notice that these estimates are asymptotically equivalent to the weighted least squares estimates and thus we may prove asymptotic results for the former to obtain the results for the latter. This result is best understood by first considering the case of one sinusoidal component ($K = 1$)

$$s(t; \beta_0) = A_0 \cos(\omega_0 t) + B_0 \sin(w_0 t)$$

with $\beta_0 = (A_0, B_0, \omega_0)'$ and then generalizing to the case of several partials.

As done in Walker (1971) for the unweighted case, we notice that if we define

$$\begin{aligned} R_T(\beta) = \sum_{t=1}^{T} w\left(\frac{t}{T}\right) y_t^2 + \tfrac{1}{2}(A^2 + B^2) \sum_{t=1}^{T} w\left(\frac{t}{T}\right) \\ - 2 \sum_{t=1}^{T} w\left(\frac{t}{T}\right) y_t \{A \cos(\omega t) + B \sin(\omega t)\} \end{aligned} \tag{10}$$

with $\beta = (A, B, \omega)$, then

$$S_T(\beta) - R_T(\beta) = \tfrac{1}{2} \sum_{t=1}^{T} w\left(\frac{t}{T}\right) \{(A^2 - B^2) \cos(2\omega t) + 2AB \sin(2\omega t)\} \tag{11}$$

Here $S_T(\beta)$ is the weighted residual sum of squares of equation (3). The difference in (11) is deterministic and, using Lemma 1, we can show it is bounded as $T \to \infty$ if $0 < \omega < \pi$.

By taking derivatives and solving when they are set to 0, we see that the ω that maximizes the periodogram of the tapered data $w(t/T)Y_t$ also minimizes $\mathrm{R}_T(\beta)$. This and (11) may be used to show that the estimates presented in (7), (8), and (9) are asymptotically equivalent to the weighted least squares estimates.

For the case of more than one frequency, model (2), the function corresponding to (10) whose minimization yields approximate weighted least squares estimators becomes

$$\begin{aligned} R_T(\beta) = \sum_{t=1}^{T} w\left(\frac{t}{T}\right) y_t^2 + \tfrac{1}{2} \sum_{k=1}^{K} (A_k^2 + B_k^2) \sum_{t=1}^{T} w\left(\frac{t}{T}\right) \\ - 2 \sum_{k=1}^{K} \sum_{t=1}^{T} w\left(\frac{t}{T}\right) y_t \{A_k \cos(\omega_k t) + B_k \sin(\omega_k t)\} \end{aligned} \tag{12}$$

Here $\beta = (\mathbf{A}, \mathbf{B}, \boldsymbol{\omega})' = (A_1, \ldots, A_K, B_1, \ldots, B_K, \omega_1, \ldots, \omega_K)'$. In this case, to obtain (12) from the weighted least squares equation (3), we need terms of the form

$$A_k A_l \sum_{t=1}^{T} w\left(\frac{t}{T}\right) \cos(\omega_k t) \cos(\omega_l t)$$

and

$$B_k B_l \sum_{t=1}^{T} w\left(\frac{t}{T}\right) \sin(\omega_k t) \sin(\omega_l t)$$

to be bounded, since they are included $S_T\ (\mathbf{A}, \mathbf{B}, \boldsymbol{\omega}) - R_T\ (\mathbf{A}, \mathbf{B}, \boldsymbol{\omega})$. Some conditions need to be imposed to avoid have the ω_k become too close together and thus prevent the estimators of two or more frequencies from converging in probability to the same value. An appropriate condition is

$$\lim_{T\to\infty} \min_{1\leq k\neq l\leq K} (T|\omega_k - \omega_l|) = \infty \tag{13}$$

See Walker (1971) for an example.

So we redefine the estimates of $\boldsymbol{\omega}_0 = (\omega_{1,0},\ldots, \omega_{k,0})'$ as the value that maximizes (9) but under a constraint satisfying (13).

We have shown that the estimates $\hat{\beta}$ defined by (7)–(9) are asymptotically equivalent to the weighted least squares estimates $\hat{\beta}$ defined by (3). Because the two main results of this paper are asymptotic results, we will abuse notation and use only $\hat{\beta}$ to denote both these estimates.

To prove consistency and asymptotic normality for the weighted least squares, or equivalently the estimates defined by (7)–(9), we need a result concerning the behaviour of the periodogram of the noise and its derivatives with respect to ω.

LEMMA 2. *Let the stationary random process $\{\epsilon_t\}$ satisfy Assumption 1 and let the weight function w (s) satisfy Assumption 2, then if*

$$p_T(\omega) = \left| T^{-(k+1)} \sum_{t=1}^{T} w\left(\frac{t}{T}\right) t^k \epsilon_t \exp(-\mathrm{i}t\omega) \right|$$

one has $k = 0, 1, \ldots$

$$\lim_{T\to\infty} \sup_{0\leq\omega\leq\pi} p_T(\omega) = 0, \text{ in probability}$$

REMARK 1. Lemma 2 has been shown to be true under different assumptions for the equally weighted case, $w(s) = 1$. In most cases the result for the weighted case follows similarly. Walker (1971) proves the Lemma for white noise with finite variance. Hannan (1973) proves it under ergodic and purely non-deterministic conditions. Brillinger (1986) proves a version of this Lemma for spatial point processes. Under Assumptions 1 and 2, Lemma 2 follows directly from Brillinger (1981, p. 98, Theorm 4.5.1)

We may now prove consistency of the weighted least squares estimates

THEOREM 1. *If $\{\epsilon_t\}$ satisfies Assumption 1 and the weight function $w(s)$ satisfies Assumption 2, then for* $0 < \omega_{k,0} < \pi$

$$\lim_{T\to\infty} \hat{A}_{k,T} = A_{k,0}$$
$$\lim_{T\to\infty} \hat{B}_{k,T} = B_{k,0}$$
$$\lim_{T\to\infty} T|\hat{\omega}_{k,T} - \omega_{k,0}| = 0$$

for $k = 1,\ldots, K$, *in probability.*

Remark 2. Using Lemmas 1 and 2, we may prove consistency in a similar way to Walker (1971) or Hannan (1973). In the Appendix, a sketch of the proof is given containing the key differences for the weighted case.

The following theorem describes the asymptotic distribution of the weighted least squares estimates. The result provides a way to construct standard errors and confidence intervals for our estimates.

Theorem 2. *Under the same conditions as Theorem 1, the vectors*

$$\{T^{\frac{1}{2}}(\hat{A}_{k,T} - A_{k,0}), T^{\frac{1}{2}}(\hat{B}_{k,T} - B_{k,0}), T^{\frac{3}{2}}(\hat{\omega}_{k,T} - \omega_{k,0})\}' \qquad k = 1, \ldots, K$$

converge in distribution to mutually independent normal vectors with zero mean and variance matrices

$$\frac{4\pi f_{\epsilon\epsilon}(\omega_{k,0})}{A_{k,0}^2 + B_{k,0}^2}\mathbf{V_k}$$

where

$$\mathbf{V_k} = \begin{bmatrix} c_1 A_{k,0}^2 + c_2 B_{k,0}^2 & -c_3 A_{k,0} B_{k,0} & -c_4 B_{k,0} \\ -c_3 A_{k,0} B_{k,0} & c_2 A_{k,0}^2 + c_1 B_{k,0}^2 & c_4 A_{k,0} \\ -c_4 B_{k,0} & c_4 A_{k,0} & c_0 \end{bmatrix} \tag{14}$$

Here

$$\begin{aligned} c_0 &= a_0 b_0 \\ c_1 &= U_0 W_0^{-2} \\ c_2 &= a_0 b_1 \\ c_3 &= a_0 W_1 W_0^{-2}(W_0^2 W_1 U_2 - W_1^3 U_0 - 2W_0^2 W_2 U_1 + 2W_0 W_1 W_2 U_0) \\ c_4 &= a_0(W_0 W_1 U_2 - W_1^2 U_1 - W_0 W_2 U_1 + W_1 W_2 U_0) \end{aligned} \tag{15}$$

where

$$\begin{aligned} a_0 &= (W_0 W_2 - W_1^2)^{-2} \\ b_n &= W_n^2 U_2 + W_{n+1}(W_{n+1} U_0 - 2W_n U_1) \quad \text{for } n = 0, 1 \end{aligned}$$

Here W_0, W_1, and W_2 are defined by (4) and U_0, U_1 and U_2 defined by

$$U_n = \int_0^1 s^n w(s)^2 \mathrm{d}s$$

Remark 3. Observe that if $w(t) = 1$ for all t, the constants in (15) reduce to $c_1 = 1$, $c_2 = 4$, $c_3 = 3$, $c_4 = 6$ and $c_0 = 12$ and the variance matrix reduces to the variance matrix obtained in the equally weighted case by, for example, Walker (1971).

4. EXAMPLES

The relation between the frequencies $\omega_{1,0},\ldots, \omega_{k,0}$, called the *partial frequencies*, and the frequency related to the pitch of the note heard when listening to the signal represented by y_t, call it λ, is of interest in musical sound signal analysis. With the estimation techniques used in sound analysis, it is common to have the smallest of the estimates of the ω relatively close to λ and the kth smallest estimate relatively close to $k\lambda$ (Rodet, 1997). As mentioned, in current sound analysis research, it is common to present these estimates without indications of their uncertainties. The fact that the smallest of the estimated ω is usually different from λ is usually interpreted as meaning that the signal is somewhat 'out of tune'. Under the assumption that the signals contain a stochastic element, the possibility exists that such variations are 'due to chance'. To explore this possibility, we redefine the frequencies in model (2) so that $\omega_{1,0} < \cdots < \omega_{k,0}$ and add the constraint

$$\hat{\omega}_{1,T} < \cdots < \hat{\omega}_{k,T} \tag{16}$$

to the estimation technique. See Remark 5 in the Appendix for a brief discussion of why, for large enough T, we can still use the asymptotic results of Section 3 when the estimation technique includes constraint (16). The asymptotic variances of the estimates resulting from this model provides a way to obtain approximate standard errors and confidence intervals for our estimates. The following examples demonstrate some applications.

A recording of a professional clarinet player playing (or trying to play) concert pitch A ($\lambda = 441$ Hz) was made. During the recording, 44100 observations of the air pressure wave were recorded per second. A one-second segment of the signal was divided into 45 non-overlapping, contiguous time frames, each with 1025 observations (approximately 23 milliseconds). The data in the first time-frame is shown in Figure 1 on page 000. Using the Splus function nls ( ), weighted least squares estimates were found for model (2), with $K = 15$, for the observed data in each of the above mentioned time frames. The choice of K and the length of time frames was made to obtain 'reasonable' fits. For example, for the first time frame, the residual mean-square is $\hat{\sigma}^2 = 0.0000134$. Comparing this to the variance of the original signal $(1/T)$ $\Sigma_t y_t^2 = 0.7609363$ shows that the fitted model explains a large amount of the variation of the original signal. Similar results where obtained for the other time-frames. For a discussion of goodness of fit and other selection procedures, see Irizarry (1998).

For many orchestral instruments, such as the clarinet, physical modelling (Fletcher and Rossing, 1991) suggests that, within short segments, the partial frequencies are harmonically related, meaning that model (2) holds, with the constraint

$$\omega_{k,0} = k\lambda \quad k = 1, \ldots, K \tag{17}$$

TABLE 1

WEIGHTED LEAST SQUARES FREQUENCY ESTIMATE FOR THE FIRST 5 FRAMES OF A CLARINET SIGNAL

Parameter	$\omega_{1,0}$	$\omega_{2,0}$	$\omega_{3,0}$	$\omega_{4,0}$	$\omega_{5,0}$	$\omega_{6,0}$	$\omega_{7,0}$	$\omega_{8,0}$	$\omega_{9,0}$	$\omega_{10,0}$	$\omega_{11,0}$	$\omega_{12,0}$	$\omega_{13,0}$	$\omega_{14,0}$	$\omega_{15,0}$
Frame 1 estimate	441.46	882.99	1324.0	1766.2	2208.2	2648.2	3090.1	3533.1	3973.0	4416.9	4855.3	5299.8	5741.6	6181.5	6621.3
Frame 2 estimate	441.51	883.07	1324.3	1765.6	2207.2	2649.6	3091.8	3532.5	3974.9	4416.5	4857.9	5300.0	5741.2	6181.4	6623.1
Frame 3 estimate	441.54	883.08	1324.5	1766.2	2208.1	2649.3	3089.7	3532.5	3972.6	4419.2	4852.3	5297.1	5736.2	6181.7	6620.9
Frame 4 estimate	441.47	883.06	1324.8	1765.3	2206.6	2649.2	3090.2	3532.8	3972.2	4415.8	4856.8	5299.0	5740.1	6181.0	6622.4
Frame 5 estimate	441.49	882.69	1324.6	1765.4	2206.4	2649.5	3092.7	3538.0	3974.7	4416.3	4854.5	5300.0	5739.7	6183.3	6625.4

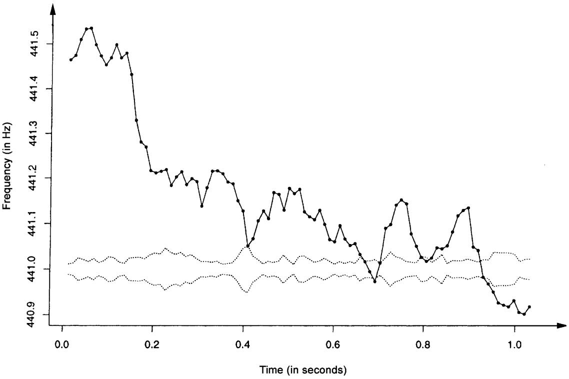

FIGURE 2. Pitch estimate for clarinet sound and confidence intervals around 441 Hz.

with λ, the frequency related to the note being played. Under this null hypothesis, with T large enough (we have $T = 1025$), we expect the weighted least squares estimate of $\omega_{k,0}$ to be 'close' to $k\lambda$ for $k = 1,\ldots, K$. Table 1 presents the resulting estimates $\hat{\omega}_{k,T}, k = 1, \ldots, 15$ for the first 5 frames. In current sound analysis research, constraint (17) is not necessarily imposed when performing estimation (Serra and Smith, 1991). The results shown on Table 1 seem to suggest that maybe constraint (17) should be considered. Our estimates and asymptotic results are a useful tool for exploring these results.

To examine the possibility that the fundamental frequency played by the instrumentalist is related to the note we hear, we study how the estimates of $\omega_{1,0}$ deviates from 441 Hz, the frequency related to concert pitch A, for each time frame. Figure 2 shows the estimate $\hat{\omega}_{1,T}$ obtained in the different time frames (corresponding to the dots in Figure 2). Theorem 2 provides the asymptotic variance of this weighted least squares estimate, which we can use as an approximation

$$\hat{\text{var}}(\hat{\omega}_{1,T}) \approx \frac{4\pi c_0 \hat{f}_{\epsilon\epsilon}(\hat{\omega}_{1,T})}{T^3(\hat{A}^2_{1,T} + \hat{B}^2_{1,T})}$$

with T, the number of observations used to perform the estimation, c_0 defined by (15), $\hat{A}_{1,T}$ and $\hat{B}_{1,T}$, the weighted least squares estimates of $A_{1,0}$ and $B_{1,0}$, and $\hat{f}_{\epsilon\epsilon}(\omega)$, a suitable estimate of the spectrum of the noise process $\{\epsilon_t\}$: see Quinn and Thomson (1991) for some examples of how to obtain this estimates. We can use this approximation, to construct marginal ± 2 s.e. limits, which we include in Figure 2. The figure suggests that, in this signal, the fundamental frequency is varying from 441 Hz for the different time-frames. We could say that, for most of the signal, the clarinet player is *statistically significantly out of tune*. But why do we hear 441 Hz? Studies show that the human ear cannot distinguish notes that are 0.03 semitones away from each other (Pierce, 1992). This implies that frequencies between 441 $\pm$ 0.76 Hz will sound the same. In

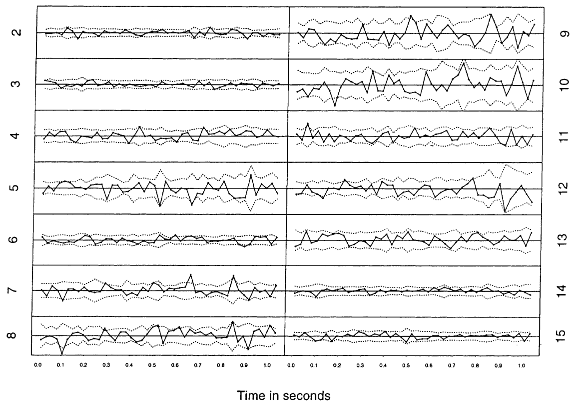

FIGURE 3. Difference between partial frequencies and respective multiples of fundamental frequency estimates for the clarinet sound with confidence intervals around 0. The range of the y-axes is -0.03–0.03 semitones in every plot.

fact, we can see that for all time-frames $\hat{\omega}_{1,T} \pm 2\sqrt{\hat{\text{var}}(\hat{\omega}_{1,T})}$ is inside 441 ± 0.76 Hz. Maybe the actual note being played *is* fluctuating between these values.

To see what the data have to say about (17), the harmonic relation suggested by physical theory, we can look at the differences $\hat{\omega}_{k,T} - k\hat{\omega}_{1,T}$ for $k = 2, \ldots, K$. In Figure 3, we plot these differences. Notice that they are, in general, not exactly equal to 0. Of course, this could simply be random variation. In current sound analysis research, this is taken as evidence that for each time frame, partials are not exactly multiples of the fundamental frequency. Again, Theorem 2 provides the asymptotic variance of weighted least squares estimates which we can use to compute approximations of the standard errors for the differences in question:

$$\hat{\text{var}}(\hat{\omega}_{k,T} - k\hat{\omega}_{1,T}) \approx 4\pi c_0 T^{-3}\left(\frac{\hat{f}_{\epsilon\epsilon}(\hat{\omega}_{k,T})}{\hat{A}^2_{k,T} + \hat{B}^2_{k,T}} + k^2\frac{\hat{f}_{\epsilon\epsilon}(\hat{\omega}_{1,T})}{\hat{A}^2_{1,T} + \hat{B}^2_{1,T}}\right) \qquad k = 1, \ldots, K$$

Using this, we construct marginal ± 2 s.e. limits about 0 and include them in Figure 3. In the present case, there seems to be no evidence that the partial frequencies are different from the respective multiples of the fundamental frequency. Notice that we are not presenting this as a formal hypothesis test, but rather as a useful exploratory method. For an example of how formal hypothesis testing can be performed, see Quinn and Thomson (1991).

For the fine example, a recording of a guitar playing D (146.8 Hz) was made. For plucked string instruments, like the guitar, physical models predict that partial frequencies will be higher than multiples of the fundamental frequency. In Fletcher and Rossing (1991) the ratio is predicted to be proportional to the partial number squared,

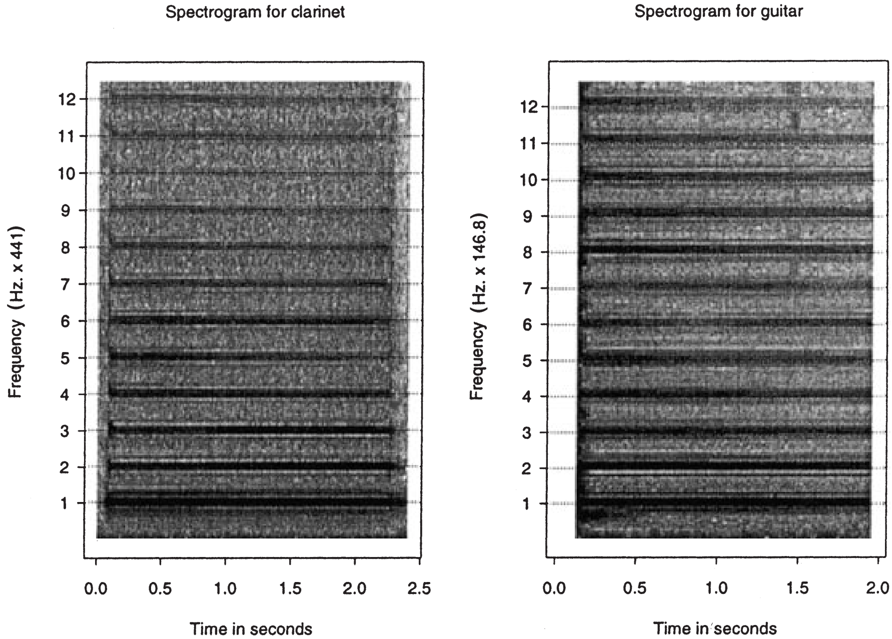

FIGURE 4. Spectrograms (sonograms) of clarinet playing A note (441 Hz) and guitar playing D note (146.8 Hz).

$$\frac{\omega_{k,0}}{\omega_{1,0}} \sim bk^2, \qquad k = 2, \ldots, K \tag{18}$$

Here b is a constant determined by the physical properties of the strings. In Figure 4, spectrograms of the guitar and clarinet sounds are compared. Dark shades of grey represent large values. For the clarinet, the fact that dark horizontal bands are centered at frequencies that are multiples of 441 Hz (denoted with dotted lines) is in agreement with our assertion that the partial frequencies are harmonically related. For the guitar, notice that dark horizontal bands are centered a bit higher than the multiples of 146.8 (this is most noticeable for the higher partials). The differences between the partial frequencies suggested by the spectrogram and the multiples of the 146.8 Hz frequency are quite small. In general, these differences are undetectable to even a 'trained ear'. However, it is of interest to assess the harmonic relationship proposed by physical models using data. The methodology described in this paper provides a way of doing this.

A two-second segment of the guitar signal was divided into 60 non-overlapping, contiguous time frames with 3000 observations each (approximately 68 milliseconds). As done for the clarinet, we find the weighted least squares estimates for model (2), with $K = 12$, for each of these time frames. In Figure 5, the differences $\hat{\omega}_{k,T} - k\hat{\omega}_{1,T}$ are shown for one of the time frames with marginal ± 2 s.e. limits around 0. All values are outside the ± 2 s.e. limits, suggesting that

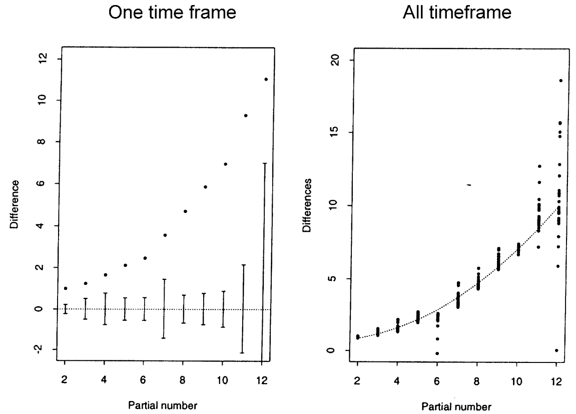

FIGURE 5. Differences between partial frequencies and respective multiples of fundamental frequency estimates for a guitar sound.

the differences are not 0. Figure 5 also shows these differences for all time frames. A parabola is fitted to these values and is also shown. These results provide evidence that the guitar produces partials that are not harmonically related and seem to be in agreement with a relationship like that of equation (18).

The asymptotic variance expressions can also be used to construct confidence intervals for the amplitude estimates $\hat{\rho}_{k,T} = (\hat{A}^2_{k,T} + \hat{B}^2_{k,T})^{\frac{1}{2}}$, $k = 1, \ldots, K$. We use the delta method to arrive at

$$\hat{\mathrm{var}}(\hat{\rho}_{k,T}) \approx 4\pi c_1 \frac{\hat{f}_{\epsilon\epsilon}(\hat{\omega}_{k,T})}{T} \qquad k = 1, \ldots, K$$

5. DISCUSSION

In this paper, we have presented an expression for the asymptotic variance of the weighted least squares estimates in a harmonic regression signal plus noise model. Useful applications in sound signal analysis were found for these results. In particular, we have examined the possibility that variations of estimates in different parts of the signal are due to chance alone. We presented evidence suggesting that the fundamental frequency of a clarinet sound departs from a fixed frequency. No evidence was found to contradict the fact that the clarinet is a harmonic instrument, with partial frequencies related to multiples of a fundamental frequency. This suggests that for the clarinet, and possibly other harmonic instruments, the additive synthesis model (1) might be improved by the constraint $\omega_{\mathrm{k}} = k\lambda$. In the case of a guitar sound, we found evidence suggesting it does not follow the same harmonic relation as the clarinet. Furthermore, the results seem to suggest that a constraint of the form $\omega_k = (\alpha + \gamma k)^2 \lambda$ may be appropriate. Further investigation of guitar sounds are of interest.

ACKNOWLEDGEMENTS

This paper is part of the author's doctoral dissertation which was written at the University of California, Berkeley, under the supervision of Professor David R. Brillinger. Professor Brillinger's suggestions and encouragement was an essential part of this work. The author would like to thank Professor David Wessel and Adrian Freed for their insights on sound analysis and Scott Zeger, Rick Schoenberg, and the referee for some useful comments. Partial support for this work was provided by the National Science Foundation Graduate Research Fellowship, the University of California's Chancellor's Predoctoral Opportunity Fellowship, and NSF grant DMS 96–25774.

APPENDIX

PROOF OF LEMMA 1. Fix n. To prove (5), notice that, for $\lambda = 0, 2\pi$, we have, from the boundedness and bounded variation of $w(s)$ that

$$\begin{aligned}\lim_{T\to\infty} T^{-(n+1)}\Delta_n^T(\lambda) &= \lim_{T\to\infty}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^n w\left(\frac{t}{T}\right)\left(\frac{1}{T}\right)\\ &= \int_0^1 u^n w(u)\mathrm{d}u\\ &= W_n\end{aligned}$$

To prove (6) let $0 < \lambda < 2\pi$ and define

$$\Delta^t(\lambda) = \sum_{s=1}^{t}\exp(\mathrm{i}\lambda s)$$

with the convention that $\Delta^0(\lambda) = 0$. Letting $h(u) = u^n w(u)$ and using summation by parts gives

$$\Delta_n^T(\lambda) = T^n\left[h(1)\Delta^T(\lambda) + \sum_{t=1}^{T-1}\left\{h\left(\frac{t}{T}\right) - h\left(\frac{t+1}{T}\right)\right\}\Delta^t(\lambda)\right]$$

Notice that, if $w(t)$ is bounded and has bounded variation on [0, 1], so does $h(s)$. Let M be $\sup_s |h(s)|$ and V be the total variation of $h(s)$:

$$V = \sup \sum_{i=1}^{k} |h(s_i) - h(s_{i-1})|$$

with the suprema taken over all possible subdivision $0 < s_0 < s_1 < \cdots < s_k < 1$ of [0, 1].

Then, we have

$$\left|\Delta_n^T(\lambda)\right| \leq T^n \left[M|\Delta^T(\lambda)| + V \max_{1 \leq t \leq T} |\Delta^t(\lambda)| \right]$$

We know – see, for example, Brillinger (1981) – that $|\Delta^t(\lambda)| \leq L = 1/|\sin(\frac{1}{2}\lambda)|$ for all t. Notice that L depends on λ, but given $0 < \lambda < 2\pi$ it is constant for all t, thus

$$\left|\Delta_n^T(\lambda)\right| \leq T^n L(M + V)$$

and this completes the proof of the Lemma.

Proof of theorem 1 Consider first the one sinusoidal case ($K = 1$). We start by proving

$$\lim_{T \to \infty} T|\hat{\omega}_T - \omega_0| = 0 \text{ in probability} \tag{19}$$

which is stronger than the usual consistency

$$|\hat{\omega}_T - \omega_0| = 0 \text{ in probability}$$

but is needed to prove the consistency of the remaining two estimates and asymptotic normality.

Letting $D_0 = \frac{1}{2}(A_0 - \mathrm{i}B_0)$ gives

$$\begin{aligned} q_T(\omega) = {} & \left|T^{-1}d^T(\omega)\right|^2 + \left|T^{-1}\{D_0\Delta_0^T(\omega_0+\omega) + \bar{D}_0\Delta_0^T(\omega_0-\omega)\}\right|^2 \\ & + 2\mathscr{R}\left(\left[T^{-1}d^T(\omega)\right]\left[T^{-1}\{D_0\Delta_0^T(\omega_0+\omega) + \bar{D}_0\Delta_0^T(\omega_0-\omega)\}\right]\right) \end{aligned} \tag{20}$$

with

$$d^T(\omega) = \sum_{t=1}^{T} w\left(\frac{t}{T}\right)\epsilon_t \exp(-\mathrm{i}\omega t)$$

By Lemma 1, for $0 < \omega < \pi$,

$$T^{-1}\Delta_1^T(\omega_0 + \omega) = o(1)$$

and

$$T^{-1}\Delta_1^T(\omega_0 - \omega) = \begin{cases} W_0 & \omega = \omega_0 \\ o(1) & \text{otherwise} \end{cases}$$

Lemma 2 implies that for $0 < \omega < \pi$, $T^{-1}d^T(\omega) = o_p(1)$. Thus we have that

$$q_T(\omega) = \frac{1}{4}\rho_0^2 \left|T^{-1}\Delta_0^T(\omega - \omega_0)\right|^2 + o_p(1)$$

and

$$q_T(\omega_0) = \frac{1}{4}\rho_0^2 W_0^2 + o_p(1)$$

To prove (19), for any $b > 0$, define

$$P_T(b) = \{\omega : T|\omega - \omega_0| \geq b\} \tag{21}$$

Notice that

$$\begin{aligned}\Pr(T|\hat{\omega}_T - \omega_0| \geq b) &\leq \Pr\left(\sup_{\omega \in P_T(b)} q_T(\omega) \geq q_T(\omega_0)\right) \\ &= \Pr\left(\sup_{\omega \in P_T(b)} \left|T^{-1}\Delta_0^T(\omega - \omega_0)\right| \geq W_0 + o_p(1)\right)\end{aligned}$$

Noticing that the expression on the left is a Reimman sum (Irizarry, 1998), we have

$$\sup_{\omega \in P_T(b)} \left|T^{-1}\Delta_0^T(\omega - \omega_0)\right| = \left|\int_0^1 w(s)\exp\{\mathrm{i}T(\omega - \omega_0)s\}\mathrm{d}s\right| + o(1)$$

Let ω^* be such that

$$\left|\int_0^1 w(s)\exp\{\mathrm{i}T(\omega^* - \omega_0)s\}\mathrm{d}s\right| = \sup_{\omega \in P_T(b)} \left|\int_0^1 w(s)\exp\{\mathrm{i}T(\omega - \omega_0)s\}\mathrm{d}s\right|$$

Let $b^* = T|\omega^* - \omega_0| \geq b > 0$. Then, by the definition of $P_T(b)$ given by (21), we have

$$\lim_{T\to\infty} \Pr(T|\hat{\omega}_T - \omega_0| \geq b) \leq \lim_{T\to\infty} \Pr\left(\left|\int_0^1 w(s)\exp(\mathrm{i}b^*s)\mathrm{d}s\right| + o(1) \geq W_0 + o_p(1)\right)$$

Since $W_0 > 0$ is a deterministic constant and $b^* > 0$

$$\begin{aligned}W_0 &= \left|\int_0^1 w(s)\mathrm{d}s\right| \\ &= \int_0^1 |w(s)\exp(\mathrm{i}b^*s)|\mathrm{d}s \\ &> \left|\int_0^1 w(s)\exp(\mathrm{i}b^*s)\mathrm{d}s\right|\end{aligned}$$

and we have (19).

To prove consistency for $\hat{A}_T$ and $\hat{B}_T$, let

$$r(t, \beta) = \{D_0 \exp(\mathrm{i}\omega_0 t) + \bar{D}_0 \exp(-\mathrm{i}\omega_0 t)\}$$

and

$$L = 2\left\{\sum_{t=1}^{T} w\left(\frac{t}{T}\right)\right\}^{-1}$$

By Lemma 1 and the mean value theorem, for some $\tilde{\omega}_T$ satisfying $|\tilde{\omega}_T - \omega_0| \leq |\hat{\omega}_T - \omega_0|$

$$|\hat{A}_T - A_0 + \mathrm{i}(\hat{B}_T - B_0)| = +\left|L\sum_{t=1}^{T} w\left(\frac{t}{T}\right) r(t;\beta)\mathrm{i}t\exp(\mathrm{i}\tilde{\omega}_T t)(\hat{\omega}_T - \omega_0)\right| + o_p(1)$$

The first term in the right-hand side of the above equation is smaller than

$$L\sum_{t=1}^{T} w\left(\frac{t}{T}\right)|r(t;\beta)|t|\hat{\omega}_T - \omega_0| \leq \rho_0 T|\hat{\omega}_T - \omega_0| = o_p(1)$$

And thus $|(\hat{A}_T - A_0) + \mathrm{i}(\hat{B}_T - B_0)| = o_p(1)$, and because both the real and imaginary parts converge in probability to 0, consistency for the one sinusoidal case is proven. The general case follows in the same way. See Irizzary (1998) for details.

REMARK 4. Notice that $q_T(\omega)$ is symmetric in its K arguments. As noticed by Walker (1971), a way of determining which component of ω is associated with a particular partial has to be found. Walker (1971) notices that if we determine the $\hat{\omega}_{k,T}$ as the kth largest local maxima of $q_T(\omega)$ subject to separation condition (13), these will, for sufficiently large T, almost certainly estimate the frequencies of the harmonic components arranged in descending order of magnitude. The same argument works in our situation.

REMARK 5. In some applications, it is convenient to have the parameters in ascending order. Define $\omega^*_{1,0} < \cdots < \omega^*_{K,0}$ so that $\omega^*_{k,0} = k$th smallest of $\omega_{1,0}, \ldots,$ $\omega_{K,0}$ *and define the* $A^*_{k,0}$ *and* $B^*_{k,0}$ *accordingly. Under this 're-parameterization', instead of using the technique described in Remark 4, we do the following:*

Notice that the proof of Theorem 1 shows that for all $\omega_{k,0}$, $k = 1, \ldots, K$, for any $\delta > 0$ there is a T_0 such that, for all $T \geq T_0$, one of the K components of $\hat{\boldsymbol{\omega}}_T$, call it $\hat{\omega}_{k,T}$, satisfies

$$\Pr(T|\hat{\omega}_{k,T} - \hat{\omega}_{k,0}| > b) < \delta \tag{22}$$

for all $b > 0$. For each T, define the estimate $\hat{\omega}^*_{k,T}$ as the kth smallest of $\hat{\omega}_{1,T}, \ldots, \omega_{\hat{K},T}$. Because K is finite, we have that for any $\delta' > 0$, we can find an appropriate $\delta(K, \delta')$ in (22) so that for $T \geq T_0$ we have

$$\Pr\left(T\left|\{\min_{1\le k\le K}\hat{\omega}_{k,T}\}-\omega_{1,0}^*\right|>b\right)<\delta'$$

for all $b > 0$. Similarly, we find that $\lim_{T\to\infty} T|\hat{\omega}_{k,T}^* - \omega_{k,0}^*| = 0$.

PROOF OF THEOREM 2. We follow in a similar way to Walker (1971) and Hannan (1973). First, consider the case of one sinusoidal component ($K=1$). Using Theorem 4.4.2 in Brillinger (1981, p. 95), we have that the vector $\mathbf{u}$, with components

$$\begin{aligned}
u_1 &= T^{-\frac{1}{2}}\sum w\left(\frac{t}{T}\right)\epsilon_t\cos\omega_0 t\\
u_2 &= T^{-\frac{1}{2}}\sum w\left(\frac{t}{T}\right)\epsilon_t\sin\omega_0 t\\
u_3 &= T^{-\frac{3}{2}}\sum w\left(\frac{t}{T}\right)\epsilon_t t\cos\omega_0 t\\
u_4 &= T^{-\frac{3}{2}}\sum w\left(\frac{t}{T}\right)\epsilon_t t\sin\omega_0 t
\end{aligned} \tag{23}$$

is asymptotically multivariate normal with zero mean and variance matrix

$$\mathbf{U} = \pi f_{\epsilon\epsilon}(\omega_0)\begin{pmatrix} U_0 & 0 & U_1 & 0\\ 0 & U_0 & 0 & U_1\\ U_1 & 0 & U_2 & 0\\ 0 & U_1 & 0 & U_2\end{pmatrix}$$

Expanding $q_T'(\omega)$ in the first two terms of its Taylor series, about ω_0 we can write

$$T^{-\frac{1}{2}}q_T'(\omega_0) = -T^{\frac{3}{2}}(\hat{\omega}_T-\omega_0)T^{-2}q_T''(\tilde{\omega}_T) \qquad |\tilde{\omega}_T-\omega_0|\le|\hat{\omega}_T-\omega_0| \tag{24}$$

Notice that, calculating the derivative and repeated use of Lemmas 1 and 2, we can show that

$$T^{-\frac{1}{2}}q_T'(\omega_0) = -W_1B_0u_1 + W_1A_0u_2 + W_0B_0u_3 - W_0A_0u_4 + o_p(1) \tag{25}$$

Since $T|\tilde{\omega}_T-\omega_0|$ converges to zero in probability, taking the second derivative, we have by repeated use of Lemmas 1 and 2

$$T^{-2}q_T''(\tilde{\omega}_T) = \tfrac{1}{2}(A_0^2+B_0^2)(W_1^2-W_0W_2)+o_p(1) \tag{26}$$

Using (24), (25) and (26), we can express the vector of standardized estimates as a linear combination of the vector $\mathbf{u}$, defined by equation (23), plus a quantity converging to 0 in probability.

$$\{T^{\frac{1}{2}}(\hat{A}_T-A_0), T^{\frac{1}{2}}(\hat{B}_T-B_0), T^{\frac{3}{2}}(\hat{\omega}_T-\omega_0)\}' = \mathbf{Au}+\mathbf{o_p(1)}$$

with

$$\mathbf{A} = \begin{pmatrix} B_0^2 W_2 + A_0^2 (W_2 - \frac{W_1^2}{W_0}) & -\frac{A_0 B_0 W_1^2}{W_0} & -B_0^2 W_1 & A_0 B_0 W_1 \\ -\frac{A_0 B_0 W_1^2}{W_0} & A_0^2 W_2 + B_0^2 \left(W_2 - \frac{W_1^2}{W_0} \right) & A_0 B_0 W_1 & -A_0^2 W_1 \\ -B_0 W_1 & A_0 W_1 & B_0 W_0 & -A_0 W_0 \end{pmatrix}$$

By Assumption 2, we know that all the denominators in the components of **A** are not 0. This implies that **Au** is asymptotically multivariate normal with variance matrix **AUA′**. By computing **AUA′**, we obtain the variance expression (14). This proves Theorem 2 for the one sinusoidal case.

Taking derivatives of $q_T(\omega)$, we notice the $\partial q_T(\omega)/\partial \omega_k$ do not depend on ω_l when $l \neq k$. Furthermore, under condition (13), the $\hat{\omega}_{k,T}$ are asymptotically independent; see, for example, Brillinger (1981). Theorem 2 now follows for the general case.

REFERENCES

Brillinger, D. R. (1997) Fitting cosines: Some procedures and some physical examples. In *Applied Probability*, *Stochastic Processes*, *and Sampling Theory* (eds I. B. MacNeil and G. J. Umphrey). Boston, MA: Reidel, pp. 75–100.

—(1981) *Time Series Data Analysis and Theory* (expanded edn). San Francisco: Holden–Day.

—(1986) Regression for randomly sampled spatial series: The trignometric case. *Essays in Time Series and Allied Processes – Papers in Honor of E. J. Hannan, Journal of Applied Probability* (23A), 275–89.

—and Irizarry, R. A. (1998) An investigation of the second- and higher-order spectra of music. *Signal Processing* 65, 161–79.

Brown, E. N. (1990) A note on the asymptotic distribution of the parameter estimates for the harmonic regression model. *Biometrika* 77 (3), 653–6.

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

Grey, J. (1977) Multidimensional perceptual scaling of musical timbre. *Journal of the Acoustical Society of America* 62, 1270–7.

Hannan, E. J. (1973) The estimation of frequency. *Journal of Applied Probability* 10, 510–19.

Hassan, T. (1982) Nonlinear time series regression for a class of amplitude modulated cosinusoids. *Journal of Time Series Analysis* 3, 109–22.

Irizarry, R. A. (1998) *Statistics and music: Fitting a local harmonic model to musical sound signals*. PhD thesis, University of California, Berkeley.

Pierce, J. R. (1992) *The Science of Musical Sound*. New York: Freeman.

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

Risset, J. C. and Mathews, M. V. (1969) Analysis of musical-instrument tones. *Physics Today* 22 (2), 23–30.

Rodet, X. (1997) Musical sound signals analysis/synthesis: Sinusoidal + residual and elementary waveform models. In *Proceedings of the IEEE Time-Frequency and Time-Scale Workshop* (*TFTS'97*). Coventry, UK: IEEE.

Serra, X. and Smith, J. O. (1991) Spectral modeling synthesis: A sound analysis/synthesis system based on deterministic plus stochastic decomposition. *Computer Music Journal* 14 (4), 12–24.

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