On frequency estimation

BY JOHN A. RICE AND MURRAY ROSENBLATT

Department of Mathematics, University of California, San Diego, La Jolla, California 92093, U.S.A.

SUMMARY

This paper discusses a least-squares procedure and the use of the periodogram for isolating a discrete harmonic of a time series. It is shown that the usual asymptotics on estimation of frequency, amplitude and phase of such a harmonic have to be used with great caution from a moderate sample perspective. Computational issues are discussed and some illustrations are provided. Bolt & Brillinger (1979) make use of these asymptotic results.

Some key words: Frequency estimation; Harmonic analysis; Periodogram; Time series.

1. INTRODUCTION

We consider a time series model of the form \( X_t = a \cos(\omega t + \phi) + \epsilon_t \), where \( \epsilon_t \) is a stationary noise sequence, and one is interested in estimating the amplitude, frequency and phase of the harmonic component. The asymptotic theory of the least-squares estimates of these parameters has a long history. Whittle (1951, 1953) obtained some of the earliest results. More recent results are by Hasan (1982), Hannan (1973) and Walker (1971), who formalize and extend Whittle's results. In these works it is shown that the asymptotic variance of the frequency estimate is of order \( n^{-3} \) and that the asymptotic variances of the other two components are of the more usual order \( n^{-1} \). These results extend when there are several harmonic components.

The rate for the estimate of \( \omega \) seems almost unbelievably good, and our work was motivated by a desire to see how reliable the asymptotic theory is. In brief, we find that the product of the amplitude and the sample size, \( n \), must be quite large in order for the asymptotic theory to be meaningful. If this product is not large, the frequency estimate is much more variable than indicated by the asymptotic theory and the amplitude estimate is severely biased. In applications in which the amplitude is small, giving rise to a small peak in the periodogram, these results suggest that naive application of the asymptotic theory to gauge resolution can be quite misleading. Section 2 of this paper is devoted to a review and examination of the asymptotic theory.

We are also concerned with computational issues arising from the least-squares problem. This problem is nonlinear in the parameters, so that some sort of iterative search must be employed. Typically, search methods start from an initial guess and then proceed by a sequence of modified Newton-Raphson steps. For this nonlinear least-squares problem, it turns out that there are many local minima with a separation in frequency about \( n^{-1} \) which makes the stationary point to which the iterative scheme converges extremely sensitive to the starting values, and this problem gets worse as the sample size increases. Furthermore, it follows from the results of § 2 that the estimate of the amplitude is very biased unless the frequency is resolved with order \( o(n^{-1}) \) so that failure to converge to the global minimum may give a very poor estimate of amplitude. The problem becomes
especially severe if one is estimating the parameters of several harmonic components simultaneously since then the iteration is taking place in a high-dimensional space with many local minima. Walker (1971) suggests finding the maximum of the periodogram and then computing the amplitude and phase estimates. However, it turns out that this approach has some drawbacks. A bias can arise for moderate sample sizes that is appreciable compared to the standard deviation suggested by asymptotic theory. These issues are further discussed in § 3. There we also report the results of some modest simulations which support the qualms about asymptotic theory raised in § 2.

The bulk of the paper deals with a general context where the procedures are suggested by the likelihood function computed for Gaussian residuals. A discussion of the likelihood function with residuals a Gaussian stationary sequence is given in the Appendix.

2. ASYMPTOTIC THEORY

We initially assume that

$$X_t = A_0 \cos(\omega_0 t) + B_0 \sin(\omega_0 t) + \epsilon_t,$$

where $\epsilon_t$ is a strictly stationary sequence with a continuous spectral density. If $\epsilon_t$ is Gaussian, the log likelihood of $X_1, \ldots, X_n$ is

$$-\frac{1}{2} \log |R_n| - \frac{1}{2} \sum_{t=1}^n \{X_t - A_0 \cos(\omega_0 t) - B_0 \sin(\omega_0 t)\} r_{t,\tau}^{-1}\{X_{\tau} - A_0 \cos(\omega_0 \tau) - B_0 \sin(\omega_0 \tau)\},$$

where $R_n$ is the covariance matrix of the $X$'s and $r_{t,\tau}^{-1}$ is the $(t, \tau)$ element of the inverse of the Toeplitz matrix $R_n$. Under a number of assumptions on the spectral density $f(\lambda)$ of the residuals $\epsilon_t$, one can give a heuristic argument showing that the part of the log likelihood above depending on $\omega_0$, after fitting $A_0$ and $B_0$, can be approximated by a constant multiplied by

$$\frac{1}{f(\omega_0)} \frac{1}{n} \left| \sum_{t=1}^n X_t e^{i\omega_0 t} \right|^2.$$

We shall give a brief heuristic discussion of this in the Appendix. In the main body of the paper we shall restrict ourselves to the case in which the $\epsilon_t$ are independent and identically distributed $N(0, \nu)$. In this case the log likelihood is

$$-\frac{1}{2} n \log (2\pi\nu) - \frac{1}{2} S_n(A, B, \omega)/\nu,$$

with

$$S_n(A, B, \omega) = \sum_{t=1}^n \{X_t - A \cos(\omega t) - B \sin(\omega t)\}^2.$$

The maximum likelihood estimates of $A, B, \omega$ are determined by minimizing the sum of squares $S_n$. The maximum likelihood estimate of $\nu$ is given by the minimum sum of squares divided by $n$.

Suppose that one replaces $S_n(A, B, \omega)$ by

$$U_n(A, B, \omega) = \sum_{t=1}^n X_t^2 - 2 \sum_{t=1}^n X_t \{A \cos(\omega t) + B \sin(\omega t)\} + \frac{1}{2} n (A^2 + B^2).$$

The difference between $S_n(A, B, \omega)$ and $U_n(A, B, \omega)$ is $O(1)$ as $n \to \infty$ if $\omega \neq 0$, $\pi$ and this suggests that if $U_n$ is minimized instead of $S_n$ the difference will be small for large
On frequency estimation

479

n if the true value \( \omega_0 \) of \( \omega \) is not zero or \( \pi \). The estimates \( A, B, \omega \) obtained by minimizing \( U_n \) instead of \( S_n \) are simpler and are given by

\[
\hat{A}_n = \frac{2}{n} \sum_{i=1}^{n} X_i \cos(\hat{\omega} t), \quad \hat{B}_n = \frac{2}{n} \sum_{i=1}^{n} X_i \sin(\hat{\omega} t),
\]

with \( \hat{\omega}_n \) such that

\[
I_n(\hat{\omega}_n) = \max_{0 < \omega < \pi} \{ I_n(\omega) \}.
\]

Following the notation of Walker (1971), the periodogram \( I_n(\omega) \) is given by

\[
\frac{1}{2} n I_n(\omega) = \left| \sum_{i=1}^{n} e^{i\omega t} [A_0 \cos(\omega_0 t) + B_0 \sin(\omega_0 t) + \epsilon_i] \right|^2.
\]

At first the analysis will be carried out for \( A_0, B_0 \) assumed constant. Later a corresponding computation will be considered where \( A_0 \) and \( B_0 \) may depend on the sample size \( n \). Walker (1971) considers estimating the unknown frequency \( \omega_0 \) first. In the asymptotic theory it is usual to assume that \( \omega_0 \) is estimated with an error \( o(n^{-1}) \). We shall finally see what happens to the estimates of the unknown parameters if \( \omega_0 \) is not estimated with this precision.

First consider \( A_0, B_0 \) to be independent of \( n \). Let \( D_0 = \frac{1}{2}(A_0 - iB_0) \) and

\[
M_n(u) = \sum_{i=1}^{n} e^{i\omega u} = \frac{\exp \{i(n+1)u\} \sin(nu)/\sin(u)}{n} \quad (0 < u < 2\pi),
\]

\[
(u = 0).
\]

Equation (1) can thus be rewritten

\[
\frac{1}{2} n I_n(\omega) = \left| \sum_{i=1}^{n} \epsilon_i e^{i\omega u} \right|^2 + 2 \Re \left[ \sum_{i=1}^{n} \epsilon_i e^{-i\omega u} \{ D_0 M_n(\omega + \omega_0) + \bar{D}_0 M_n(\omega - \omega_0) \} \right] + |D_0 M_n(\omega + \omega_0) + \bar{D}_0 M_n(\omega - \omega_0)|^2.
\]

Let \( 0 < \omega_0 < \pi \). Following Walker (1971) it can be shown that when \( \omega = \omega_0 \) expression (2) is given to the first order by

\[
|\bar{D}_0 M_n(0)|^2 = \frac{1}{4} n^2 (A_0^2 + B_0^2),
\]

which arises from the last term on the right of (2). It is clear that

\[
\max_{0 < \omega < \pi} \{ |M_n(\omega + \omega_0)| \} = O(1), \quad \max_{0 < \omega < \pi} \{ |M_n(\omega - \omega_0)| \} = n.
\]

Under assumptions of short range dependence on the sequence \( \{\epsilon_i\} \) (Brillinger, 1981; Turkman & Walker, 1984) one can show that

\[
\max_{0 < \omega < \pi} \left( \left| \sum_{i=1}^{n} \epsilon_i e^{i\omega u} \right|^2 \right) = O(n \log n)^{\frac{1}{2}}.
\]

It then follows that

\[
\max_{0 < \omega < \pi} \{ |I_n(\omega) - \frac{1}{n} n^{-1}(A_0^2 + B_0^2)| M_n(\omega - \omega_0)|^2 \} = O(n \log n)^{\frac{1}{2}}.
\]

The argument given by Walker (1971) shows that if \( \delta > 0 \) is chosen small enough so that \( \{\sin(\frac{1}{n}\delta)/\sin(\frac{1}{n}\delta)\}^2 > 1/\pi^2 \) then

\[
K(n, \delta) = \max_{|\omega - \omega_0| > n^{-1}\delta} \{ I_n(\omega) \}
\]

\[
\leq \frac{1}{2} n^{-1}(A_0^2 + B_0^2) \sin^2(\frac{1}{n}\delta)/\sin^2(\frac{1}{n}\delta) + O(n \log n)^{\frac{1}{2}},
\]
using estimate (3). Since
\[ \lim_{n \to \infty} \Pr \{ K(n, \delta) < I_n(\omega_0) \} = 1 \]
it follows that
\[ \lim_{n \to \infty} \Pr (n|\hat{\omega}_n - \omega_0| < \delta) = 1, \]
where \( \hat{\omega}_n \) is the location of the absolute maximum of \( I_n(\omega) \) on \( 0 \leq \omega \leq \pi \). This itself suggests that if one wishes effectively to estimate the maximum one would have to sample the periodogram deterministically by a grid mesh with the mesh length \( o(n^{-1}) \) as \( n \to \infty \). Because of the many local maxima referred to in \( \S \) 1, it is clear that the use of a nonlinear least-squares procedure implemented by a recursive local computation has great disadvantages.

Walker (1971) goes on to show that \( A_n, B_n, \omega_n \) are asymptotically normally distributed. This result is established by a linearization. It is part of the point of the present paper to note that this linearization only holds in a very narrow range.

To examine the situation in which the amplitude of the harmonic component is small, let us now consider the analysis again where \( A_0 \) and \( B_0 \) can depend on \( n \). Estimate (3) implies again that the first term on the right-hand side of (2) is \( O(n \log n) \). We shall estimate the remaining terms on the right-hand side of (2) over the range \( |\omega - \omega_0| > h \), where \( h \) decreases to zero as \( n \to \infty \). The second term on the right-hand side of (2) is of order of magnitude
\[ O(n \log n) \] and the last term is
\[ O(|D_0| \min (nh, 1)/h) \] These estimates imply that if \( |D_0|^2 \geq (\log n)/n \) asymptotically as \( n \to \infty \) one will be able to estimate \( \omega_0 \) with an error \( o(n^{-1}) \).

Suppose that one had obtained an estimate \( \omega_n \) of \( \omega_0 \) by searching along a grid mesh of mesh length \( h \) which was not \( o(n^{-1}) \) or that an iterative procedure had converged to a local maximum which was not within \( o(n^{-1}) \) away from \( \omega_0 \). One would still be tempted to use the estimates \( \hat{A}_n, \hat{B}_n \) of the coefficients \( A_0 \) and \( B_0 \)
\[ \hat{A}_n + i\hat{B}_n = \frac{2}{n} \sum_{i=1}^{n} X_i e^{i\omega t_i} = \frac{2}{n} \sum_{i=1}^{n} (D_0 e^{i\omega t_i} + D_0 e^{-i\omega t_i} + \epsilon_i) e^{i\omega t_i}. \]
Then
\[ \hat{A}_n - A_0 + i(\hat{B}_n - B_0) = \frac{2}{n} \left[ D_0 M_n(\omega_n + \omega_0) + D_0 (M_n(\omega_n - \omega_0) - n) + \sum_{i=1}^{n} \epsilon_i e^{i\omega t_i} \right]. \]
Assuming that \( \omega_n - \omega_0 = o(1) \), it then follows that
\[ \hat{A}_n - A_0 + i(\hat{B}_n - B_0) = \frac{2}{n} \left[ D_0 \{ M_n(\omega_n - \omega_0) - n \} + O(n^{-1} \log n) \right]. \]
If \( \omega_n - \omega_0 \sim cn^{-\beta} \), for \( 0 < \beta < 1 \), \( c \neq 0 \), then
\[ \hat{A}_n - A_0 + i(\hat{B}_n - B_0) = -2D_0 + O(n^{-1+\beta} + n^{-1} \log n) \]
On frequency estimation

481

and this suggests that instead of using the conjugate of $\hat{D}_n$, $\hat{D}_n^* = \hat{A}_n + i\hat{B}_n$, as the appropriate estimate one ought to use $-\hat{D}_n^*$. However, if $\omega_n - \omega_0 \sim cn^{-1}$, one would obtain

$$\hat{D}_n^* - \hat{D}_0 = 2\hat{D}_0 \exp \left( \frac{1}{2}ic \right) \sin \left( \frac{1}{2}c \right) - 1 + o(1).$$

Since one could not determine how close $\omega_n$ is to $\omega_0$ such relations are meaningless in obtaining an estimate of $\hat{D}_0$. We thus see that, if $\omega_n$ is not within $o(n^{-1})$ of $\omega_0$, the estimates of $A_0$ and $B_0$ are inconsistent. Note that the fast Fourier transform allows us only to compute the periodogram at the points $2\pi j/n$ which is $O(n^{-1})$ and not $o(n^{-1})$.

3. Computational considerations

In this section we consider some computational aspects and the results of some modest simulations for the simple model $x_t = a_0 \cos (\omega_0 t + \phi_0) + e_t$, where the noise sequence $e_t$ is Gaussian with mean 0 and variance 1. We will see that certain difficulties arise already in fitting this simple model.

As mentioned in the introduction, computation of the least-squares estimates is complicated by the presence of many local minima. Figure 1 illustrates this problem. It shows minus the residual sum of squares for one realization of the model above with $n = 100$, $a_0 = 1$, $\omega_0 = 0.5$, and $\phi_0 = 0$ as a function of the variables $a$, $\phi$, $\omega$, for $a = 1$, $-\pi \leq \phi \leq \pi$ and $0.3 \leq \omega \leq 0.7$. A Newton-type algorithm iterating in a three-dimensional space runs a risk of converging to a local extrema. The situation is not as bad as it could be since the ridges corresponding to local extrema in frequency make a small angle with the $\phi$ axis and since the estimate of $a$ is asymptotically uncorrelated with the estimates of frequency and phase. Nevertheless, it is clear that real problems can arise, especially if one were fitting two or more peaks simultaneously. The initial values provided to the search algorithm are thus critical; in particular initial values suggested by complex demodulation may be unsuitable because of the low pass filter employed in this process. We have found it necessary to use a large number of random starting points in our computations.

A direct search of the periodogram at a fine grid of points substantially finer than that given by the frequencies $2\pi j/n$ used by the fast Fourier transform is thus appealing, but unfortunately has its drawbacks as well. In the previous section it was shown that the contribution to the expected value of the periodogram by the harmonic component is proportional to $|M_n(\omega + \omega_0) + M_n(\omega - \omega_0)|^2$, and that for large $n$ this expression is dominated by the second factor. However, for small $n$, the first factor can make a nonnegligible
contribution. For example, for \( n = 100 \) and \( \omega_0 = 0.5 \), the maximum of this expression occurs at \( \omega = 0.5013 \) rather than at 0.5. According to our simulations, if \( a = 8 \), the bias of \( \hat{\omega} \) is consequently more than twice its standard error, and similarly for the estimate of phase. While this is noticeable, it is not as large as that caused by using a local rather than the global minimum of the sum of squares. As we will see below, the least-squares estimate does not appear to be similarly affected.

We now turn to a discussion of some simulations. For the simple model above we took \( n = 100 \) and varied \( a \) as 0.5, 1, 2, 4, 8. For each such model, 100 realizations were constructed, and the least-squares estimates were calculated by a modified Levenberg-Marquardt algorithm, \texttt{SNLSE} from \texttt{MINPACK}. Because of the many local minima alluded to above we found it necessary to use many starting points to locate the global minimum. Thus 50 starting values for the algorithm were generated on each run as \( a + a(U_1 - \frac{1}{2}), \omega + 8\pi(U_2 - \frac{1}{2})n, \) and \( 2\pi U_3 \), where the \( U_i \) were pseudo random uniform variates on \([0,1] \). The choice of 50 may have been unduly conservative, but the only penalty can have been in computer time. The proper choice of a range for the random starting points is more problematical. The fast Fourier transform computes the periodogram at frequencies which are multiples of \( 2\pi/n \), and such a periodogram would often be the starting point of an investigation of periodicities; we thus generate the starting values for \( \hat{\omega} \) in an interval of width \( 8\pi/n \) around \( \omega_0 \) assuming that a crude guess is obtained from the periodogram. We assume that some knowledge, but not much, about amplitude is obtained from the periodogram, and that there is no information about phase.

To compare the results to the asymptotic theory, we standardized the least-squares estimates by subtracting the true values and dividing by the asymptotic standard deviations. Table 1 gives the means and standard deviations of the standardized variables for each model. Table 1 has been supplemented by graphical analysis of the frequency distributions. Estimates of amplitude are positively biased and the bias becomes smaller with increasing amplitude.

| Table 1. Least-squares estimates of \( a, \omega, \phi \) minus true value divided by asymptotic standard error. Mean and standard deviation of resulting standardized variable |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| \( a = \frac{1}{2} \) | \( a = 1 \) | \( a = 2 \) | \( a = 4 \) | \( a = 8 \) |
| \( a \) | 0.94, 1.03 | 0.45, 1.30 | 0.31, 1.33 | 0.28, 1.34 | 0.20, 1.34 |
| \( \omega \) | -0.71, 6.98 | 0.02, 1.26 | 0.05, 1.25 | 0.06, 1.25 | 0.07, 1.25 |
| \( \phi \) | 0.05, 1.34 | -0.05, 1.31 | -0.08, 1.28 | -0.10, 1.28 | -0.10, 1.28 |

When \( a = \frac{1}{2} \) the frequency is not well resolved, and the asymptotics fail to give a reasonable indication of variability. Direct examination of periodograms shows that in this case the periodogram frequently has a local maximum near the true value, but larger maxima at other locations. In this case the contribution of the noise sequence to the expected value of the periodogram, the first term in (2), is 2 and the contribution from the harmonic component, the last term in (2), is \( \frac{1}{2} na^2 = 12.5 \). At \( \omega = 0.5 \) the contribution to the variance of the periodogram from the product of the noise sequence and the harmonic component, the middle term in (2), is about \( (7.07)^2 \), which is large enough to occasionally result in a false dominant peak. For larger amplitudes, the frequency resolution is in rough accord with the asymptotics, although there are occasional outlying values. Apart from some inflation of the standard deviation, the distribution of the phase estimates is about as predicted by the asymptotic theory.
Appendix

An approximation relating to the likelihood function for a Gaussian process

We consider a Gaussian stationary process and note that if the spectral density \( f(\lambda) \) is continuously differentiable and bounded away from zero the following approximations appear to be appropriate. The basic approximation is that of replacing \( r_{i,\tau}^{(-1)} (i, \tau = 1, \ldots, n) \) by

\[
\int_{-\pi}^{\pi} e^{it(\pi-\tau)} \frac{1}{4\pi^2} f^{-1}(\lambda) \, d\lambda.
\]  

(A1)

Aside from the term \(-\frac{1}{2} \log |R_\omega|\) the log likelihood can be written as

\[
\sum_{i,\tau=1}^{n} X_i X_\tau r_{i,\tau}^{(-1)} - 2 \sum_{i,\tau=1}^{n} X_i \{ A \cos (\omega \tau) + B \sin (\omega \tau) \} r_{i,\tau}^{(-1)} + A^2 \sum_{i,\tau=1}^{n} \cos (\omega \tau) \cos (\omega \tau) r_{i,\tau}^{(-1)}
\]

\[
+ B^2 \sum_{i,\tau=1}^{n} \sin (\omega \tau) \sin (\omega \tau) r_{i,\tau}^{(-1)} + 2AB \sum_{i,\tau=1}^{n} \cos (\omega \tau) \sin (\omega \tau) r_{i,\tau}^{(-1)}.
\]

The third term, on using the approximation (A1) mentioned above, is to first order

\[
\frac{A^2}{4\pi^2} \int_{-\pi}^{\pi} \frac{1}{f(\lambda)} \left| \sum_{i=1}^{n} \cos (\omega \tau) e^{it\lambda} \right|^2 \, d\lambda
\]

and if \( 0 < \omega < \pi \) this is approximated by \( \frac{1}{2\pi} \frac{n^2 \pi^{-1} A^2}{f(\omega)} \). The same argument shows that the fourth term is approximated by \( \frac{1}{2\pi} \frac{n^2 \pi^{-1} B^2}{f(\omega)} \). The fifth term can be shown to be \( o(n) \) if \( 0 < \omega < \pi \), since it is to the first order

\[
\frac{AB}{4\pi^2} \int_{-\pi}^{\pi} \frac{1}{f(\lambda)} \left\{ \sum_{i=1}^{n} \cos (\omega \tau) e^{it\lambda} \right\} \left\{ \sum_{i=1}^{n} \sin (\omega \tau) e^{-it\lambda} \right\} \, d\lambda.
\]

If we insert these approximations and minimize the resulting expressions as a function of \( A, B, \omega \) the estimates of \( A, B, \omega \) resulting are

\[
\hat{A} = \frac{4\pi f(\hat{\omega})}{n} \sum_{i,\tau=1}^{n} X_i r_{i,\tau}^{(-1)} \cos (\hat{\omega} \tau), \quad \hat{B} = \frac{4\pi f(\hat{\omega})}{n} \sum_{i,\tau=1}^{n} X_i r_{i,\tau}^{(-1)} \sin (\hat{\omega} \tau),
\]

where \( \hat{\omega} \) maximizes

\[
\frac{4\pi f(\omega)}{n} \left[ \left\{ \sum_{i,\tau=1}^{n} X_i r_{i,\tau}^{(-1)} \cos (\omega \tau) \right\}^2 + \left\{ \sum_{i,\tau=1}^{n} X_i r_{i,\tau}^{(-1)} \sin (\omega \tau) \right\}^2 \right].
\]  

(A2)

Again using the approximation (A1), it would appear to be the case that (A2) can be approximated by

\[
\hat{\omega} = \frac{2}{n f(\hat{\omega})} \left| \sum_{i=1}^{n} X_i e^{-\hat{\omega} i} \right|^2.
\]

From the same approximation, the estimates of \( A_0 \) and \( B_0 \) are given approximately by

\[
\hat{A} = \frac{2}{n} \sum_{i=1}^{n} X_i \cos (\hat{\omega} i), \quad \hat{B} = \frac{2}{n} \sum_{i=1}^{n} X_i \sin (\hat{\omega} i).
\]

Acknowledgement

This research is supported in part by the Office of Naval Research and National Science Foundation of the U.S.A.
REFERENCES


[Received August 1987. Revised January 1988]