Frequency Estimation Using Unequally-Spaced Astronomical Data

by

James Dennis Reimann

B.S. (University of Adelaide) 1990
M.A. (University of California at Berkeley) 1992

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Statistics

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor John Rice, Chair
Professor David R. Brillinger
Professor Donald C. Backer

1994

The dissertation of James Dennis Reimann is approved:

Chair Date

Date

Date

University of California at Berkeley

1994

ABSTRACT

F E U U S A D

by

James Dennis Reimann

Doctor of Philosophy in Statistics
University of California at Berkeley
Professor John Rice, Chair

This thesis studies estimation of the frequency of a periodic function of time, when the function is observed with noise at a collection of unequally-spaced times. This research was motivated by the detection and classification of variable stars in astronomy. Most of the statistical literature on frequency estimation assumes equally-spaced times, but observation times in astronomy are often unequally-spaced with a sampling distribution that contains periodic effects due to being able to collect data only at certain times of day.

In Chapter 1 we describe the database of variable stars collected by the MACHO collaboration and present examples which illustrate the common types of variable stars and the nature of the estimation problem.

In Chapter 2 we provide background material and give models for the periodic function and sampling times. We derive the asymptotic behavior of frequency estimates based on periodogram and least-squares estimation methods for sinusoidal periodic curves and sample times that are randomly distributed about equally-spaced values, and evaluate these estimators using a simulation study. We also discuss bounds on the variance of frequency estimates for general periodic functions under two sampling models.

In Chapter 3 we outline various methods for estimating frequency in practice, apply these methods to some example data, and compare their precision through the use of a simulation study.

To my parents,

Dennis and Gwen

Contents

Acknowledgements

This work would not have been possible without the guidance of my advisor, John Rice, who has taught me much, and given generously of his time, support, and encouragement.

I would like to thank the members of the MACHO collaboration for providing a fascinating problem and allowing me to use their hard-earned data. Kim Griest introduced me to some of the statistical issues in the MACHO data and shared many enjoyable discussions. Kem Cook greatly assisted in the application of these methods through his experience with variable stars. I am grateful to Charles Alcock and Lawrence Livermore National Laboratory for their support during 1994.

My dissertation committee, Don Backer, David Brillinger, and John Rice, suggested many improvements to the first draft which greatly enhanced the final product. Thanks also to Peter Bickel and David Brillinger for their interest in my research.

Thank you to the Statistics Department for financial support of my graduate studies and for help in many other ways. Sara Wong cheerfully and efficiently guided me through the administrative maze of the university, and Leslie Leonard went out of her way in preparing and explaining the departmental support. Thanks also to Chris Bush, Debbie Haaxman, Pat Hardy, Norma Hernandez, Myrna Hutcherson, and Jane Muirhead.

The Statistical Computing Facility is a tremendous asset of the department. I am grateful to Phil Spector for giving sound, immediate advice on almost any area of statistical computing. Thanks also to James Blakly, Rick Kawin, Ofer Licht, Keir Morgan, and Judd Reiffin for their assistance.

My fellow graduate students introduced me to the department and the computer system, helped me with teaching and research, and most importantly, provided friendship and comradery. Thank you to Helge Blaker, John Boscardin, Karl Broman, Babette Brumback, Sam Buttrey, Bill Forrest, Mark Hansen, Chad Heilig, Hank Ibser, Bret Larget, David Levin, Tony Lin, Dave Nelson, Steve Rein, Melissa Sanchez, Shaji Sebastian, Barathi Sethuraman, Nong Shang, Duncan Temple Lang, and Li-Ping Yang. I want to specially thank my office-mates in 305 Evans, Charles Kooperberg and HongYu Zhao, for their company and friendship.

My parents, Dennis and Gwen Reimann, have given me constant encouragement and backing through the many stages of my education.

Lastly, and therefore most importantly, thank you to my wife, Anita, for her unconditional love and support during the ups and downs of the Ph.D. experience.

Chapter 1

Introduction

In this thesis we consider estimation of the frequency of a periodic function of time, when the function is observed with noise at a collection of unequally-spaced times.

This research grew out of work on a problem in astronomy: the detection and classification of periodic variable stars. *Variable stars* are stars in which the intensity of the emitted energy changes over time, and in *periodic variable stars* the change of intensity is periodic over time. In this chapter we describe the substantial collection of light curve data that has been collected by one astronomical star survey. We summarize the common types of variable stars and present examples of some of these stars to illustrate the nature of the estimation problem.

In Chapter 2 we provide some background material on frequency estimation in time series and give models for the periodic function shape and the distribution of the sampling times. We derive the asymptotic behavior of two frequency estimators under the assumptions that the periodic function is sinusoidal and that the sampling times are distributed about equally-spaced values, and evaluate these estimators using a simulation study. Finally, we discuss frequency estimation for general periodic functions, obtain variance bounds for frequency estimators for two sampling models, and propose a variance bound for other sampling schemes.

In Chapter 3 we outline various methods for estimating frequency in practice and characterize the optimization problem inherent in these methods. We apply these methods to light curves of variable stars, and compare their precision through a simulation study. Finally, we note further issues in frequency estimation that were unable to be developed in this work.

The MACHO Project

The data upon which this research is based were collected at the Mount Stromlo Observatory near Canberra, Australia, by the MACHO collaboration, a group of scientists from the Center for Particle Astrophysics of the University of California at Berkeley, Lawrence Livermore National Laboratory, and Mount Stromlo and Siding Spring Observatories of the Australian National University. The collaboration is probing the halo of our galaxy in order to detect dark matter in the form of Massive Compact Halo Objects, commonly known as MACHOs. These are astronomical bodies that emit negligible visible light, such as dwarf or neutron stars, large planets, and black holes. Detection of a MACHO is achieved by observing its gravitational lensing effect on a chance background star as the MACHO crosses near the line of sight between the observer and this star. In order to detect a sufficiently large number of MACHOs, the collaboration needs to collect observations on an large number of distant stars over an extended period of time. Data are being collected daily over a 4-year period (weather permitting), on approximately 8 million stars in the Large Magellanic Cloud (LMC) and the bulge of the Milky Way. Approximately 250-300 observations have been collected per star over a 400-day observing period, and by the end of the project there should be more than 1200 observations per star. This database is a valuable resource for many other types of astronomical research. It is the most comprehensive catalog of stars in the LMC and contains stars much dimmer than those covered by previous surveys. The length of data-taking is unusually long compared to most star surveys, which permits a comprehensive study of star variability, including long periods and transient phenomena. Over 40,000 variable stars have been detected in the LMC data. The author has been collaborating with the MACHO group on methods for estimating the periodicity of the periodic variable stars in this group, which will be used to prepare a catalog of the estimated periods and amplitudes.

Data Collection

Observations are taken from the 1.27m "Great Melbourne" telescope at the Mount Stromlo Observatory. The incoming light stream is split by filters into two color bands, a blue band from 4500–6300 Angstroms and a red band from 6300–7600 Angstroms. Two large charge-coupled device (CCD) cameras are positioned at the focal points of the streams; each camera contains a 2×2 array of 2048×2048 pixel CCD imagers. Each digitized image thus contains over 16 million pieces of information in each color band. The photon counts at each pixel are adjusted to compensate for differing efficiencies between pixels; this is known as *flat-fielding*.

The digital images are reduced to a collection of star brightnesses by a photometry program “SoDophot”. It was designed for photometry (calculation of light intensity) on crowded sky images, which is the case with the images from the LMC and the galactic bulge. Incoming light from a point source is blurred by the atmosphere; the shape of the distribution of scattered photons is called the *point-spread function (PSF)*. SoDophot first fits a 7-parameter nonlinear model to the scattered light for a collection of bright stars that are isolated from their neighbors on the image. The seven parameters are the background intensity, the additional intensity at the center of the star image (the *star intensity*), and five parameters for modeling the shape of the point-spread function. The pooled estimates of the five PSF shape parameters from the bright stars are used in the estimation of the background intensity and star intensity for the remaining stars. As each star intensity is estimated in turn, the fitted point-spread function is subtracted from the data before the next star is fitted; if two stars have overlapping images, a bimodal estimation is done on the two together. The intensities of each star are converted to fluxes to compensate for the observing conditions of the night; this is done by comparing the intensity of non-varying bright stars with their previously estimated value.

Error in the flux measurements can come from multiple sources, which can generally be described as *additive* or *multiplicative*. The size of additive errors does not depend on the flux, while the size of multiplicative errors is proportional to the flux. Poisson-like behavior in the photon counts contributes additional error which is inversely proportional to the square root of the flux. An estimate of the standard error of each flux measurement (combining all components of error) is produced by the photometry program. The standard error estimates come from the curvature of the residual sum of squares (RSS) surface for the estimation of the background and star intensities, with a lower bound specified on the errors to prevent unrealistically low values. There is reason to believe that the error estimates are appropriate for the weaker stars and are slightly too large for the stronger stars. Systematic bias is also present in the flux measurements. As this bias varies between stars but not within multiple measurements on the same star, this is not a problem when considering changes in flux of a given star over time. There is prior evidence that the noise in the flux measurements is correlated with the *seeing conditions*, which is the state of the atmosphere that determines the shape and spread of the point-spread function. When the seeing conditions are bad, i.e., the incoming light is severely blurred by the atmosphere, it is difficult to resolve nearby stars and the quality of the flux estimate deteriorates. As seeing conditions a few hours apart are effectively unrelated, and repeat measurements in the MACHO data are spaced at least a few hours apart, the noise in the flux measurements can be considered independent. It is observed that the measurement noise is also approximately normally distributed.

Variable Stars

In order to find the variable stars in the overall population of stars, the average flux of each star is subtracted from the sequence of measurements and the weighted residual sum of squares (WRSS) is calculated. Stars are considered variable if the WRSS falls above a given value, which changes with star baseline flux to compensate for the bias in the standard error estimates. Up to the present time, this has yielded over 40,000 variable stars.

The most common types of periodic variable stars that will be detectable in the MACHO data are eclipsing binaries, RR Lyraes, Cepheids, and Long Period Variables. *Eclipsing binaries* are binary stars (two stars orbiting each other) for which brightness variability occurs because one star passes in front of the other in turn; as the stars may be of different brightnesses, the drop in light flux depends on which star is in the front. These stars have periods of between 3 hours and 24 years, although 0.5 to 10 days is the most common range. The brightness changes in the remaining classes of periodic variables is caused by periodic pulsation (contraction and expansion) of the stars and their outer layers. *RR Lyrae stars* are the second most common class of known periodic variable stars; they have periods in the range 0.2 to 0.9 days. There are two common types, RRab stars which have an asymmetric signal and RRc stars which have a symmetric signal. *Cepheids* are rare, very bright stars with periods of 1-70 days. The light curve has an asymmetric shape, and rises more rapidly than it falls. Cepheids with periods of about 1 week tend to have a bump in the descending part of the curve. For periods of about 10 days, the bump is at the peak of the curve, and for longer periods it is on the rising part of the curve. The most common class of known periodic variable stars are the *Long Period Variables* (LPVs). These are red giant stars with periods in the range 30-1000 days. The period can vary by about 10% with accompanying changes in average flux and amplitude of the harmonic component, and so these are more accurately semiperiodic stars. Other types of periodic variable stars are not expected to be detected in any great number by the experiment.

Examples

We present five examples of common types of periodic variable stars which illustrate typical periodic curve shapes. The first star is a cepheid variable of magnitude 16.1 in the red band

and of magnitude 16.5 in the blue band. [The *magnitude* scale is a logarithmic measure of flux,

$$m = -2.5 \log(f/f_s),$$

where f is the measured flux and f_s is a standardization constant; note that weaker stars have *larger* magnitudes.] Figure 1.1 shows the brightness plotted against time for the two bands. There are 323 observations in the red band and 203 observations in the blue band. The observation times are measured in days since January 2, 1992, and span approximately 400 days. There are substantially less observations in the early part of the time period when the experiment was not yet running smoothly, and indeed near day 250 there was no data taken for 25 days. The vertical axis, A-1, is the ratio of the measured flux to a average flux for that star, shifted to have mean zero. This quantity is correctly called the *normalized flux*, but for convenience we shall refer to it as "brightness" in the remainder of this document. At each point, the small dot represents the brightness estimate and the two bars lie at the ± 1 SE values.

Denote the brightness and observation time of the jth observation by y_j and t_j for $j = 1, \ldots, n$. If the brightness (without noise) has a strict periodic dependence on time with period p, then the brightness y should also be dependent on the circular variable t mod p, and a plot of y_j versus $\frac{1}{p}t_j$ mod 1 should describe the nature of the dependence of y_j on t_j at period p. We call this the *phase plot* at period p. Figure 1.2 shows the phase plot of the same star at period 1.44324 days. The phase axis has been extended by 0.2 units at each end so that the shape of the plot near phase values zero and one can be seen more easily. Dependence of y_j on t_j can be seen clearly in these plots; the shape of the curve is approximately sinusoidal and there are few outlying points.

The second example is also a cepheid star, but is brighter than the first: it is of magnitude 15.3 in the red and 15.8 in the blue. Figure 1.3 shows the brightness plotted against time. The standard errors are lower than in the previous example, and the error bars increase in length as the measured flux increases. There appears to be four horizontal "bands" in each scatter plot. There are 303 observations in the red band and 254 observations in the blue band. Figure 1.4 shows the phase plot for these data at period 3.9862 days. The shape of the dependence of brightness on time is non-sinusoidal and there are four gaps in the data between phases 0 and 1. These gaps are due to the frequency of the oscillation: as the period is very close to four days, parts of the signal cycle occur only during the day, when astronomical observations cannot be taken. Thus any method used to detect and estimate periodicity in data of this type must handle missing information on parts of the curve.

The third example is an eclipsing binary star of magnitude 16.4 both in the red and blue

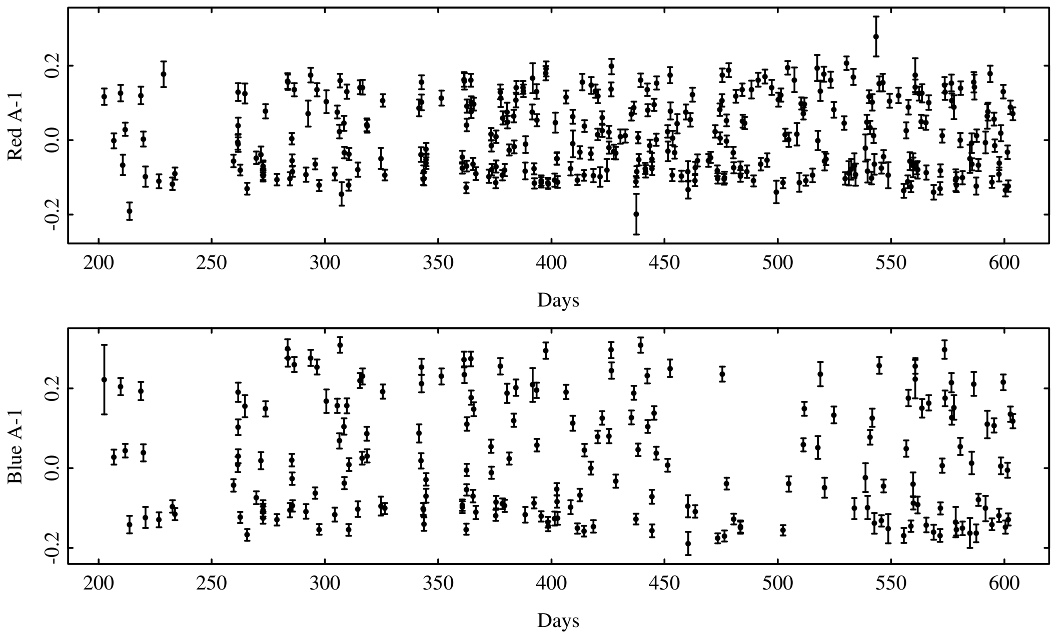

Figure 1.1: Brightness (normalized flux) plotted against time, for the red and blue bands of star 77021:1992.

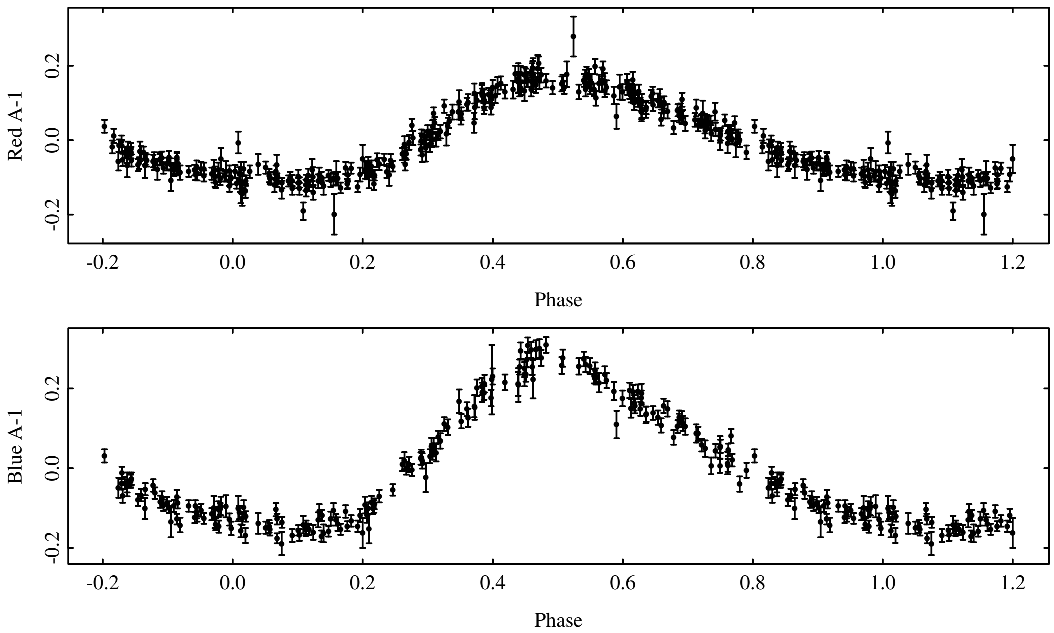

Figure 1.2: Phase plot of star 77021:1992 at period 1.44324 days.

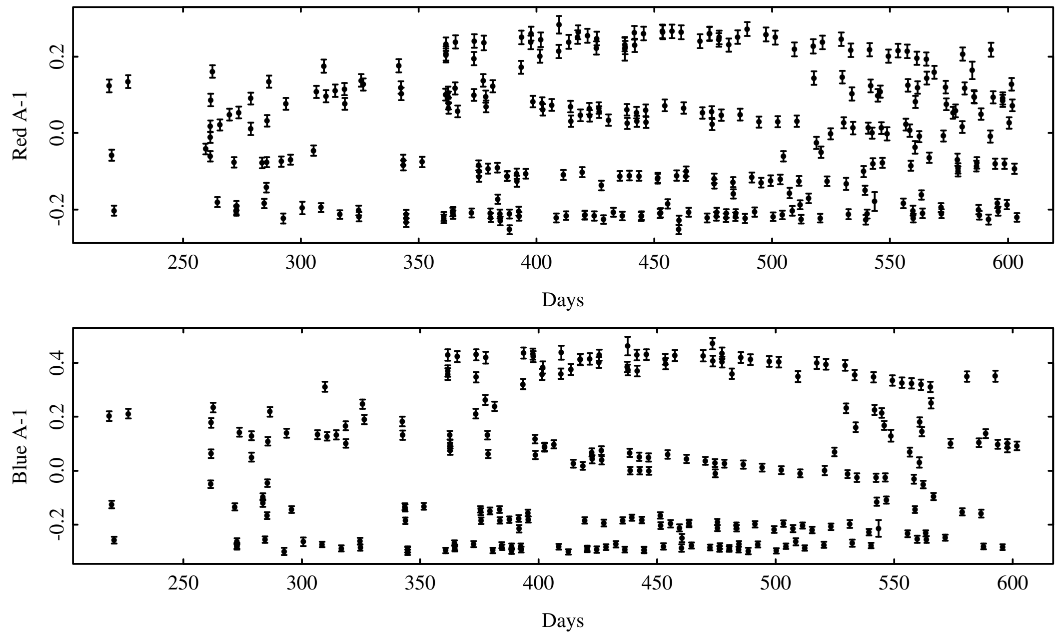

Figure 1.3: Brightness plotted against time for the red and blue bands of star 77010:788.

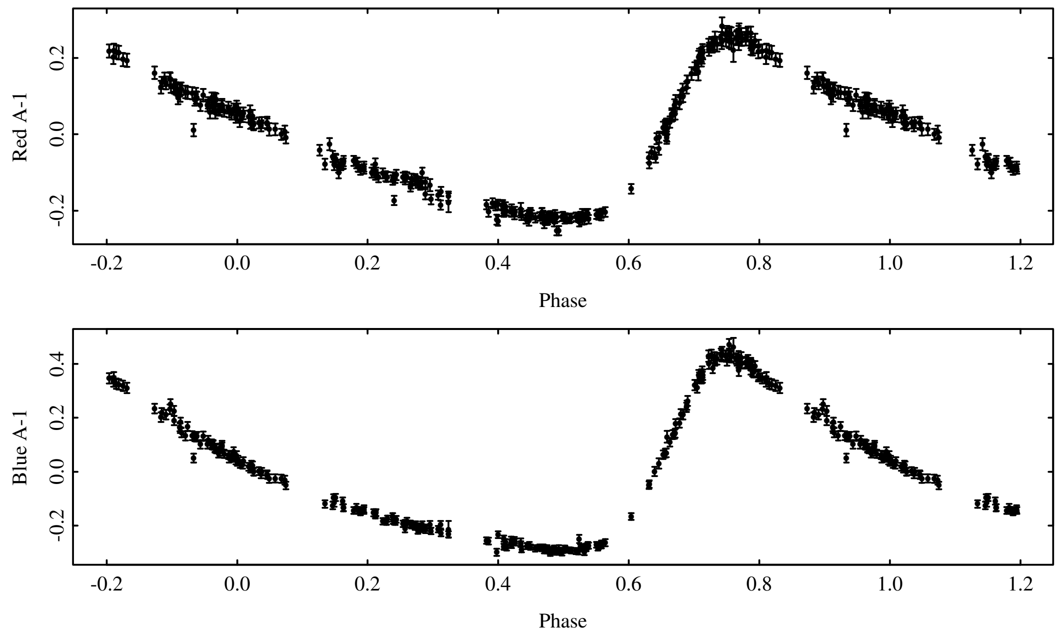

Figure 1.4: Phase plot of star 77010:788 at period 3.9862 days.

bands. The plots of the brightness against time are shown in Figure 1.5; there are 300 observations in the red band and 305 observations in the blue band. Note that there is a skew in the observed brightness: there are many observations with brightness close to zero, and a smaller number with negative brightness, but few with large positive values. This is typical of eclipsing binary stars, and provides a method of identifying them without even estimating the periodicity. Figure 1.6 shows the phase plot for period 2.47133 days. The curve shape is typical of eclipsing binaries: the brightness is at a constant level while the two stars are side by side, with less light emitted when one of the stars is positioned behind the other. The differing depths of the troughs is caused by one star being brighter then the other; when the bright star is in front, the trough is less deep. This signal shape can be more difficult to detect than a unimodal cycle shape, and a general period estimation approach should be able to accommodate eclipsing binaries.

The fourth example is an RR Lyrae star which is much dimmer then the previous examples. It has magnitude 19.1 in the red band and 19.5 in the blue band, which is very close to the limit of resolution of the experiment. The plots of brightness against times are shown in Figure 1.7; the red band had 304 observations and the blue band had 310. The error bars are much wider than in the other examples, with standard deviation of about 0.1, compared with about 0.015 for the second example and about 0.02 for the first and third examples. This means that the amplitude of the oscillation must be very large to be seen above the noise. The phase plots for these data at period 0.52715 days are shown in Figure 1.8. The periodic dependence appears clearly in the blue band but less strongly in the red band. RR Lyrae stars typically have periods of less than one day, and so methods for detecting them must be able to search for periods well under the sampling rate, which for these data is about one observation every two days.

The final example is a relatively bright LPV star, of magnitude 15.0 in the red band and 16.2 in the blue band. Plots of the brightness against time are shown in Figure 1.9; there are 322 observations in the red band and 277 observations in the blue. The brightness varies slowly over time and seems not to be strictly periodic, with changes in amplitude and baseline over time. Because there are only a few cycles present in these data, many more observations are needed to describe the non-periodic behavior of stars like this one.

In these examples, we have seen that the shape of the oscillation is not necessarily sinusoidal, and that the curve may even be bimodal over a single cycle. We saw also that when the period is close to a multiple of days, there may not be information available on all parts of the curve, and that for stars such as RR Lyraes, the noise in the data can be large in comparison with the amplitude of the oscillation. Finally, there is a class of stars which are obviously variable, but

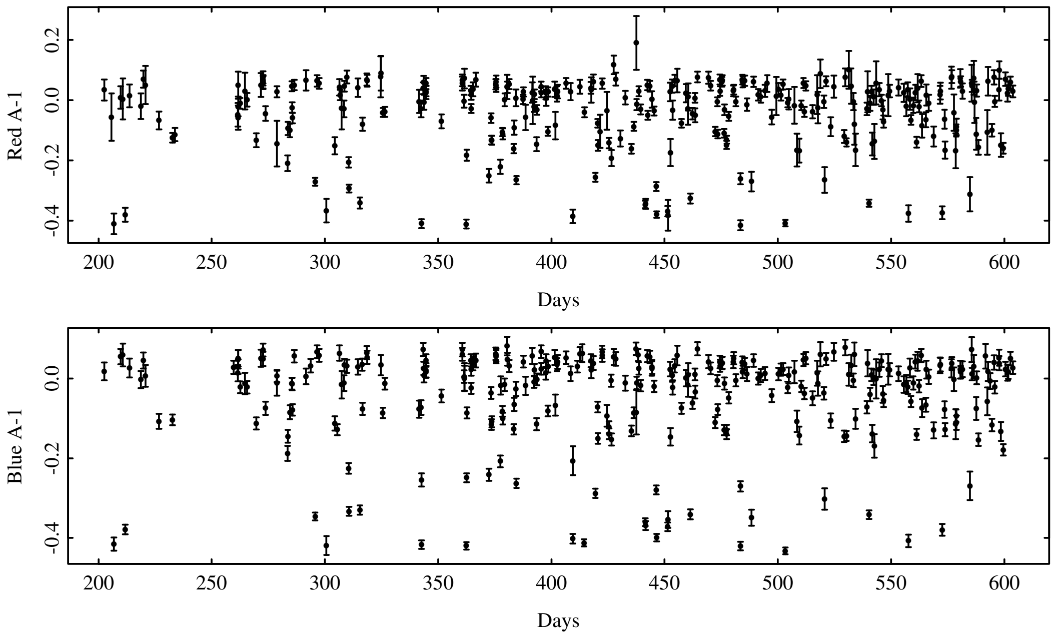

Figure 1.5: Brightness plotted against time for the red and blue bands of star 77043:4317.

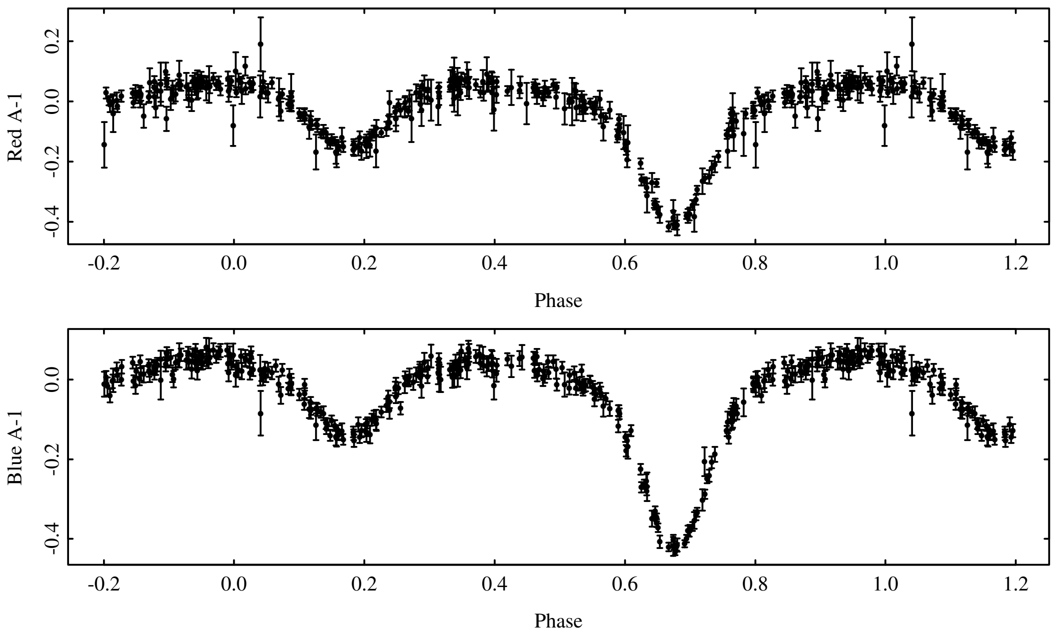

Figure 1.6: Phase plot of star 77043:4317 at period 2.47133 days.

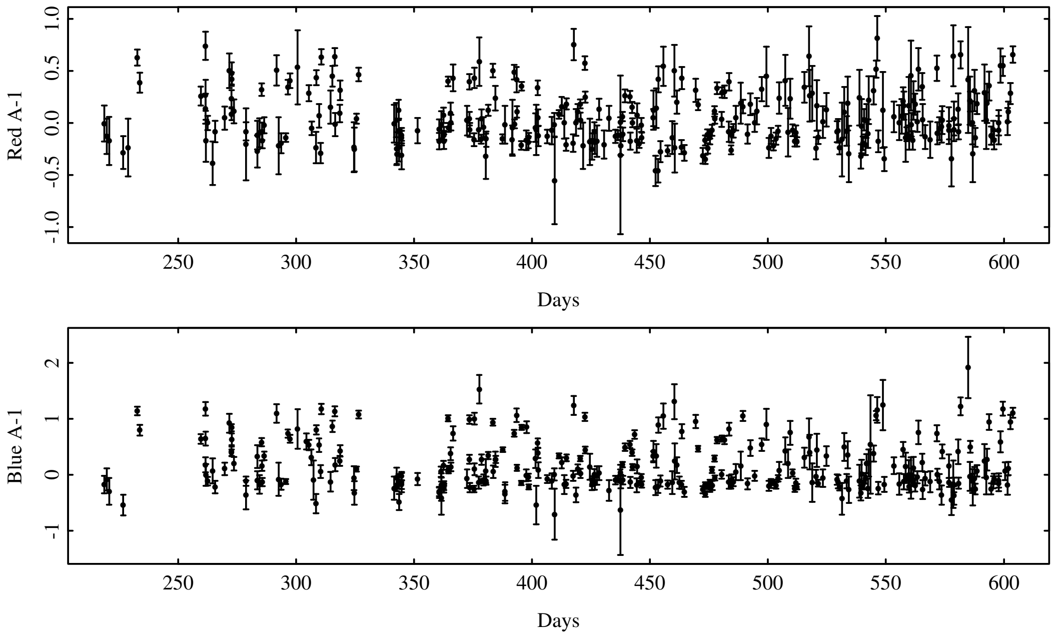

Figure 1.7: Brightness plotted against time for the red and blue bands of star 77009:64163.

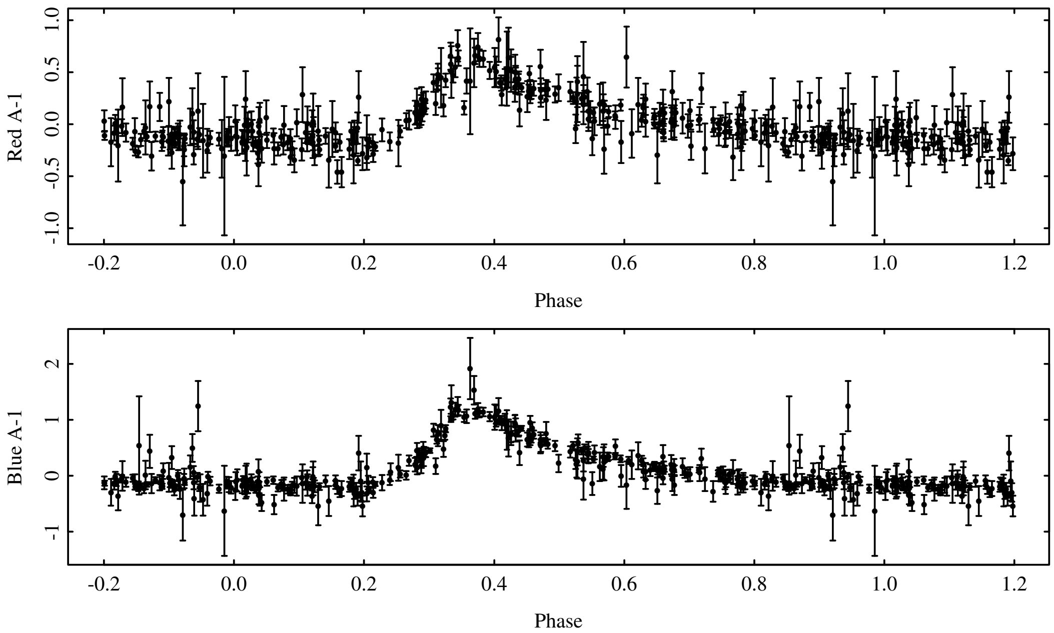

Figure 1.8: Phase plot of star 77009:64163 at period 0.52715 days.

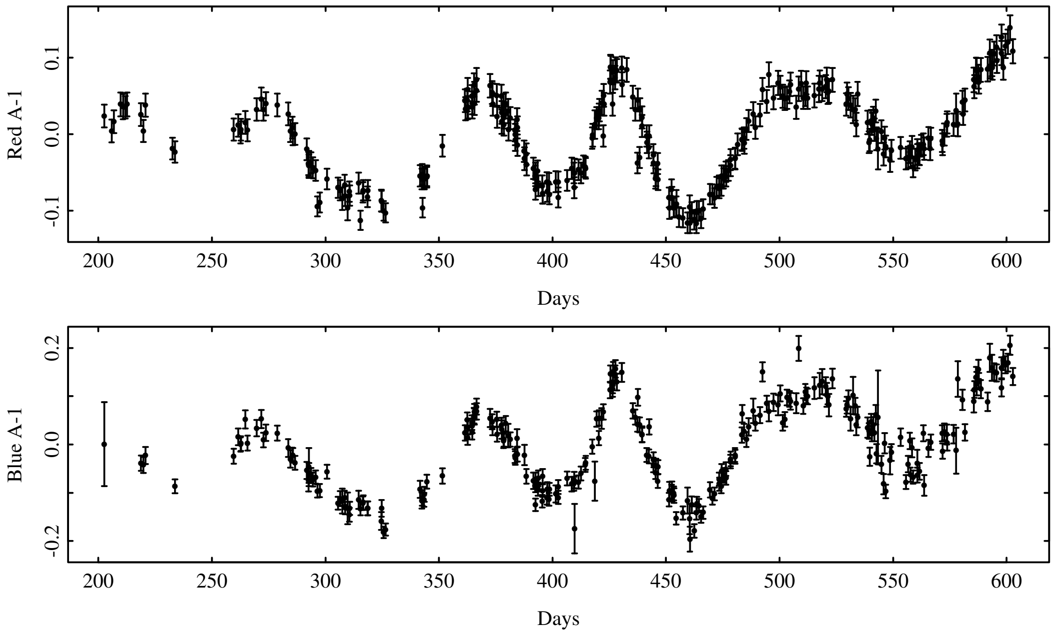

Figure 1.9: Brightness plotted against time for the red and blue bands of star 78017:497.

that are not perfectly describable by periodic functions.

Chapter 2

Theory of Frequency Estimation

2.1 Introduction

In this chapter we determine properties of frequency estimation methods, including consistency of the frequency estimate, rate of convergence, and influence of the sampling times on the estimation. In its most general form, the periodic regression problem can be expressed as

$$y_j = s(wt_j) + \epsilon_j,\ j = 1, \ldots, n, \tag{2.1}$$

in which t_j denotes the jth sampling time, w is the frequency, ϵ_j is the measurement noise in the jth observation, s is a continuous periodic function of period one, and y_j is the jth brightness measurement. This model is discussed in Bickel, Klassen, Ritov & Wellner[6], p. 107, which presented it as an example of a semiparametric regression model and investigated its asymptotic behavior. McDonald [30] also discussed this model, as well as estimation of w by the use of an edge-preserving smoother. In a different context, that of estimation of motion in meteorological data, Brillinger [12] estimated the frequency of a periodic function of space and time and compared the performance of Fourier and nonparametric (spline-based) approaches. There is little else in the literature for the general model (2.1), but much work has been done on the *simple harmonic model*,

$$y_j = a\cos(wt_j) + b\sin(wt_j) + \epsilon_j, \tag{2.2}$$

and the generalization to multiple frequencies,

$$y_j = \sum_{k=1}^{p} \{a_k \cos(w_k t_j) + b_k \sin(w_k t_j)\} + \epsilon_j. \tag{2.3}$$

Note that the above model with $w_k = kw$ is contained within (2.1) as the set of functions s that can be expressed as a p-term Fourier expansion.

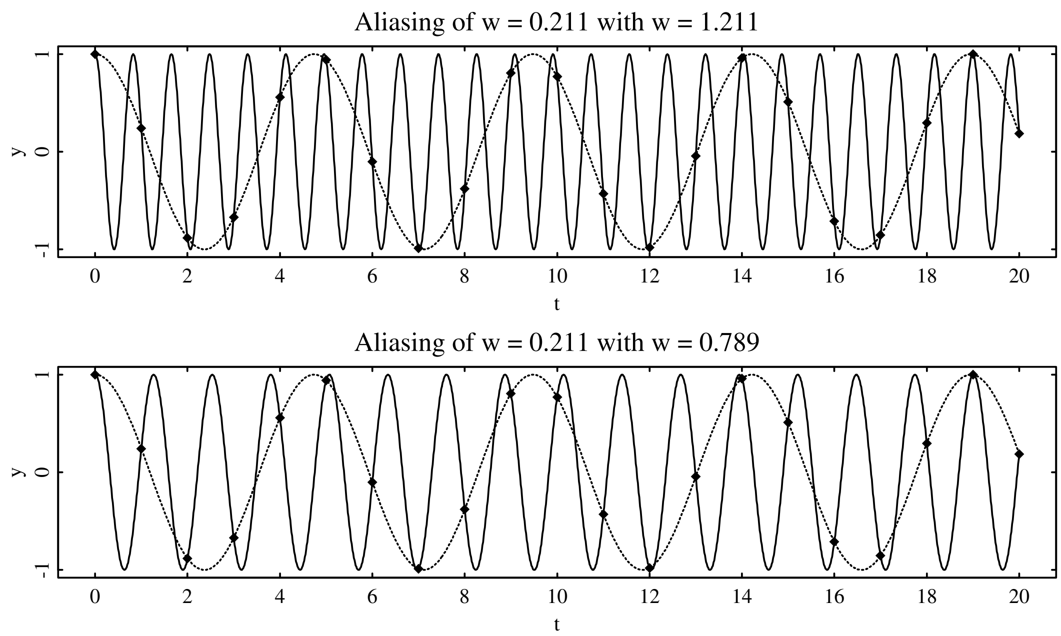

Figure 2.1: The curve $y = \cos\,(0.211\,2\pi t)$, as a dotted line, displayed with $y = \cos(1.211\,2\pi t)$ in the upper plot and $y = \cos(0.789\,2\pi t)$ in the lower plot.

Background

Extensive theoretical work has been done on the simple harmonic model with equally-spaced sampling times,

$$t_j = j,\ j = 1, \ldots, n,$$

beginning with Whittle [50]. Hannan [22] and Walker [48] formalized and generalized Whittle's results. These works used the periodogram as an approximation to least squares, and found that the estimate of w is consistent for frequencies in the range $[0, \pi]$, that the asymptotic variance of the frequency estimate is of order n^{-3}, and that the asymptotic variances of the estimates of a and b are of order n^{-1}. These results extend to the model with multiple harmonic components (2.3). Hannan [22] also showed that the rate of convergence of the frequency estimate is faster than n^{-3} for $w = 0, \pi$ (see Section 2.2.4 for further discussion).

The frequency estimation is restricted to the range $[0, \pi]$ for equally-spaced times because of *aliasing*. One frequency is an alias of another if the signals at the two frequencies have identical values at the sample times. This is illustrated in Figure 2.1. The dotted line in each plot is the curve $y = \cos(0.211\,2\pi t)$, plotted over $t \in [0, 20]$. The solid line in the upper plot is the curve $y = \cos(1.211\,2\pi t)$, and the solid line in the lower plot is the curve $y = \cos(0.789\,2\pi t)$. The value of the curves at integer times is marked with a diamond, and is identical for the two curves in each plot. The frequencies $w = 1.211$ and $w = 0.789$ are thus indistinguishable from $w = 0.211$ at these times. Indeed, this is true for all frequencies of the form $w = \pm 0.211 + k,\ k \in \mathcal{Z}$, as $\cos((\pm w + k)2\pi t) = \cos(w\, 2\pi t)$ for integers k. Aliasing is also discussed in Bloomfield [8].

A few writers discuss sampling models for (2.2) other than equally-spaced times. Ivanov [26] studies continuously-sampled time, and showed consistency of the least squares estimate of frequency for $w \in (0, \Omega),\ \Omega < \infty$, with the asymptotic variance being of order n^{-3}. Kutoyants [27] also considers the continuous-time process, as well as frequency estimation in the intensity function of a point process. Isokawa [25] and Brillinger [11] study the frequency estimate obtained from the periodogram for sample times generated by a point process that is stationary and mixing. These works find that the frequency estimate is consistent for $0 < w < \Omega < \infty$, and has asymptotic variance of order n^{-3}. Thrall [44] derives spectral estimates for a sampling model in which the probability of an observation being taken on a given day depends on the day of the week.

None of the above models are appropriate for our data, in which the sampling times are not equally-spaced, but there is periodicity in the sampling distribution (as observations can be taken only at night) that violates the mixing assumption for the point process model. We need to know how the sampling scheme affects the frequency estimation and how this differs from what occurs in the above models.

Spectral Windows

Consider the Fourier transform of the periodic regression function s. If $s(wt)$ is a simple harmonic component at frequency w, then the Fourier transform of $s(wt)$,

$$F(v) = \int_{-\infty}^{\infty} s(wt) e^{i2\pi vt}\, dt,$$

is zero everywhere but at $v = w$ and $v = -w$. We only have information about s at the sample times t_j, and the discrete Fourier transform,

$$F_n(v) = \tfrac{1}{n}\sum_{j=1}^{n} s(wt_j) e^{i2\pi vt_j},$$

is a natural measure of the periodicity of the observed function. $F_n(v)$ is related to $F(v)$ by a convolution result,

$$\begin{aligned} F_n(v) &= F(v) * \delta_n(v) \\ &= \int_{-\infty}^{\infty} F(v-u)\delta_n(u)\, du, \end{aligned}$$

where δ_n, the *spectral window* of times t_j, is defined by

$$\delta_n(v) = \tfrac{1}{n}\sum_{j=1}^{n} e^{i2\pi v t_j}.$$

If $s(wt)$ is a simple harmonic component with Fourier transform $F(w) = c$, $F(-w) = c^*$ and $F(v) = 0$ otherwise, then

$$F_n(v) = c\,\delta_n(v - w) + c^*\,\delta_n(v + w).$$

If the spectral window is significantly different from zero at frequencies other than $v = 0$, then the discrete Fourier transform will be large at frequencies other than at $v = \pm w$. A typical spectral window has a well defined peak at $v = 0$ with approximate width of $2T^{-1}$, where T is the span of the sample times, and numerous smaller peaks coming from the spacing of the sample times. Deconvolution methods can be used to approximately remove the effect of the time sampling from the discrete Fourier transform if the correct frequency of the harmonic components is known. One implementation of this is the CLEAN algorithm (Schwarz [41]).

Figure 2.2 shows the modulus of the spectral window of 300 equally-spaced sample times. There is a sequence of peaks of height one at the frequencies $v = 1, 2, 3, \ldots$ with smaller side peaks around the main peaks. Thus the modulus of the discrete Fourier transform of a simple harmonic component at these times will have peaks of equal height at the frequencies $\pm w + k,\ k \in \mathcal{Z}$. Furthermore, the modulus of $F_n(v)$ would be the same if s was a simple harmonic component with the same amplitude as before but now with frequency at any of $\pm w = k$, and we see as discussed above that the simple harmonic components with these frequencies are indistinguishable on equally-spaced times. For continuous data and a stationary and mixing point process, the spectral window tends to zero everywhere but at $v = 0$, so aliasing effects become unimportant as n becomes large.

What does this mean for our observation times? The spectral window of the 300 observation times of the red band of star 77043:4317 are plotted over the frequency intervals $[0, 5.5]$ and $[0, 0.15]$ in Figure 2.3. There are significant peaks at the frequencies $v = k$, but the size of the peaks decreases with k. There are numerous smaller peaks visible which are larger than the small peaks in the spectral window of the equally-spaced times. We see from this that the Fourier transform of a simple harmonic component would have large modulus at frequencies $\pm w + k$, with height decreasing with k. This means that a harmonic component at some frequency $\pm w + k$ would not have an identical Fourier transform to the signal with frequency w, but that it would have

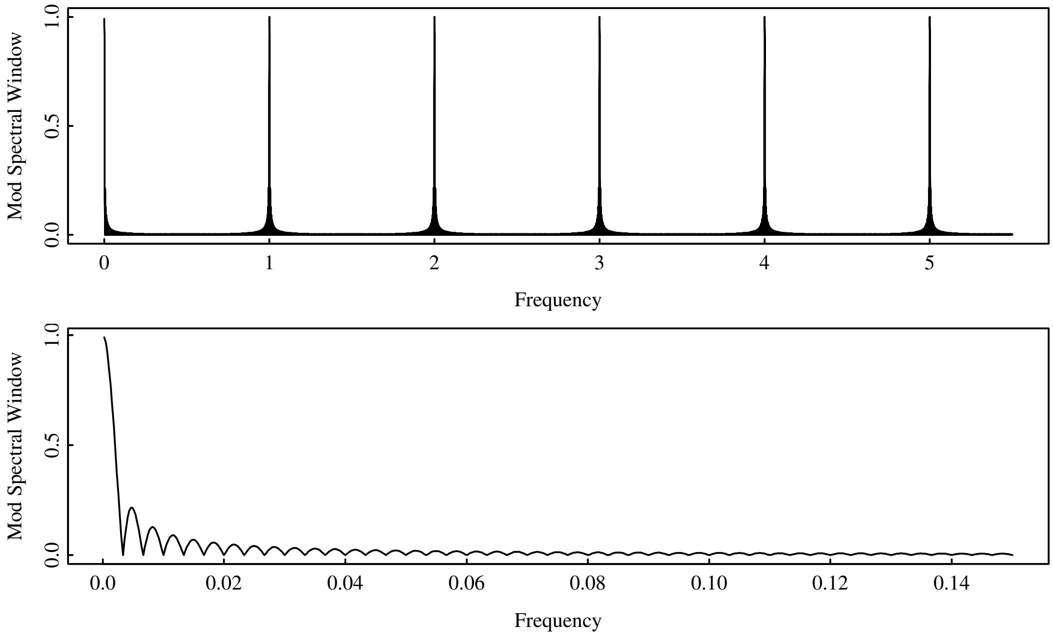

Figure 2.2: Modulus spectral window of 300 equally-spaced sample times, displayed over the frequency intervals $[0, 5.5]$ and $[0, 0.15]$.

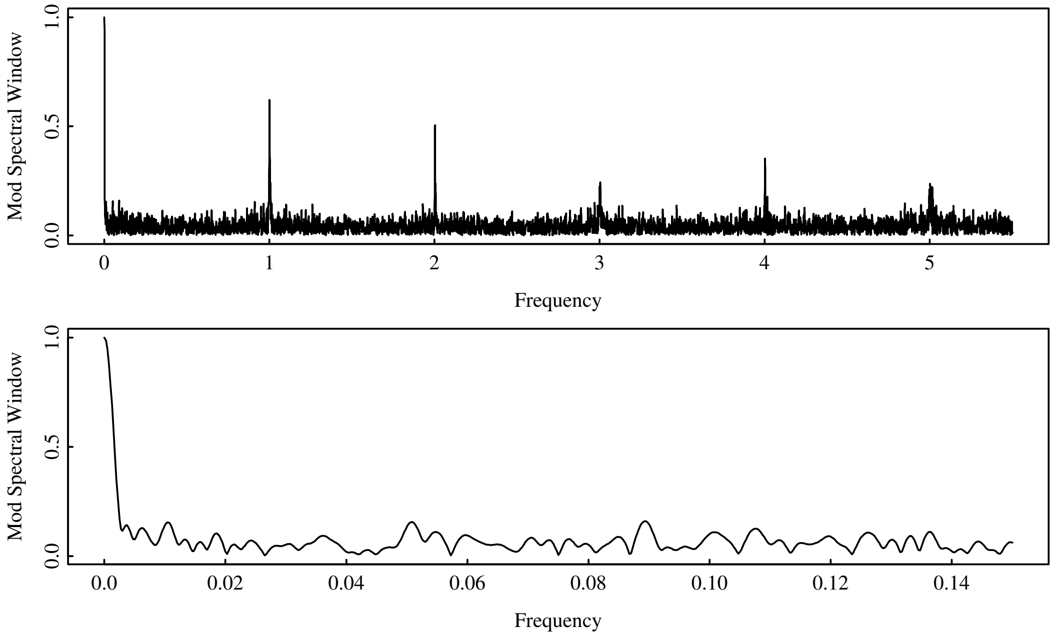

Figure 2.3: Modulus spectral window of 300 sample times of the red band of star 77043:4317, displayed over the frequency intervals $[0, 5.5]$ and $[0, 0.15]$.

peaks in the modulus for many of the same frequencies. We call this effect *pseudo-aliasing* (after Scargle [36]) since the signals at these frequencies will not be identical to the signal at the true frequency, as in aliasing, but will still be similar to this signal. Similarly, the frequencies $\pm w + k$ for $k \neq 0$ are called *pseudo-aliases*.

Jittered Sampling Times

We need a model that captures the essence of this behavior: that there is a periodicity in the distribution of the sampling times but that they are not equally-spaced. Consider the sampling model in which the observation times are randomly perturbed about uniformly-spaced values,

$$t_j = j + \delta_j, \quad |\delta_j| < \Delta, \quad j = 1, \ldots, n, \tag{2.4}$$

where the $\{\delta_j\}$ are independently and identically distributed (IID) with probability density function h, have mean zero, and are distributed independently of $\{\epsilon_j\}$. This observation model is called *randomly jittered sampling* in Beutler [4], and is also discussed in the context of spectral estimation in Akaike [2]. As presented, the jittered sampling process is not stationary; it can easily be made stationary by making the time scale arbitrary (e.g., by adding the same $U[0,1]$ random variable to all the sample times). The resulting process is not mixing because the autocovariance function $C_{NN}(u)$ is periodic for large u (This is discussed in more detail in the Comments of Section 2.2.1). This model mimics astronomical observations which are taken daily but at varying times. This is only a first approximation to real life, in which there can be multiple observations taken per night, as well as stretches of days when no data can be collected due to bad weather. Nevertheless, it captures the most important feature of the sampling: that the observations come from continuous time with a strong day effect in the sampling distribution.

The modulus of the spectral window for 300 times generated from the jittered sampling model with $\delta_i \sim U[-\frac{1}{6}, \frac{1}{6}]$ is displayed in Figure 2.4 over the frequency ranges $[0, 5.5]$ and $[0, 0.15]$. The modulus spectral window takes on the value $|\phi(4\pi v)|$ at the frequency $v = k, k \in \mathcal{Z}$, where $\phi(.)$ is the characteristic function of h, and tends to zero otherwise as n gets large. The heights of the side peaks are small when density h has a large variance, are large when h has a small variance, and are of height one when the δ_j are identically zero (the equally-spaced case). Note that over the range $[0, 0.15]$, the spectral window looks more like that of Figure 2.2 than that of Figure 2.3. This is because at the short frequencies (long periods), the jittered sampling is effectively equally-spaced, with one observation per day, while the actual sampling times have varying numbers of

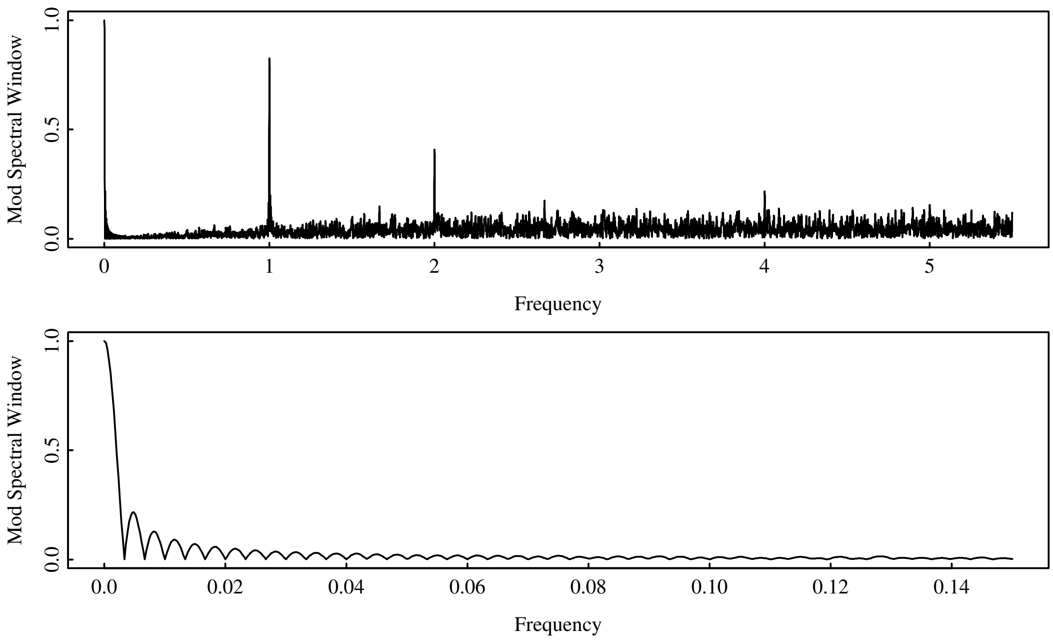

Figure 2.4: Modulus spectral window of 300 times generated from the jittered sampling model with $\delta_i \sim U[-\frac{1}{6}, \frac{1}{6}]$, shown over the frequency ranges $[0, 5.5]$ and $[0, 0.15]$.

observations per day. Limitations of the jittered sampling model are discussed further in Section 2.4.2.

Much of the theory that we present is based on the simple harmonic model (2.2) with parameter vector $\theta = (a, b, w)$ satisfying

$$a^2 + b^2 \in (0, \infty), \quad w \in [0, \infty), \tag{2.5}$$

and in which the measurement noise ϵ_j are IID with mean zero and variance σ^2. We call this response model the *jittered cosine model* when the sampling times are distributed as in (2.4).

Overview of the Chapter

The properties of frequency estimation under the simple harmonic model are investigated in Sections 2.2–2.4. Frequency estimation based on the periodogram is discussed in Section 2.2, and the maximum likelihood estimator is discussed in Section 2.3. Conditions are given for consistency of the estimators and the asymptotic distributions of the estimates are calculated. Both estimators are found to be consistent for frequencies in the space

$$w \in (0, \Omega], \ w \neq k\pi, \ k \in \mathcal{Z},$$

but the maximum likelihood method is asymptotically efficient while the periodogram estimator is inefficient. We also discuss extensions to the simple harmonic model: the addition of a constant term and consideration of several harmonic components.

Section 2.4 presents the results of simulations which evaluate the behavior of the periodogram and maximum likelihood estimates at and near the frequencies $w = k\pi$, and compares the observed finite-sample distribution of the frequency estimates with the asymptotic distributions for frequencies not close to $w = k\pi$. Two types of sampling times were used in the simulations: simulated times from the jittered model (2.4), and actual times from one of the MACHO light curves. The simulations found that the maximum likelihood estimator appeared to be consistent at $w = k\pi$ while the periodogram estimator was sometimes inconsistent, and that the finite sample distributions agreed closely with those predicted by asymptotic theory.

In Section 2.5 we discuss issues that arise in estimation using the semiparametric model (2.1), derive variance bounds for frequency estimators for a couple of sampling models, and propose an estimate of the best attainable precision of the frequency estimate under more general sampling schemes.

2.2 Estimating Frequency Using the Periodogram

In this section we derive the asymptotic properties of a frequency estimate based on the *periodogram*, defined by

$$I_n(w) = \frac{2}{n}\left|\sum_{j=1}^{n} y_j\, e^{iwt_j}\right|^2 . \tag{2.6}$$

This is a natural extension to the definition of the periodogram for equally-spaced times, which is the modulus-squared of the finite Fourier transform of the time series. Assume the simple harmonic model (2.2) and that the errors ϵ_j are IID with mean zero and variance σ^2. Define the *periodogram estimate of frequency* $\hat{w}_n$ on some interval $[0, \Omega]$ to be the frequency which maximizes the periodogram on that interval,

$$I_n(\hat{w}_n) = \sup_{v\in[0,\Omega]} \{I_n(v)\}. \tag{2.7}$$

Further define estimators of a, b and σ^2

$$\hat{a}_n \;=\; \frac{2}{n}\sum_{j=1}^{n} y_j \cos(\hat{w}_n t_j)$$

$$\hat{b}_n = \frac{2}{n}\sum_{j=1}^{n} y_j \sin(\hat{w}_n t_j)$$

$$\hat{\sigma}_n^2 = \frac{1}{n}\left\{\sum_{j=1}^{n} y_j^2 - I_n(\hat{w}_n)\right\} \tag{2.8}$$

The periodogram was introduced by Schuster [38], who applied it to the analysis of various data sets [39, 40].

The periodogram estimate is asymptotically equivalent to the least squares estimate under certain conditions. If the errors are normally distributed and the distribution of the sample times does not depend on the parameters, then the log-likelihood of the observations $y_1, \ldots, y_n$ is equal to

$$L_n(a, b, w, \sigma^2) = -\frac{1}{2}n\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\mathrm{SS}_n(a, b, w),$$

plus a constant term. Here, $\mathrm{SS}_n()$ is the residual sum of squares (RSS)

$$\mathrm{SS}_n(a, b, w) = \sum_{j=1}^{n}\{y_j - a\cos(wt_j) - b\sin(wt_j)\}^2.$$

Thus the maximum likelihood estimate of $[w, a, b]$ is obtained by minimizing the RSS, and we call this the *least squares estimate*. In comparison, the periodogram estimate is the estimate obtained by minimizing U_n, a modified form of the RSS

$$\mathrm{U}_n(a, b, w) = \sum_{j=1}^{n} y_j^2 - 2\sum_{j=1}^{n} y_j\{a\cos(wt_j) + b\sin(wt_j)\} + \frac{1}{2}n(a^2 + b^2).$$

The difference between the RSS and U_n is

$$\mathrm{U}_n(a, b, w) - \mathrm{SS}_n(a, b, w) = \frac{1}{2}\sum_{j=1}^{n}\{(a^2 - b^2)\cos(2wt_j) + 2ab\sin(2wt_j)\}. \tag{2.9}$$

This is uniformly O(1) for w bounded away from $0, \pi$ when the sampling times are equally-spaced, and so the periodogram and least squares estimates are equivalent for this model. Ibragimov and Has'minskii [24] found that the least squares estimate was asymptotically efficient for $w \neq 0, \pi$. As the asymptotic variance they obtained was the same as that derived by Hannan [22] and Walker [48] for the periodogram, the periodogram estimate is also efficient for equally-spaced times. For sampling times distributed according to the jittered model (2.4), the difference (2.9) is $O_p(n^{1/2})$ for $w \neq k\pi$, and we will see in Sections 2.2 and 2.3 that the periodogram estimate has larger asymptotic variance than the least squares estimate.

In Section 2.2.1 we present consistency and asymptotic normality results for the periodogram estimates under the jittered model; the proofs of these are given in Sections 2.2.2 and 2.2.3. We find that the periodogram estimate is consistent for $w \neq k\pi$, and has asymptotic variance larger than that for equally-spaced times and which depends on the jittering distribution. In Section 2.2.4 we discuss the case $w = k\pi$, and extend the results of Section 2.2.1 to models incorporating a constant term or multiple periodic components.

2.2.1 Asymptotic Results

Theorem 1 *Consider the jittered cosine model (2.4), with $\theta \in \Theta$*

$$\Theta = \{(a, b, w) : \ a^2 + b^2 \in (0, \infty); \ w \in [0, \Omega]; \ w \neq k\pi, \ k \in \mathcal{Z}\}.$$

Also assume:

(1)

$$E\,|\epsilon_1|^5 < \infty. \tag{2.10}$$

(2) Denoting the characteristic function of δ_1 by $\phi(.)$,

$$\forall \eta > 0, \ \sup_{|t|>\eta} |\phi(t)| < 1. \tag{2.11}$$

Then the estimators defined in (2.7) and (2.8) satisfy

$$\begin{aligned} p \lim_{n\to\infty} n\hat{w}_n &= nw, \\ p \lim_{n\to\infty} (\hat{a}_n, \hat{b}_n, \hat{\sigma^2}) &= (a, b, \sigma^2). \end{aligned} \tag{2.12}$$

Here, $p\lim_{n\to\infty}$ denotes convergence in probability: we write $p\lim_{n\to\infty} y_n = a_n$ if, for all $\delta > 0$,

$$\lim_{n\to\infty} \Pr[|y_n - a_n| > \delta] = 0.$$

Comments

- Almost sure results of this type have been shown for equally-spaced observation times (Hannan [22]), and should also be obtainable for the jittered model.
- Condition (1) is used in Lemma 1(a) to study the stochastic order of

$$\sup_{v\in[0,\Omega]} |\sum_{j=1}^{n} \epsilon_j e^{ivt_j}|^2.$$

We obtained the order $o_p(n^2)$, but feel this could be improved. For equally-spaced sample times, assuming $\sigma^2 < \infty$ gives the order $O_p(n^{3/2})$ (Walker [48]), and assuming $\mathrm{E}|\epsilon_1|^r < \infty$ for some $r > 4$ gives the order $O_p(n \log n)$ (Whittle [51]).

- Condition (2) is known as *Cramér's condition*, and ensures that the limiting value of the periodogram at the true frequency w is larger than the periodogram at the pseudo-alias frequencies $\{\pm w + k2\pi,\ k \in \mathcal{Z}\}$. Any probability measure having a nontrivial, absolutely continuous component satisfies this condition (Bhattacharya & Rao [5],p.207).

Theorem 2 *Under the conditions of Theorem 1,* $[n^{1/2}(\hat{a}_n - a), n^{1/2}(\hat{b}_n - b), n^{3/2}(\hat{w}_n - w)]$ *converges in distribution to a multivariate normal distribution with zero mean and variance matrix* Σ *given by*

(a) *For* $w \neq \frac{k\pi}{2},\ k \in \mathcal{Z}^+$,

$$\Sigma = \left\{2\sigma^2 + \frac{1}{2}(a^2+b^2)\left(1 - |\phi(2w)|^2\right)\right\}\Gamma^{-1}, \tag{2.13}$$

(b) *For* $w = \frac{k\pi}{2},\ k = 1, 3, 5, \ldots$,

$$\Sigma = 2\sigma^2\Gamma^{-1} + \Gamma^{-1}C\begin{bmatrix} \frac{1}{2}\Upsilon & \frac{1}{4}\Upsilon \\ \frac{1}{4}\Upsilon & \frac{1}{6}\Upsilon \end{bmatrix}C^T\Gamma^{-1}. \tag{2.14}$$

Here, the matrices Γ, C, *and* Υ *are defined by*

$$\Gamma = \begin{bmatrix} 1 & 0 & \frac{b}{2} \\ 0 & 1 & -\frac{a}{2} \\ \frac{b}{2} & -\frac{a}{2} & \frac{a^2+b^2}{3} \end{bmatrix}, \quad C = \begin{bmatrix} -a & -b & 0 & 0 \\ b & -a & 0 & 0 \\ 0 & 0 & -2ab & a^2-b^2 \end{bmatrix},$$

$$\Upsilon = \begin{bmatrix} 1 + Re\phi(4w) - 2Re^2\phi(2w) & Im\phi(4w) - Re\phi(2w)Im\phi(2w) \\ Im\phi(4w) - Re\phi(2w)Im\phi(2w) & 1 - Re\phi(4w) - 2Im^2\phi(2w) \end{bmatrix}. \tag{2.15}$$

Comments

Substituting the explicit value of Γ^{-1} in (2.13) for case **(a)**, the asymptotic variance of the scaled frequency estimate $n^{3/2}\hat{w}_n$ is

$$\frac{24\sigma^2}{a^2+b^2} + 6\{1 - |\phi(2w)|^2\}, \tag{2.16}$$

and so the asymptotic standard error of $\hat{w}_n$ is of order $n^{-3/2}$. In addition to the usual $n^{-1/2}$ rate (due to averaging out of the observational noise), the additional order n in the precision is due to the lengthening of the series. For example, a signal of frequency $\frac{\pi}{2}$ goes through 2.5 cycles over a ten-day period, and 25 cycles over 100 days. If one tries to fit a signal of frequency $\frac{\pi}{2} + \delta$ to this, it would be out of phase by 2.5δ at 10 days, and by 25δ at 100 days. Thus estimates based on the longer estimation period should be 10 times more precise.

The comparable asymptotic variance matrix for the scaled periodogram estimate of frequency for equally-spaced times (derived by Walker [48] and Hannan [22]) is

$$\frac{24\sigma^2}{a^2 + b^2}, \tag{2.17}$$

and so the asymptotic variance for jittered sampling contains an additional term, $6\{1 - |\phi(2w)|^2\}$, which depends on the sampling distribution through the quantity $\phi(2w)$. [If we let $|\phi(2w)| \uparrow 1$, the sampling distribution approaches that of equally-spaced times and the two variances (2.16) and (2.17) agree.]

Case **(b)** is a little more difficult to interpret. The contribution to the variance from the observation times comes through the variance of the sums $\sum_{j=1}^{n} e^{i2wt_j}$ and $\sum_{j=1} t_j e^{i2wt_j}$. For $w \neq k\pi/2$, the variances of the real and imaginary parts of these are equal, but if $w = k\pi/2$, the symmetry in the formulae is disturbed, leading to the complicated form in (2.14).

Another model in which the observation times are random is a point process that is stationary and mixing (Isokawa [25] and Brillinger [11]). Denote the point process by $N(dt)$, the sampling rate by $\beta = \mathrm{E}[(0, 1]]$, and the spectral densities of $\epsilon(t)$ and $N(dt)$ by $f_{\epsilon\epsilon}(\lambda)$ and $f_{NN}(\lambda)$ respectively. Then the asymptotic variance of the scaled periodogram estimates is given by

$$\begin{aligned} \Sigma &= \{4\pi f_{\epsilon\epsilon}(w) + 4\pi\beta^{-2}(f_{\epsilon\epsilon} * f_{NN})(w) + \pi(a^2+b^2)\beta^{-2} f_{NN}(2w)\}\Gamma^{-1} \\ &+ 2\pi\beta^{-2} f_{NN}(0) \begin{bmatrix} 0 & 0 & 0 \\ 0 & a^2 & ab \\ 0 & ab & b^2 \end{bmatrix}. \end{aligned} \tag{2.18}$$

In the jittered cosine model, $\epsilon(t)$ is a white noise process with variance σ^2 which gives $f_{\epsilon\epsilon}(v) = \frac{\sigma^2}{2\pi}$. If the time scale of the sampling is made arbitrary (for example, by adding the same U[0,1] variable to all the times), then the jittered sampling is stationary with sampling rate $\beta = 1$. The remaining unknown term in (2.18) is the spectral density of the jittered sampling process, $f_{NN}(\lambda)$. The autocovariance function of the jittered process, $C_{NN}(u)$, is equal to one at $u = 0$, and otherwise

$$
\begin{aligned}
C_{NN}(u) &= \sum_{j=-\infty}^{\infty} \int h(t)h(t+u-j)\,dt - \int h(t)h(t+u)\,dt - 1 \\
&= p(u) - \int h(t)h(t+u)\,dt,
\end{aligned}
$$

where $p(u)$ is a periodic function with period one and integral zero over $(-\infty, \infty)$. Integrating $\frac{1}{2\pi} C_{NN}(u)$ against $e^{-i\lambda u}$, we find that the spectral density of the jitter process is

$$
f_{NN}(\lambda) = \frac{1}{2\pi}\{1 - |\phi(\lambda)|^2\} + \frac{1}{2\pi} \sum_{k \neq 0} \delta(\lambda - 2\pi k) \sum_{j=-\infty}^{\infty} |\phi(2\pi k)|^2,
$$

where $\delta()$ is the Dirac delta function, and so $f_{NN}(2\pi k)$ is infinite when $|\phi(2\pi k)| > 0$.

Substituting $f_{\epsilon\epsilon}$, f_{NN}, and β into (2.18), the asymptotic variance of the scaled estimates for $w \neq k\pi$ is

$$
\Sigma = \left\{ 2\sigma^2 + \frac{1}{2}(a^2 + b^2)(1 - |\phi(2w)|^2) + 2\sigma^2 \int f_{NN}(\lambda)\,d\lambda \right\} \Gamma^{-1}.
$$

Although the third term in this equation is infinite, the first two terms match the asymptotic variance in Theorem 2.

2.2.2 Consistency

This section contains the derivation of Theorem 1, which is modeled on Walker [48], pp. 25-27, the equivalent result for equally-spaced sample times.

We first need to show

$$
n(\hat{w}_n - w) = o_p(1). \tag{2.19}
$$

From the definition of the periodogram (2.6), and (2.2),

$$
I_n(v) = \frac{2}{n} \left| \sum_{j=1}^{n} e^{ivt_j} \{a\cos(wt_j) + b\sin(wt_j) + \epsilon_j\} \right|^2. \tag{2.20}
$$

Rewriting

$$
a\cos(wt_j) + b\sin(wt_j) = ce^{iwt_j} + c^* e^{-iwt_j},
$$

where $c = \frac{1}{2}(a - ib)$ and c^* is the complex conjugate of c, and defining

$$
A_n(v) = \sum_{j=1}^{n} \epsilon_j\, e^{ivt_j}, \quad B_n(v) = \sum_{j=1}^{n} e^{ivt_j}, \tag{2.21}
$$

we can expand (2.20) as

$$I_n(v) = \frac{2}{n}\left\{|A_n(v)|^2 + 2\,\mathrm{Re}\left[A_n(-v)\left\{cB_n(v+w) + c^*B_n(v-w)\right\}\right] + |cB_n(v+w) + c^*B_n(v-w)|^2\right\}. \tag{2.22}$$

We will look at the limiting behavior of $I_n(w)$ as $n \to \infty$, and compare it with that of

$$K(n,\lambda) \overset{def}{=} \sup_{|v-w|\geq n^{-1}\lambda} \{I_n(v)\}$$

where λ can be arbitrarily small. Consider the expansion (2.22) at $v = w$. As $\mathrm{E}\,|A_n(v)|^2 = n\,\sigma^2$, then

$$|A_n(v)|^2 = O_p(n). \tag{2.23}$$

This result and $|B_n(v)| \leq n$ gives

$$2\,\mathrm{Re}\left[A_n(-v)\left\{cB_n(v+w) + c^*B_n(v-w)\right\}\right] = O_p(n^{3/2}). \tag{2.24}$$

Regarding the third term in (2.22), note that

$$\mathrm{Var}[\mathrm{Re}B_n(w)] = O(n), \quad \mathrm{Var}[\mathrm{Im}B_n(w)] = O(n),$$

and so $B_n(2w) = O_p(n^{1/2})$. This with $B_n(0) = n$ gives

$$|cB_n(2w) + c^*B_n(0)|^2 = \frac{n^2}{4}(a^2+b^2) + O_p(n^{3/2}). \tag{2.25}$$

Substituting (2.23),(2.24), and (2.25) into (2.22), the limiting value of the periodogram at $v = w$ is

$$I_n(w) = \frac{n}{2}(a^2+b^2) + O_p(n^{1/2}). \tag{2.26}$$

Now we consider the behavior of $K(n,\lambda)$ as $n \to \infty$:

Lemma 1 *Consider the jittered cosine model of Section 2.1.*

(a) Under assumption (2.10)

$$\sup_{v\in[0,\Omega]} |A_n(v)|^2 = o_p(n^2). \tag{2.27}$$

(b) Under assumption (2.11)

$$p\lim_{n\to\infty} \sup_{|v-w|\geq n^{-1}\lambda} |cB_n(v+w) + c^*B_n(v-w)|^2 < \frac{n^2}{4}(a^2+b^2) \tag{2.28}$$

Using Lemma 1 and (2.22), we see that

$$p \lim_{n\to\infty} n^{-1} K(n,\lambda) < \frac{1}{2}(a^2+b^2).$$

So with (2.26) we obtain

$$\lim_{n\to\infty} P[K(n,\lambda) < I_n(w)] = 1$$

which, as λ can be arbitrarily small, is equivalent to (2.19).

The consistency of $(\hat{a}_n, \hat{b}_n, \hat{\sigma}_n^2)$ is shown, with only minor changes, in Walker [48], p.27. This completes the proof of Theorem 1.

Proof of Lemma 1.

(a) We need to show that for any constant $s > 0$,

$$\lim_{n\to\infty} P\left[\sup_{v\in[0,\Omega]} |A_n(v)|^2 > s\,n^2\right] = 0.$$

First we divide $[0,\Omega]$ into r_n subintervals

$$R_l = \left[\frac{l-1}{r_n}\Omega, \frac{l}{r_n}\Omega\right], \quad l = 1,\ldots,r_n,$$

and let $v_l = (l-\frac{1}{2})\Omega/r_n$, the center point of R_l. Then

$$P\left[\sup_{v\in[0,\Omega]} |A_n(v)|^2 > s\,n^2\right] \le r_n P\left[\sup_{v\in R_l} |A_n(v)|^2 > s\,n^2\right]. \tag{2.29}$$

Consider the derivative

$$\begin{aligned}\frac{d}{dv}|A_n(v)|^2 &= \frac{d}{dv}\sum_{j,k=1}^{n} \epsilon_j\epsilon_k\, e^{iv(t_j-t_k)} \\ &= \sum_{j,k=1}^{n} \epsilon_j\epsilon_k\, e^{iv(t_j-t_k)}\, i(t_j-t_k).\end{aligned}$$

Thus the absolute value of the derivative satisfies

$$\left|\frac{d}{dv}|A_n(v)|^2\right| \le (n+2\Delta)\left\{\sum_{j=1}^{n}|\epsilon_j|\right\}^2 \stackrel{def}{=} D_n,$$

noting that D_n is a random quantity not depending on v. Thus from (2.29), we obtain the further bound

$$\begin{aligned}P\left[\sup_{v\in[0,\Omega]} |A_n(v)|^2 > sn^2\right] &\le r_n\,P\left[|A_n(v_l)|^2 + \frac{\Omega}{2r_n}D_n > s\,n^2\right] \\ &\le r_n\,P\left[|A_n(v_l)|^2 > \frac{sn^2}{2}\right] + r_n\,P\left[D_n > \frac{sr_nn^2}{\Omega}\right]. \end{aligned}\tag{2.30}$$

Using the Markov Inequality and (2.10),

$$\begin{aligned} P\left[|A_n(v_l)|^2 > \frac{s\,n^2}{2}\right] &\leq \frac{4}{s^2n^4}\mathrm{E}[|A_n(v_l)|^4] \\ &\leq \frac{4}{s^2n^4}\{n\,\mathrm{E}[|\epsilon_1|^4] + n(n-1)\sigma^4\} \\ &\leq \frac{L}{s^2n^2}, \end{aligned} \tag{2.31}$$

for some finite constant L. Similarly,

$$\begin{aligned} P\left[D_n > \frac{sr_nn^2}{\Omega}\right] &= P\left[\left\{\sum_{j=1}^{n}|\epsilon_j|\right\}^2 \geq \frac{sr_nn^2}{\Omega(n+2\Delta)}\right] \\ &\leq \left(\frac{\Omega(n+2\Delta)}{sr_nn^2}\right)^{5/2}\mathrm{E}\left[\left\{\sum_{j=1}^{n}|\epsilon_j|\right\}^5\right] \\ &\leq M\left(\frac{n}{sr_n}\right)^{5/2} \end{aligned} \tag{2.32}$$

for some finite M.

Thus from (2.30), (2.31) and (2.32), and choosing $r_n = n^{9/5}$,

$$\begin{aligned} P\left[\sup_{v\in[0,\Omega]}|A_n(v)|^2 > s\,n^2\right] &\leq \frac{Lr_n}{s^2n^2} + \frac{Mn^{5/2}}{s^{5/2}r_n^{3/2}} \\ &\overset{n\to\infty}{\longrightarrow} 0, \end{aligned}$$

which establishes (2.27).

(b) To establish (2.28), we need to find $s < 1$ such that $\lim_{n\to\infty} T(n, \lambda, s) = 0$, where

$$T(n,\lambda,s) \overset{def}{=} P\left[\sup_{|v-w|\geq n^{-1}\lambda}|cB_n(v+w) + c^*B_n(v-w)|^2 > \frac{sn^2}{4}(a^2+b^2)\right].$$

Parallel to the approach in (a), we divide the interval $[n^{-1}\lambda, \Omega]$ into r_n equal length subintervals $\{R_l\}$ with center points $\{v_j\}$. The derivative of the function being maximized is

$$\begin{aligned} &\frac{d}{dv}|cB_n(v+w) + c^*B_n(v-w)|^2 \\ &\quad= \frac{d}{dv}\sum_{j,k=1}^{n} e^{iv(t_j-t_k)}\{ce^{ivt_j} + c^*e^{-ivt_j}\}\{c^*e^{-ivt_k} + ce^{ivt_k}\} \\ &\quad= \sum_{j,k=1}^{n} i(t_j-t_k)e^{iv(t_j-t_k)}\{ce^{ivt_j} + c^*e^{-ivt_j}\}\{c^*e^{-ivt_k} + ce^{ivt_k}\}, \end{aligned}$$

and so the absolute value of the derivative satisfies

$$\begin{aligned}\left|\frac{d}{dv}|cB_n(v+w)+c^*B_n(v-w)|^2\right| &\leq \sum_{j,k=1}^{n}|t_j-t_k|(a^2+b^2)\\ &\leq (n+2\Delta)n^2(a^2+b^2). \qquad (2.33)\end{aligned}$$

Thus by using the subintervals $\{R_l\}$ and the derivative bound (2.33), we can bound $T(n,\lambda,s)$ by a sum of probabilities

$$\begin{aligned}T(n,\lambda,s) &\leq \sum_{l=1}^{r_n} P\left[\sup_{v\in R_l}|cB_n(v+w)+c^*B_n(v-w)|^2 > \frac{sn^2}{4}(a^2+b^2)\right]\\ &\leq \sum_{l=1}^{r_n} P\left[|cB_n(v_l+w)+c^*B_n(v_l-w)|^2 > \frac{n^2(a^2+b^2)}{4}\left\{s-\frac{2\Omega}{r_n}(n+2\Delta)\right\}\right].\end{aligned}$$

As $|cB_n(v+w)+c^*B_n(v-w)|^2 \leq \frac{1}{4}(a^2+b^2)\{|B_n(v+w)|+|B_n(v-w)|\}^2$, we can simplify the above to

$$T(n,\lambda,s) \leq \sum_{l=1}^{r_n} P\left[\{|\tfrac{1}{n}B_n(v+w)|+|\tfrac{1}{n}B_n(v-w)|\}^2 > s-\frac{2\Omega}{r_n}(n+2\Delta)\right]. \qquad (2.34)$$

Next we look at the behavior of $B_n(v)$. The mean of $B_n(v)$ is

$$\mathrm{E}\left[B_n(v)\right] = \phi(v)\sum_{j=1}^{n} e^{ivt_j} = \phi(v)\frac{e^{\frac{i}{2}(n+1)v}\sin(\frac{nv}{2})}{\sin(\frac{v}{2})}. \qquad (2.35)$$

In order to evaluate (2.34), we need to bound $|\frac{1}{n}\mathrm{E}B_n(v)|$. By (2.11) and (2.35), and as

$$|\sin(\frac{nv}{2})/n\sin(\frac{v}{2})| \leq 1,$$

there exists $\gamma_\pi < 1$ such that

$$\sup_{|v|>\pi}|\tfrac{1}{n}\mathrm{E}B_n(v)| < \gamma_\pi.$$

Also by Walker [48], p.26, for λ small enough so that $\{\sin(\lambda/2)/(\lambda/2)\}^2 > 1/\pi^2$,

$$\sup_{n^{-1}\lambda\leq|v|\leq\pi}|\tfrac{1}{n}\mathrm{E}B_n(v)| \leq \left|\frac{\sin(\frac{\lambda}{2})}{n\sin(\frac{\lambda}{2n})}\right|.$$

For all such λ, the function $\sin(\frac{\lambda}{2n})/(\frac{\lambda}{2n})$ is strictly increasing with n. Thus for all $n \geq 2$,

$$\left|\frac{\sin(\frac{\lambda}{2})}{n\sin(\frac{\lambda}{2n})}\right| \leq \left|\frac{\sin(\frac{\lambda}{2})}{2\sin(\frac{\lambda}{4})}\right| < 1.$$

So combining the above bounds and setting $\gamma_\lambda = \sup\{\gamma_\pi, |\sin(\frac{\lambda}{2})/2\sin(\frac{\lambda}{4})|\}$, we obtain

$$\sup_{|v|\geq n^{-1}\lambda} |\tfrac{1}{n}\mathrm{E}B_n(v)| < \gamma_\lambda < 1. \tag{2.36}$$

This is only a crude bound, as for $v \neq 2k\pi$ and $k \in \mathcal{Z}$, $\mathrm{E}[B_n(v)] = O(1)$. For arbitrarily small $\eta, \epsilon > 0$, and $n > n_* = \{\epsilon \sin(\eta/2)\}^{-1}$,

$$|\tfrac{1}{n}\mathrm{E}B_n(v)| < \epsilon, \quad |v - 2k\pi| > \eta, \quad k \in \mathcal{Z}. \tag{2.37}$$

As $w \neq k\pi, \quad k \in \mathcal{Z}^+$, $v + w$ and $v - w$ cannot simultaneously be integer multiples of 2π. Further, letting η satisfy $|w - k\pi| > \eta, \quad k \in \mathcal{Z}^+$, then from (2.37) either $|\frac{1}{n}\mathrm{E}B_n(v + w)| < \epsilon$ or $|\frac{1}{n}\mathrm{E}B_n(v - w)| < \epsilon$. Together with (2.36), this gives

$$|\tfrac{1}{n}\mathrm{E}B_n(v + w)| + |\tfrac{1}{n}\mathrm{E}B_n(v - w)| < \gamma_\lambda + \epsilon \quad \forall v. \tag{2.38}$$

Also, define $B_n^0(v) = B_n(v) - \mathrm{E}B_n(v)$, where $\mathrm{E}B_n(v) = \phi(v)\sum_{j=1}^n e^{ivj}$. Then, using the Markov Inequality,

$$\begin{aligned} P\left[|B_n^0(v)| > \beta n\right] &\leq \frac{\mathrm{E}[|B_n^0(v)|^4]}{n^4\beta^4} \\ &\leq \frac{1}{n^4\beta^4}\sum_{j,k=1}^{n} \mathrm{E}[|e^{ivt_j} - \phi(v)e^{ivj}|^2|e^{ivt_k} - \phi(v)e^{ivk}|^2] \\ &\leq \frac{16}{n^2\beta^4} \end{aligned} \tag{2.39}$$

Now we return to (2.34). Let us choose $\epsilon, \alpha > 0$ so that $s \stackrel{def}{=} (\gamma_\lambda + 3\epsilon)^2 + \alpha < 1$, and set $r_n = n\log n$. Then using (2.38) and (2.39),

$$\begin{aligned} \lim_{n\to\infty} T(n, \lambda, s) &\leq \lim_{n\to\infty}\sum_{l=1}^{r_n} P[|\tfrac{1}{n}B_n(v_l + w)| + |\tfrac{1}{n}B_n(v_l - w)| > \gamma_\lambda + 3\epsilon] \\ &\leq \lim_{n\to\infty}\sum_{l=1}^{r_n} P[|\tfrac{1}{n}B_n^0(v_l + w)| > \epsilon] + P[|\tfrac{1}{n}B_n^0(v_l - w)| > \epsilon] \\ &\leq \lim_{n\to\infty}\frac{32r_n}{n^2\epsilon^4} = 0, \end{aligned}$$

which completes Lemma 1(b).

□

2.2.3 Asymptotic Normality

In this section we provide the proof of Theorem 2. Following Walker [48], pp.28–31, we define

$$U_n(\alpha,\beta,v)=\sum_{j=1}^{n}y_j^2-2\sum_{j=1}^{n}y_j\{\alpha\cos(vt_j)+\beta\sin(vt_j)\}+\frac{n}{2}(\alpha^2+\beta^2).$$

This function is minimized at $\alpha=\hat{a}_n$, $\beta=\hat{b}_n$, and $v=\hat{w}_n$. Use of the mean value theorem yields

$$(U_n)_a=(U_n)_{a_na_n}(a-\hat{a}_n)+(U_n)_{a_nb_n}(b-\hat{b}_n)+(U_n)_{a_nw_n}(w-\hat{w}_n), \tag{2.40}$$

where $(U_n)_\alpha$ denotes $\partial U_n(\alpha,\beta,v)/\partial\alpha$, $(U_n)_{\alpha\beta}$ denotes $\partial^2U_n(\alpha,\beta,v)/\partial\alpha\partial\beta$, and so on, and (a_n,b_n,w_n) is some point on the line segment joining (a,b,w) and $(\hat{a}_n,\hat{b}_n,\hat{w}_n)$. Similarly

$$\begin{aligned}
(U_n)_b &= (U_n)_{b_na_n}(a-\hat{a}_n)+(U_n)_{b_nb_n}(b-\hat{b}_n)+(U_n)_{b_nw_n}(w-\hat{w}_n)\\
(U_n)_w &= (U_n)_{w_na_n}(a-\hat{a}_n)+(U_n)_{w_nb_n}(b-\hat{b}_n)+(U_n)_{w_nw_n}(w-\hat{w}_n).
\end{aligned} \tag{2.41}$$

Now we look at the first order partial derivatives of U_n. Firstly,

$$\begin{aligned}
(U_n)_a &= na-2\sum_{j=1}^{n}y_j\cos(wt_j)\\
&= na-2a\sum_{j=1}^{n}\cos(wt_j)^2-2b\sum_{j=1}^{n}\sin(wt_j)\cos(wt_j)-2\sum_{j=1}^{n}\epsilon_j\cos(wt_j)\\
&= -a\sum_{j=1}^{n}\cos(2wt_j)-b\sum_{j=1}^{n}\sin(2wt_j)-2\sum_{j=1}^{n}\epsilon_j\cos(wt_j),
\end{aligned} \tag{2.42}$$

and similarly

$$\begin{aligned}
(U_n)_b &= b\sum_{j=1}^{n}\cos(2wt_j)-a\sum_{j=1}^{n}\sin(2wt_j)-2\sum_{j=1}^{n}\epsilon_j\sin(wt_j),\\
(U_n)_w &= -2ab\sum_{j=1}^{n}t_j\cos(2wt_j)+(a^2-b^2)\sum_{j=1}^{n}t_j\sin(2wt_j)-2b\sum_{j=1}^{n}t_j\epsilon_j\cos(wt_j)\\
&\quad +2a\sum_{j=1}^{n}t_j\epsilon_j\sin(wt_j).
\end{aligned} \tag{2.43}$$

Defining

$$\begin{aligned}
V &= [V_1,V_2,V_3,V_4]^T\\
&= n^{-1/2}\sum_{j=1}^{n}[\cos(2wt_j),\sin(2wt_j),\epsilon_j\cos(wt_j),\epsilon_j\sin(wt_j)]^T,
\end{aligned}$$

$$
\begin{aligned}
W &= [W_1, W_2, W_3, W_4]^T \\
&= n^{-3/2}\sum_{j=1}^{n}[t_j\cos(2wt_j), t_j\sin(2wt_j), t_j\epsilon_j\cos(wt_j), t_j\epsilon_j\sin(wt_j)]^T,
\end{aligned}
$$

we can write (2.42) and (2.43) as

$$
\begin{aligned}
[n^{-1/2}(U_n)_a, n^{-1/2}(U_n)_b] &= \begin{bmatrix} -a & -b & -2 & 0 \\ b & -a & 0 & -2 \end{bmatrix} V \\
n^{-3/2}(U_n)_w &= \begin{bmatrix} -2ab & (a^2-b^2) & -2b & 2a \end{bmatrix} W. \qquad (2.44)
\end{aligned}
$$

We need to find the limiting distribution of (V, W), first calculating the means and covariances. The mean of V_1 is

$$
\begin{aligned}
\mathrm{E}[V_1] &= n^{-1/2}\sum_{j=1}^{n}\{\mathrm{Re}\phi(2w)\cos(2wj) - \mathrm{Im}\phi(2w)\sin(2wj)\} \\
&= O(n^{-1/2}),
\end{aligned}
$$

as $w \neq k\pi,\ k \in \mathcal{Z}^+$. Using similar working on the other elements of V, one can obtain

$$
\lim_{n\to\infty}\mathrm{E}[V] = [0,0,0,0]^T. \qquad (2.45)
$$

Consider the mean of W_1

$$
\begin{aligned}
\mathrm{E}[W_1] &= n^{-3/2}\sum_{j=1}^{n}\mathrm{Re}\,\mathrm{E}\left[t_j e^{i2wt_j}\right] \\
&= n^{-3/2}\sum_{j=1}^{n}\mathrm{Re}\{j\phi(2w)e^{i2wj} - \mathrm{E}[\delta_j e^{i2wt_j}]\} \\
&= n^{-3/2}\sum_{j=1}^{n}j\{\mathrm{Re}\phi(2w)\cos(2wj) - \mathrm{Im}\phi(2w)\sin(2wj)\} + O(n^{-1/2}) \\
&= O(n^{-1/2}),
\end{aligned}
$$

remembering that $|\delta_j| < \Delta$. Extending these ideas to the remaining elements of W gives

$$
\lim_{n\to\infty}\mathrm{E}[W] = [0,0,0,0]^T. \qquad (2.46)
$$

The (1,1) term of the variance matrix of V is

$$
\mathrm{Var}[V_1] = n^{-1}\sum_{j=1}^{n}\{\mathrm{E}[\cos^2(2wt_j)] - \mathrm{E}^2[\cos(2wt_j)]\}
$$

$$
\begin{aligned}
&= n^{-1}\sum_{j=1}^{n}\Big[\tfrac{1}{2}\{1+\mathrm{Re}\phi(4w)\cos(4wj)-\mathrm{Im}\phi(4w)\sin(4wj)\} \\
&\qquad\qquad -\{\mathrm{Re}\phi(2w)\cos(2wj)-\mathrm{Im}\phi(2w)\sin(2wj)\}^2\Big] \\
&= \tfrac{1}{2}\left(1-|\phi(2w)|^2\right)+\tfrac{1}{2n}\{\mathrm{Re}\phi(4w)-\mathrm{Re}^2\phi(2w)+\mathrm{Im}^2\phi(2w)\}\sum_{j=1}^{n}\cos(4wj) \\
&\quad +O(n^{-1}),
\end{aligned}
$$

and so

$$
\lim_{n\to\infty}\mathrm{Var}[V_1]=\begin{cases}\frac{1}{2}\left(1-|\phi(2w)|^2\right), & w\neq\frac{k\pi}{2}, k\in\mathcal{Z}^+,\\ \frac{1}{2}\Big(1+\mathrm{Re}\phi(4w)-2\mathrm{Re}^2\phi(2w)\Big), & w=\frac{k\pi}{2}, k=1,3,5,\ldots,\end{cases} \tag{2.47}
$$

Similarly, $\mathrm{Var}[W_1]$ can be calculated by

$$
\begin{aligned}
\mathrm{Var}[W_1] &= n^{-3}\sum_{j=1}^{n}\{\mathrm{E}[t_j\cos^2(2wt_j)]-\mathrm{E}^2[t_j\cos(2wt_j)]\} \\
&= n^{-3}\sum_{j=1}^{n}j^2\Big[\tfrac{1}{2}\{1+\mathrm{Re}\phi(4w)\cos(4wj)-\mathrm{Im}\phi(4w)\sin(4wj)\} \\
&\qquad\qquad -\{\mathrm{Re}\phi(2w)\cos(2wj)-\mathrm{Im}\phi(2w)\sin(2wj)\}^2\Big]+O(n^{-1}),
\end{aligned}
$$

so that

$$
\lim_{n\to\infty}\mathrm{Var}[W_1]=\begin{cases}\frac{1}{6}\left(1-|\phi(2w)|^2\right), & w\neq\frac{k\pi}{2}, k\in\mathcal{Z}^+,\\ \frac{1}{6}\Big(1+\mathrm{Re}\phi(4w)-2\mathrm{Re}^2\phi(2w)\Big), & w=\frac{k\pi}{2}, k=1,3,5,\ldots,\end{cases} \tag{2.48}
$$

Similar working applied to the other elements of the variance-covariance matrix of (V, W) yields $\mathrm{Var}[V]=\frac{1}{2}\Lambda$, $\mathrm{Var}[W]=\frac{1}{6}\Lambda$, and $\mathrm{Cov}[V,W]=\frac{1}{4}\Lambda$, where

$$
\Lambda=\begin{cases}\begin{bmatrix}\left(1-|\phi(2w)|^2\right)I_2 & 0\\ 0 & \sigma^2 I_2\end{bmatrix}, & w\neq\frac{k\pi}{2}, k\in\mathcal{Z}^+,\\ \begin{bmatrix}\Upsilon & 0\\ 0 & \sigma^2 I_2\end{bmatrix}, & w=\frac{k\pi}{2}, k=1,3,5,\ldots,\end{cases} \tag{2.49}
$$

I_2 is the 2×2 identity matrix, and Υ is defined in (2.15).
Now consider $\lambda^T V+\eta^T W$, where λ, η are vectors from $\Re^4$ such that $0<\lambda^T\lambda+\eta^T\eta<\infty$. Then

$$
\begin{aligned}
\lim_{n\to\infty}\mathrm{E}[\lambda^T V+\eta^T W] &= 0,\\
\lim_{n\to\infty}\mathrm{Var}[\lambda^T V+\eta^T W] &= \frac{1}{2}\lambda^T\Lambda\lambda+\frac{2}{4}\lambda^T\Lambda\eta+\frac{1}{6}\eta^T\Lambda\eta.
\end{aligned} \tag{2.50}
$$

Expressing $\lambda^T V + \eta^T W = \sum_{j=1}^n X_j$, where

$$\begin{aligned} X_j \quad &\stackrel{def}{=} \quad n^{-1/2}\{\lambda_1 \cos(2wt_j) + \lambda_2 \sin(2wt_j) + \lambda_3 \epsilon_j \cos(wt_j) + \lambda_4 \epsilon_j \sin(wt_j)\} \\ &\quad + n^{-3/2} t_j \{\eta_1 \cos(2wt_j) + \eta_2 \sin(2wt_j) + \eta_3 \epsilon_j \cos(wt_j) + \eta_4 \epsilon_j \sin(wt_j)\}, \end{aligned}$$

the sum of third absolute moments of $(X_j - \mathrm{E}X_j)$ satisfies

$$\sum_{j=1}^n \mathrm{E}|X_j - \mathrm{E}X_j|^3 \quad \leq \quad n^{-1/2}\mathrm{E}\,|L + M|\epsilon_1||^3$$

for some finite constants $L,\ M$. Thus the $\{X_j\}$ satisfy the Lyapounov condition (Billingsley [7], p.371),

$$\lim_{n\to\infty} \frac{\sum_{j=1}^n \mathrm{E}|X_j - \mathrm{E}X_j|^3}{\sum_{j=1}^n \mathrm{Var}[X_j]^{3/2}} = 0,$$

and $\lambda^T V + \eta^T W$ converges in distribution to a Normal distribution with mean and variance given in (2.50). As convergence in distribution of a vector-valued random variable is implied by convergence of all linear combinations of its terms (Billingsley [7], p.397), the joint distribution of V, W converges to

$$\mathrm{N}\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{2}\Lambda & \frac{1}{4}\Lambda \\ \frac{1}{4}\Lambda & \frac{1}{6}\Lambda \end{bmatrix} \right),$$

with Λ given in (2.49).

Returning to (2.44) and using the above result, after some calculation we find that

$$[n^{-\frac{1}{2}}(U_n)_a,\ n^{-\frac{1}{2}}(U_n)_b,\ n^{-\frac{3}{2}}(U_n)_w]$$

converges to a multivariate normal distribution with mean $[0, 0, 0]$ and variance

$$\begin{cases} \left\{2\sigma^2 + \frac{1}{2}(a^2+b^2)\left(1 - |\phi(2w)|^2\right)\right\}\Gamma, & w \neq \frac{k\pi}{2},\ k \in \mathcal{Z}^+, \\ 2\sigma^2\Gamma + C \begin{bmatrix} \frac{1}{2}\Upsilon & \frac{1}{4}\Upsilon \\ \frac{1}{4}\Upsilon & \frac{1}{6}\Upsilon \end{bmatrix} C^T, & w = \frac{k\pi}{2},\ k = 1, 3, 5, \ldots, \end{cases} \tag{2.51}$$

where $\Gamma,\ C$ is defined in (2.15).

Next consider the second order partial derivatives in (2.40) and (2.41). It is trivial to see that

$$(U_n)_{a_n a_n} = n,\ \ (U_n)_{b_n b_n} = n,\ \ (U_n)_{a_n b_n} = 0. \tag{2.52}$$

Also,

$$
\begin{aligned}
(U_n)_{a_n w_n} &= 2\sum_{j=1}^{n} y_j t_j \sin(w_n t_j) \\
&= a\sum_{j=1}^{n} t_j \sin((w_n + w)t_j) + a\sum_{j=1}^{n} t_j \sin((w_n - w)t_j) - b\sum_{j=1}^{n} t_j \cos((w_n + w)t_j) \\
&\quad + b\sum_{j=1}^{n} t_j \cos((w_n - w)t_j) + 2\sum_{j=1}^{n} \epsilon_j t_j \sin(w_n t_j). \qquad (2.53)
\end{aligned}
$$

We will apply the mean value theorem to the real and imaginary parts of $A_n'(v)$ and $B_n'(v)$, where $A_n(v)$, $B_n(v)$ are defined in (2.21). As $|A_n''(v)| = O_p(n^3)$ and $|B_n''(v)| = O(n^3)$, applying the mean value theorem gives

$$
\begin{aligned}
i\sum_{j=1}^{n} t_j e^{i(w_n - w)t_j} &= B_n'(0) + o_p(n^{-1})O(n^3) \\
&= \frac{in^2}{2} + o_p(n^2), \\
i\sum_{j=1}^{n} t_j e^{i(w_n + w)t_j} &= B_n'(2w) + o_p(n^{-1})O(n^3) \\
&= o_p(n^2), \\
i\sum_{j=1}^{n} \epsilon_j t_j e^{i w_n t_j} &= A_n'(w) + o_p(n^{-1})O(n^3) \\
&= o_p(n^2), \qquad (2.54)
\end{aligned}
$$

remembering that $w_n - w = o_p(n^{-1})$ from Theorem 1. Inserting these results into (2.53) yields

$$
p\lim_{n\to\infty} n^{-2}(U_n)_{a_n w_n} = \frac{b}{2}. \qquad (2.55)
$$

Similarly, as

$$
\begin{aligned}
(U_n)_{b_n w_n} &= -2\sum_{j=1}^{n} y_j t_j \cos(w_n t_j) \\
&= -b\sum_{j=1}^{n} t_j \sin((w_n + w)t_j) - b\sum_{j=1}^{n} t_j \sin((w_n - w)t_j) - a\sum_{j=1}^{n} t_j \cos((w_n + w)t_j) \\
&\quad - a\sum_{j=1}^{n} t_j \cos((w_n - w)t_j) - 2\sum_{j=1}^{n} \epsilon_j t_j \cos(w_n t_j),
\end{aligned}
$$

we obtain

$$
p\lim_{n\to\infty} n^{-2}(U_n)_{b_n w_n} = -\frac{a}{2}. \qquad (2.56)
$$

Now,

$$
\begin{aligned}
(U_n)_{w_nw_n} &= 2\sum_{j=1}^{n} y_j t_j^2\{a_n\cos(w_nt_j)+b_n\sin(w_nt_j)\} \\
&= (aa_n+bb_n)\sum_{j=1}^{n} t_j^2\cos((w_n+w)t_j)+(aa_n+bb_n)\sum_{j=1}^{n} t_j^2\cos((w_n-w)t_j) \\
&\quad +(a_nb+ab_n)\sum_{j=1}^{n} t_j^2\sin((w_n+w)t_j)+(ab_n-a_nb)\sum_{j=1}^{n} t_j^2\sin((w_n-w)t_j) \\
&\quad +2a_n\sum_{j=1}^{n}\epsilon_j t_j^2\cos(w_nt_j)+2b_n\sum_{j=1}^{n}\epsilon_j t_j^2\sin(w_nt_j). \qquad (2.57)
\end{aligned}
$$

Applying the mean value theorem to $A_n''(v)$ and $B_n''(v)$ as above gives

$$
\begin{aligned}
\sum_{j=1}^{n} t_j^2 e^{i(w_n-w)t_j} &= \frac{n^3}{3}+o_p(n^3), \\
\sum_{j=1}^{n} t_j^2 e^{i(w_n+w)t_j} &= o_p(n^3), \\
\sum_{j=1}^{n} \epsilon_j t_j^2 e^{iw_nt_j} &= o_p(n^3),
\end{aligned}
$$

and so from (2.57), remembering that $a_n=a+o_p(1)$ and $b_n=b+o_p(1)$,

$$
p\lim_{n\to\infty} n^{-3}(U_n)_{w_nw_n}=\frac{a^2+b^2}{3}. \qquad (2.58)
$$

Thus if we define

$$
\Gamma_n=\begin{bmatrix} n^{-1}(U_n)_{a_na_n} & n^{-1}(U_n)_{a_nb_n} & n^{-2}(U_n)_{a_nw_n} \\ n^{-1}(U_n)_{b_na_n} & n^{-1}(U_n)_{b_nb_n} & n^{-2}(U_n)_{b_nw_n} \\ n^{-2}(U_n)_{w_na_n} & n^{-2}(U_n)_{w_nb_n} & n^{-3}(U_n)_{w_nw_n} \end{bmatrix},
$$

from (2.52),(2.55),(2.56) and (2.58), we obtain

$$
p\lim_{n\to\infty}\Gamma_n=\Gamma, \qquad (2.59)
$$

with Γ defined in (2.15). As Γ is nonsingular, Γ_n is also nonsingular with probability tending to 1 as $n\to\infty$. So for large n we can re-express (2.40) and (2.41) as

$$
\left[n^{\frac{1}{2}}(\hat{a}_n-a), n^{\frac{1}{2}}(\hat{b}_n-b), n^{\frac{3}{2}}(\hat{w}_n-w)\right]=-\left[n^{-\frac{1}{2}}(U_n)_a, n^{-\frac{1}{2}}(U_n)_b, n^{-\frac{3}{2}}(U_n)_w,\right]\Gamma_n^{-1}. \qquad (2.60)
$$

We showed that the limiting distribution of the vector on the right hand side of (2.60) is multivariate normal with zero mean and variance given in (2.51). So use of the Continuous Transformation Theorem (Barndorff-Nielsen & Cox [3]) with (2.59) shows

$$[n^{1/2}(\hat{a}_n - a),\ n^{1/2}(\hat{b}_n - b),\ n^{3/2}(\hat{w}_n - w)]$$

tends in distribution to a multivariate normal distribution with mean zero and variance given in (2.13) and (2.14). This completes the proof of Theorem 2.

2.2.4 Extensions

Theorems 1 and 2 can be extended in a number of natural ways: considering frequencies that are multiples of π, including a constant in the response model, and estimating multiple frequency components.

Frequency Multiple of π

Proposition 1 *Assume the conditions of Theorem 1, except set $w = k\pi$ for some integer k. Then*

(1) *if $w = 0$, $p\lim_{n\to\infty} n\hat{w}_n = 0$;*

(2) *otherwise, $p\lim_{n\to\infty} n\{\hat{w}_n \pmod{2\pi}\} = n\{w \pmod{2\pi}\}$.*

Proof.
(1) Following the proof of Theorem 1, the result equivalent to (2.26) is

$$I_n(0) = 2a^2 n + O_p(n^{1/2}).$$

Lemma 1(a) still applies, and Lemma 1(b) becomes

$$p\lim_{n\to\infty} \max_{|v|\geq n^{-1}\lambda} |cB_n(v) + c^* B_n(v)|^2 < 2a^2 n. \tag{2.61}$$

As the left hand side of this is

$$p\lim_{n\to\infty} \max_{|v|\geq n^{-1}\lambda} a^2 |B_n(v)|^2,$$

using (2.34),(2.36) and (2.39) gives (2.61) and completes the proof of (1).

Note that the estimates (2.8) are no longer consistent, in fact tending to $[2a, 0, \sigma^2 - a^2]$ in probability. If instead the estimates are obtained by regressing y_j on $\cos(\hat{w}_n t_j)$, $\sin(\hat{w}_n t_j)$ (where there is no estimate of b if $\hat{w}_n = 0$), then consistency is preserved. Hannan [22] shows that for equally-spaced times and $w = 0$, there exists finite n^* such that $\hat{w}_n = w$ for all $n > n^*$; this is because the periodogram has a local maximum exactly at $w = 0$. A result of this type may be obtainable for jittered sampling times.

(2) For the case $w = k\pi,\ k \neq 0$, the periodogram has the value (as $n \to \infty$)

$$I_n(w) = 2n|c\phi(2k\pi) + c^*|^2 + O_p(n^{1/2}). \tag{2.62}$$

The periodogram is also order n at the frequencies $w \pm k2\pi,\ k \in \mathcal{Z}$. Without loss of generality, assume that w is an even multiple of π, $w = k2\pi$. Then at $v = l2\pi$,

$$I_n(v) = 2n|c\phi((l+k)2\pi) + c^*\phi((l-k)2\pi)|^2 + O_p(n^{1/2}), \tag{2.63}$$

which depending on $(a, b, \phi(.))$ is not necessarily smaller than (2.62) asymptotically. So $\hat{w}_n$ is not always consistent, but using the methods in the proof of Lemma 1(b), it is possible to show

$$p \lim_{n\to\infty} \max_{\substack{|v-l2\pi|\geq n^{-1}\lambda \\ l\in\mathcal{Z}}} |cB_n(v+w) + c^*B_n(v-w)|^2 < \max_{l\in\mathcal{Z}} n^2|c\phi((l+k)2\pi) + c^*\phi((l-k)2\pi)|^2,$$

which completes (2) for $w = k2\pi$. A similar approach is used for $w = (2k+1)\pi$. □

Subtracting out the series mean

Proposition 2 *Suppose that the conditions of Theorem 1 apply, except that the response model is*

$$y_j = m + a\cos(wt_j) + b\sin(wt_j) + \epsilon_j.$$

Define $\tilde{y}_j = y_j - \bar{y}$, $\bar{y} = \sum_{j=1}^n y_j$, and estimate the parameters $[w, a, b, \sigma^2]$ as before, except replacing y_j in the estimation formulae by $\tilde{y}_j$. Then the results of Theorems 1 and 2 apply to these estimates.

Proof. Let $\tilde{I}_n(v)$ be the periodogram for the mean-corrected data, i.e.

$$\tilde{I}_n(v) = \frac{2}{n}\left|\sum_{j=1}^n \tilde{y}_j e^{ivt_j}\right|^2.$$

Then the difference between the original and mean-corrected periodogram is

$$\tilde{I}_n(v) - I_n(v) = \frac{2}{n}|\bar{y}B_n(v)|^2 - \frac{4}{n}\text{Re}\left[\bar{y}B_n(v)\sum_{j=1}^n y_j e^{-ivt_j}\right].$$

As $I_n(v)$ is invariant under change of origin, m can be set to zero for simplicity. Now

$$\begin{aligned}\bar{y} &= n^{-1}\{cB_n(w) + c^*B_n(-w)\} + n^{-1}\sum_{j=1}^{n}\epsilon_j \\ &= O_p(n^{-1/2}),\end{aligned}$$

$B_n(v) = O(n)$, and $\sum_{j=1}^{n} y_j e^{-ivt_j} = O_p(n)$, where the $O(.)$ are uniform for $v \in [0, \Omega]$. Thus

$$\tilde{I}_n(v) - I_n(v) = O_p(n^{1/2})$$

uniformly, and the rest of the proof of consistency follows as in Theorem 1. Asymptotic normality is obtained using a similar approach on $\tilde{U}_n(a, b, w)$ (Walker [48],p.35). □

Several-Frequency Case

Instead of the simple harmonic model (2.2), consider the regression function

$$y_j = \sum_{l=1}^{p}\{a_l\cos(w_lt_j) + b_l\sin(w_lt_j)\} + \epsilon_j. \tag{2.64}$$

Quinn [33] and Wang [47] study estimation of the number of terms (p) for equally-spaced sample times. Under the jittered sampling model and fixed p, an estimation procedure can be constructed which gives consistent estimates of $[a_l, b_l, w_l]$ for $l = 1, \ldots, p$. Conditions (1) and (2) of Theorem 1 are required, as well as assuming the parameters come from the space defined by

$$a_1^2 + b_1^2 > a_2^2 + b_2^2 > \ldots > a_p^2 + b_p^2 > 0, \tag{2.65}$$

$$w_l \in [0, \Omega],\ w_l \pm w_m \neq k2\pi,\ \text{for } l, m \in 1, \ldots, p,\ k \in \mathcal{Z}. \tag{2.66}$$

Note that the latter includes the condition $w_l \neq k\pi,\ k \in \mathcal{Z}^+$. The periodogram (2.6) has the limiting behavior

$$I_n(v) = \frac{2}{n}\left|\sum_{l=1}^{p}\{c_lB_n(v + w_l) + c_l^*B_n(v - w_l)\}\right|^2 + O_p(n^{1/2}), \tag{2.67}$$

where $c_l = \frac{1}{2}(a_l - ib_l)$ parallels the notation in Theorem 1. Thus $I_n(v)$ is $O_p(n^{1/2})$ everywhere but near the frequencies

$$\pm w_l + k2\pi,\ l = 1, \ldots, p,\ k \in \mathcal{Z},$$

and the condition (2.66) ensures that not more than one of the terms in the sum in (2.67) can be of order n.

Let the estimate of w_1, $\hat{w}_{1,n}$, be the frequency in $[0, \Omega]$ that maximizes the periodogram. The condition (2.65) ensures that this tends towards w_1 instead of one of $w_2, \ldots, w_p$. In estimating the rest of the frequencies, the estimates $\hat{w}_{l,n}$ must not be too close together, otherwise they would all tend to the frequency with the largest amplitude. For equally-spaced observation times, the required condition is (Walker [48], p.32; Brillinger [11], p.282)

$$\lim_{n\to\infty} \min_{l,m\in 1\ldots,p} n|w_l - w_m| \to \infty.$$

For the jittered observation times, we also need to keep an estimate from being too close to the pseudo-alias frequencies of other estimates. So for $l = 2, \ldots, p$ define $\hat{w}_{l,n}$ as that frequency in

$$\{v : |v \pm \hat{w}_{m,n} + k2\pi| \geq cn^{-3/4},\ m = 1, \ldots, l-1,\ k \in \mathcal{Z}\}$$

that maximizes $I_n(v)$, where c is a positive constant. Also defining

$$\hat{a}_{l,n} = \frac{2}{n}\sum_{j=1}^{n} y_j \cos(\hat{w}_{l,n}t_j),\quad \hat{b}_{l,n} = \frac{2}{n}\sum_{j=1}^{n} y_j \sin(\hat{w}_{l,n}t_j),$$

the results of Theorems 1 and 2 apply to the estimates $\{[\hat{a}_{l,n}, \hat{b}_{l,n}, \hat{w}_{l,n}],\ l = 1, \ldots, p\}$.

Alternatively, equivalent results should be achievable by subtracting out the fitted signal at one frequency before estimating the next, although verification of this is not presented here. In other words, $\hat{w}_{1,n}$ is estimated as above, but for $l = 2, \ldots, p$, $\hat{w}_{l,n}$ is estimated by maximizing the periodogram calculated on the data

$$y_j^{l-1} = y_j - \sum_{m=1}^{l-1} \{\hat{a}_{l,n}\cos(\hat{w}_{l,n}) + \hat{b}_{l,n}\sin(\hat{w}_{l,n})\}.$$

2.3 Maximum Likelihood Estimate of Frequency

In this section we derive asymptotic properties of the maximum likelihood estimator of frequency for the simple harmonic model and randomly jittered sample times. The main results are presented in Section 2.3.1 and the derivation given in Section 2.3.2. We find that the maximum likelihood estimate is consistent for $w \neq k\pi$ and is asymptotically efficient with variance which is lower than that of the periodogram estimate. In Section 2.3.3 we discuss the case $w = k\pi$ and extend the asymptotic results to the multiple-frequency model.

Assume the simple harmonic model (2.2) with IID errors and randomly jittered sample times (2.4). Assume also that the distribution of the jitter variables δ_j does not depend on the parameters $[a, b, w, \sigma^2]$. Then assuming the probability densities of ϵ_j and δ_j are known and denoted by $g(\epsilon)$ and $h(\delta)$, and if we define $x_j = [y_j, t_j]$, the probability density of x_j is

$$f_j(x_j; \theta) = g(y_j - a\cos(wt_j) - b\sin(wt_j))\, h(t_j - j). \tag{2.68}$$

Then for θ in a suitable parameter space Θ, the *maximum likelihood estimate* (MLE) of θ, $\check{\theta}_n$, is any $\theta \in \Theta$ such that

$$\prod_{j=1}^{n} f_j(x_j; \theta) = \sup_{\tau \in \Theta} \prod_{j=1}^{n} f_j(x_j; \tau). \tag{2.69}$$

Under normality of the observational noise, maximizing the likelihood is equivalent to minimizing the residual sum of squares

$$\sum_{j=1}^{n} \{y_j - a\cos(wt_j) - b\sin(wt_j)\}^2.$$

Thus maximum likelihood theory can be used to study the behavior of the least squares estimate under normality. Other noise distributions lead to different estimates; for example, if the noise is assumed to come from the Laplace distribution

$$g(\epsilon) = \frac{1}{2} e^{-|\epsilon|/\sigma},$$

then the MLE is obtained by minimizing the sum of absolute residuals

$$\sum_{j=1}^{n} |y_j - a\cos(wt_j) - b\sin(wt_j)|.$$

2.3.1 Asymptotic Results

Theorem 3 *Consider the jittered cosine model of Section 2.1, with $\theta \in \Theta$*

$$\Theta = \{[a, b, w] : a^2 + b^2 \in (m, M),\ w \in (m, M),\ w \notin [k\pi - m, k\pi + m] \text{ for } k \in \mathcal{Z}^+\}, \tag{2.70}$$

where $m > 0$ can be arbitrarily small and M arbitrarily large. Also assume

(1) *For some $\rho > 0$, the probability density of ϵ_1, denoted $g(\epsilon)$, satisfies*

$$\int \left|\frac{g'}{g}(\epsilon)\right|^{2+\rho} g(\epsilon)\, d\epsilon < \infty; \quad \int \left[\frac{\partial^2}{\partial \epsilon^2} g^{1/2}(\epsilon)\right]^2 d\epsilon < \infty. \tag{2.71}$$

(2) *The characteristic function of δ_1, $\phi(.)$, satisfies*

$$\forall \eta > 0,\ \sup_{|t| > \eta} |\phi(t)| < 1. \tag{2.72}$$

Then the maximum likelihood estimate has the following properties:

(i) $\check{\theta}_n$ *is consistent.*

(ii) $\check{\theta}_n$ *is asymptotically normal with mean* θ *and covariance matrix given by*

$$Cov\,[N^{1/2}(\check{\theta}_n - \theta)] = 2I^{-1}\Gamma^{-1}, \tag{2.73}$$

where $N = diag[n^{1/2}, n^{1/2}, n^{3/2}]$, Γ *is defined in (2.15) and*

$$I = \int g'(\epsilon)^2 \frac{1}{g(\epsilon)}\, d\epsilon,$$

the information in density g*. Further, the convergence is uniform in* T*, any compact set in* Θ.

(iii) *There exists a matrix* Ψ_n *so that* $\check{\theta}_n$ *is asymptotically efficient for loss functions* $w(\Psi_n(\tau - \theta))$*, where* τ *is an estimate of* θ *and* $w(.)$ *is a loss function with a polynomial majorant.*

Comments

- The variance of the periodogram estimate is asymptotically larger than that of the least squares estimate. Under normality distributed noise, Theorem 3 shows that the asymptotic distribution of the scaled MLE (also the least squares estimate in this case) is given by

$$n^{3/2}\check{w}_n \Rightarrow_{n\to\infty} N\left(w, \frac{24\sigma^2}{(a^2 + b^2)}\right), \tag{2.74}$$

  where $\Rightarrow$ denotes convergence in distribution. This is the same asymptotic distribution as for equally-spaced times (Walker [48]). In comparison, we saw in Theorem 2 that the asymptotic distribution of the periodogram estimate of frequency is

$$n^{3/2}\hat{w}_n \Rightarrow_{n\to\infty} N\left(w, \frac{24\sigma^2}{(a^2 + b^2)} + 6\{1 - |\phi(2w)|^2\}\right). \tag{2.75}$$

  The additional term in the variance of the periodogram estimate depends on the jittered sampling distribution through the value $|\phi(2w)|$, which suggests that the periodogram estimate is sensitive to variability in the sample times, while the MLE is not.

- Under the conditions of the theorem, properties **(i)**–**(iii)** are also valid for certain Bayesian estimators; see Ibragimov & Has'minskii [24], p. 191 for more details.

- In the definition of Θ, the bounding of $a^2 + b^2$ away from 0 and ∞, and the bounding of w away from multiples of π is necessary to show uniformity of convergence.
- The requirements of condition **(1)** exclude discrete and partially continuous noise distributions, and ensure the smoothness of the likelihood surface.
- As in Theorem 1, condition **(2)** prevents the RSS at the pseudo-aliases $\pm w + k2\pi$ from being as small as that at the true frequency w.

2.3.2 Derivation of Results

The central tool in this proof is Theorem III.4.1 of Ibragimov & Has'minskii [24], p. 191. Denote Fisher's Information Matrix by

$$I_j(\theta) = \int \left[\tfrac{\partial}{\partial\theta} f_j(x_j;\theta)\right] \left[\tfrac{\partial}{\partial\theta} f_j(x_j;\theta)\right]^T \frac{dx_j}{f_j(x_j;\theta)}, \tag{2.76}$$

and the information from the first n observations by

$$\Psi^2(n,\theta) = \sum_{j=1}^{n} I_j(\theta). \tag{2.77}$$

Sometimes $\Psi_n^2 = \Psi^2(n,\theta)$ is used when the context is clear.

To make use of the result of Ibragimov & Has'minskii, Ψ_n^2 must be positive definite for $\theta \in \Theta$, and the following conditions satisfied:

(A) Denoting $|A| = \sup_{|\lambda|=1} |\lambda^T A \lambda|$,

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} \operatorname{tr} \Psi_n^{-2} = 0, \quad \lim_{n\to\infty} \sup_{\theta,\tau\in\Theta} |\Psi_n^{-1}\Psi^2(n,\tau)\Psi_n^{-1}| < \infty.$$

(B) For some $\rho > 0$ and $u \in \Re^3$,

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} \sum_{j=1}^{n} \mathrm{E} \left| \left[\tfrac{\partial}{\partial\theta} \ln f_j(x_j;\theta)\right]^T \Psi_n^{-1} u \right|^{2+\rho} = 0.$$

The next two conditions use a sequence $\lambda(n) \to \infty$ as $n \to \infty$:

(C) $f_j^{1/2}(x_j;\theta)$ is twice continuously differentiable with respect to θ, and as $n \to \infty$,

$$\sup_{\theta\in\Theta} \sup_{\substack{|u|<\lambda(n) \\ \theta+\Psi_n^{-1}u\in\Theta}} \sum_{j=1}^{n} \int \left| \Psi_n^{-1} \tfrac{\partial^2}{\partial\theta^2} f_j^{1/2}(x_j, \theta + \Psi_n^{-1}u) \Psi_n^{-1} \right|^2 dx_j = o\left(\lambda(n)^{-2}\right).$$

(D) For some $\beta > 0$,

$$\inf_{\theta\in\Theta} \inf_{\substack{|u|<\lambda(n) \\ \theta+\Psi_n^{-1}u\in\Theta}} |\Psi_n|^{-\beta} \sum_{j=1}^{n} \int \left[f_j^{1/2}(x_j;\theta+\Psi_n^{-1}u) - f_j^{1/2}(x_j;\theta) \right]^2 dx_j > 0.$$

(These correspond to relations (4.2), (4.8), (4.9) and (4.4) in Ibragimov & Has'minskii.)

In order to check these conditions, Ψ_n^2, Ψ_n^{-2}, Ψ_n, and Ψ_n^{-1} must first be calculated. From (2.76) and (2.68),

$$I_j(\theta) = I\,\mathrm{E}\begin{bmatrix} \cos^2(wt_j) & \cos(wt_j)\sin(wt_j) & t_j\cos(wt_j)\{b\cos(wt_j) - a\sin(wt_j)\} \\ * & \sin^2(wt_j) & t_j\sin(wt_j)\{b\cos(wt_j) - a\sin(wt_j)\} \\ * & * & t_j^2\{b\cos(wt_j) - a\sin(wt_j)\}^2 \end{bmatrix},$$

where I is defined in the statement of the theorem. Thus the [1,1] element of Ψ_n^2 is

$$\begin{aligned} \left[\Psi_n^2\right]_{1,1} &= I\sum_{j=1}^{n} \mathrm{E}[\cos^2(wt_j)] \\ &= \frac{I}{2}\sum_{j=1}^{n}\{1 + \mathrm{Re}\phi(2w)\cos(2wj) - \mathrm{Im}\phi(2w)\sin(2wj)\} \\ &= \frac{n}{2}I + O(1) \end{aligned}$$

as $n \to \infty$, because $w \neq k\pi$. Note also that the $O(.)$ is uniform for $\theta \in \Theta$. Similar working for the other elements of Ψ_n^2 leads to

$$\Psi_n^2 = \frac{I}{2}\begin{bmatrix} n + O(1) & O(1) & \frac{b}{2}n^2 + O(n) \\ * & n + O(1) & -\frac{a}{2}n^2 + O(n) \\ * & * & \frac{a^2+b^2}{3}n^3 + O(n^2) \end{bmatrix},$$

which simplifies to

$$\Psi_n^2 = \frac{I}{2}N^{1/2}\left\{\Gamma + O(n^{-1})\right\}N^{1/2}, \tag{2.78}$$

where again the $O(.)$ are uniform for $\theta \in \Theta$. Note that Ψ_n^2 is positive definite for $a^2 + b^2 \neq 0$ and $w \neq k\pi$, and uniformly positive definite for $\theta \in \Theta$. This implies

$$\Psi_n^{-2} = 2I^{-1}N^{-1/2}\left\{\Gamma^{-1} + O(n^{-1})\right\}N^{-1/2}, \tag{2.79}$$

where as before the $O(.)$ are uniform for $\theta \in \Theta$. Some tedious calculations give the equivalent results for Ψ_n and Ψ_n^{-1}, the symmetric square roots of Ψ_n^2 and Ψ_n^{-2}:

$$\Psi_n = \frac{\sqrt{\frac{1}{8}In}}{a^2+b^2}\begin{bmatrix} 2a^2+b^2 + O(n^{-1}) & ab + O(n^{-1}) & b\sqrt{3a^2+3b^2} + O(n^{-1}) \\ * & a^2+2b^2 + O(n^{-1}) & -a\sqrt{3a^2+3b^2} + O(n^{-1}) \\ * & * & \frac{2n}{3^{1/2}}(a^2+b^2)^{3/2} + O(1) \end{bmatrix}, \tag{2.80}$$

$$\Psi_n^{-1} = \frac{\sqrt{\frac{2}{In}}}{a^2+b^2}\begin{bmatrix} a^2+2b^2+O(n^{-1}) & -ab+O(n^{-1}) & -3bn^{-1}+O(n^{-2}) \\ * & 2a^2+b^2+O(n^{-1}) & 3an^{-1}+O(n^{-2}) \\ * & * & n^{-1}\sqrt{3a^2+3b^2}+O(n^{-2}) \end{bmatrix}, \tag{2.81}$$

with $O(.)$ uniform for $\theta \in \Theta$.

Consider condition (A). Firstly,

$$\text{tr}\Psi_n^{-2} = \frac{2I^{-1}}{a^2+b^2}\{n^{-1}(5a^2+5b^2)+O(n^{-2})\},$$

and so $\lim_{n\to\infty}\sup_{\theta\in\Theta}\text{tr}\Psi_n^{-2} = 0$ as required. Secondly,

$$\Psi_n^{-1}\Psi^2(n,\tau)\Psi_n^{-1} = \frac{n^{-1}}{a^2+b^2}A\{\Gamma_\tau + O(n^{-1})\}A^T,$$

where

$$A = \begin{bmatrix} O(n^{1/2}) & O(n^{1/2}) & O(n^{1/2}) \\ O(n^{1/2}) & O(n^{1/2}) & O(n^{1/2}) \\ O(n^{-1/2}) & O(n^{-1/2}) & O(n^{1/2}) \end{bmatrix}.$$

So $|\Psi_n^{-1}\Psi^2(n,\tau)\Psi_n^{-1}| = O(1)$ uniformly for $\theta, \tau \in \Theta$. This completes condition (A).

For condition (B), consider

$$\tfrac{\partial}{\partial\theta}\ln f_j(x_j;\theta) = \frac{g'}{g}(y_j - a\cos(wt_j) - b\sin(wt_j))\begin{bmatrix} -\cos(wt_j) \\ -\sin(wt_j) \\ at_j\sin(wt_j) - bt_j\cos(wt_j) \end{bmatrix},$$

and so

$$\begin{aligned} \mathrm{E}\left|\left[\tfrac{\partial}{\partial\theta}\ln f_j(x;\theta)\right]^T\Psi_n^{-1}u\right|^{2+\rho} &= \mathrm{E}\left|\frac{g'}{g}(\epsilon_1)\,O(n^{-1/2})\right|^{2+\rho} \\ &= Cn^{-1-\rho/2}\,\mathrm{E}\left|\frac{g'}{g}(\epsilon_1)\right|^{2+\rho} \end{aligned}$$

for some finite constant C, uniformly for $\theta \in \Theta$. Taking ρ to be the same as that in (2.71), condition (B) is established.

In condition (C), let $\tau = [\alpha, \beta, v]$ be a shorthand for $\theta + \Psi_n^{-1}u$. As

$$\tfrac{\partial^2}{\partial\tau^2}f_j^{1/2}(x_j;\tau) = h^{1/2}(t_j - j)\tfrac{\partial^2}{\partial\tau^2}g^{1/2}(y_j - \alpha\cos(vt_j) - \beta\sin(vt_j)),$$

the $[1,1]$ element is

$$\tfrac{\partial^2}{\partial\alpha^2} f_j^{1/2}(x_j;\tau) \quad = \quad h^{1/2}(t_j - j)\ddot{g}(\epsilon_j)\cos^2(vt_j),$$

using the notation $\dot{g} = \frac{\partial}{\partial\epsilon} g^{1/2}(\epsilon)$, $\ddot{g} = \frac{\partial^2}{\partial\epsilon^2} g^{1/2}(\epsilon)$. Similar working on the other elements of $\frac{\partial^2}{\partial\tau^2} f_j^{1/2}(x_j;\tau)$ yields

$$\tfrac{\partial^2}{\partial\tau^2} f_j^{1/2}(x_j;\tau) = h^{1/2}(t_j - j)O(1)\begin{bmatrix} \ddot{g}(\epsilon_j) & \ddot{g}(\epsilon_j) & n\{\dot{g}(\epsilon_j) + \ddot{g}(\epsilon_j)\} \\ * & \ddot{g}(\epsilon_j) & n\{\dot{g}(\epsilon_j) + \ddot{g}(\epsilon_j)\} \\ * & * & n^2\{\dot{g}(\epsilon_j) + \ddot{g}(\epsilon_j)\} \end{bmatrix},$$

and thus

$$\left|\mathbf{\Psi}_n^{-1} \tfrac{\partial^2}{\partial\tau^2} f_j^{1/2}(x_j,\tau)\mathbf{\Psi}_n^{-1}\right|^2 \le h(t_j - j)\,|\dot{g}(\epsilon_j) + \ddot{g}(\epsilon_j)|^2\, O(n^{-2}).$$

So for all $\theta, \tau \in \Theta$, there exists a constant C such that

$$\begin{aligned} \sum_{j=1}^n \int \left|\mathbf{\Psi}_n^{-1} \tfrac{\partial^2}{\partial\theta^2} f_j^{1/2}(x_j;\tau)\mathbf{\Psi}_n^{-1}\right|^2 dx_j \quad &\le \quad \frac{C}{n}\int h(t_j - j)\,|\dot{g}(\epsilon_j) + \ddot{g}(\epsilon_j)|^2 dx_j \\ &\le \quad \frac{C}{n}\int |\ddot{g}(\epsilon)|^2 + |\dot{g}(\epsilon)|^2\, d\epsilon. \end{aligned}$$

So taking $\lambda(n) = n^{1/2}$ and using (2.71), condition (C) is verified.

In condition (D), the summation term can be rewritten as

$$\sum_{j=1}^n \int h(t_j - j)\int \left\{g^{1/2}(y_j - s(t_j;\tau)) - g^{1/2}(y_j - s(t_j;\theta))\right\}^2 d(y_j|t_j)\, dt_j,$$

where $\tau = \theta + \mathbf{\Psi}_n^{-1}u$ and $s(t_j;\theta) = a\cos(wt_j) + b\sin(wt_j)$. As $s(t_j;\theta)$ is bounded for $\theta \in \Theta$, by Ibragimov & Has'minskii [24], p.195, there exists a constant C uniform for $\theta, \tau \in \Theta$ such that

$$\int \left\{g^{1/2}(y_j - s(t_j;\tau)) - g^{1/2}(y_j - s(t_j;\theta))\right\}^2 d(y_j|t_j) \ge C\{s(t_j;\tau) - s(t_j;\theta)\}^2.$$

Thus condition (D) can be verified by showing

$$\inf_{\theta\in\Theta} \inf_{\substack{|u|<\lambda(n) \\ \theta+\mathbf{\Psi}_n^{-1}u\in\Theta}} |\mathbf{\Psi}_n|^{-\beta} \sum_{j=1}^n \mathrm{E}\left[\{s(t_j;\tau) - s(t_j;\theta)\}^2\right] > 0.$$

Further, as $|\mathbf{\Psi}_n|^{-\beta} = O(n^{-3/2\beta})$, the above is equivalent to showing for some $\beta > 0$,

$$\inf_{\theta\in\Theta} \inf_{\substack{|u|<\lambda(n) \\ \theta+\mathbf{\Psi}_n^{-1}u\in\Theta}} n^{-\beta} \sum_{j=1}^n \mathrm{E}\left[\{s(t_j;\tau) - s(t_j;\theta)\}^2\right] > 0.$$

Looking at (2.81), $\theta + \Psi_n^{-1}u$ can be replaced by $[a + n^{-1/2}u_1, b + n^{-1/2}u_2, w + n^{-3/2}u_3]$, and so condition (D) can be verified by showing

$$\inf_{\theta\in\Theta} \inf_{\substack{|u|<\lambda(n) \\ \theta+N^{1/2}u\in\Theta}} n^{-\beta} \sum_{j=1}^{n} \mathrm{E}\left[\{s(t_j;\tau) - s(t_j;\theta)\}^2\right] > 0. \tag{2.82}$$

Lemma 2 *Let $t_j,\ j = 1, \ldots, n$ be distributed according to randomly jittered sampling (2.4), ϵ be a number in $(0, 1)$, and define*

$$A_n = \{\theta = [a, b, w] \in \Theta,\ \tau = [\alpha, \beta, v] \in \Theta : |a - \alpha| + |b - \beta| + n|w - v| > n^{-\epsilon}\}.$$

Then there exists a finite constant C for which

$$\inf_{A_n} \sum_{j=1}^{n} E\left[\{a\cos(wt_j) + b\sin(wt_j) - \alpha\cos(vt_j) - \beta\sin(vt_j)\}^2\right] \geq Cn^{1-2\epsilon}. \tag{2.83}$$

Setting $\lambda(n) = n^{-\epsilon+1/2}$ in Lemma 2, with $0 < \epsilon < \frac{1}{2}$ and $\beta < 1 - 2\epsilon$, condition (D) is verified. Thus from Theorem 4.1 of Ibragimov & Has'minskii, $\check{\theta}_n$ is consistent and $\Psi_n(\check{\theta}_n - \theta)$ is asymptotically normal with mean zero and unit variance matrix. It is apparent from (2.80) that the asymptotic normality result is not changed by replacing Ψ_n with

$$\frac{\sqrt{\frac{1}{8}In}}{a^2 + b^2} \begin{bmatrix} 2a^2{+}b^2 & ab & b\sqrt{3a^2{+}3b^2} \\ * & a^2{+}2b^2 & -a\sqrt{3a^2{+}3b^2} \\ * & * & \frac{2n}{3^{1/2}}(a^2{+}b^2)^{3/2} \end{bmatrix}.$$

This gives the variance matrix (2.73), and completes Theorem 3.

Proof of Lemma 2. Define

$$\begin{aligned} H(\theta,\tau) &= \sum_{j=1}^{n} \mathrm{E}\left[\{a\cos(wt_j) + b\sin(wt_j) - \alpha\cos(vt_j) - \beta\sin(vt_j)\}^2\right] \\ &= \frac{1}{4}\sum_{j=1}^{n} \mathrm{E}\left|\left(ce^{iwt_j} + c^*e^{-iwt_j}\right) - \left(\gamma e^{ivt_j} + \gamma^* e^{-ivt_j}\right)\right|^2, \end{aligned}$$

in which $c = a + bi$ and $\gamma = \alpha + \beta i$. Having $\theta, \tau \in \Theta$ means that $2w$ and $2v$ are bounded away from $2k\pi$. Hence with some working, $H(\theta, \tau)$ can be rewritten as

$$H(\theta,\tau) = \frac{n}{2}\left(|c|^2 + |\gamma|^2\right) - \mathrm{Re}\left[c\gamma^*\phi(w - v)\sum_{j=1}^{n} e^{i(w-v)j} + c\gamma\phi(w + v)\sum_{j=1}^{n} e^{i(w+v)j}\right] + O(1), \tag{2.84}$$

where the $O(1)$ is uniform for $\theta, \tau \in A_n$.

Consider the following cases in A_n:

(i) $|w-v| < \frac{1}{2}n^{-1-\epsilon}$,

(ii) $\frac{1}{2}n^{-1-\epsilon} < |w-v| < m$,

(iii) $|w-v-k2\pi| < m, \ \ k = \pm 1, \pm 2, \ldots$,

(iv) $|w+v-k2\pi| < m, \ \ k = \pm 1, \pm 2, \ldots$,

(v) all other w, v in A_n.

Under case **(i)**, $\sum_{j=1}^{n} e^{i(w-v)j} = n + O(1)$ and $\phi(w-v) = 1 + O(n^{-1-\epsilon})$ (Billingsley [7], p. 354). As in this case $w+v$ is bounded away from $k2\pi$, $\sum_{j=1}^{n} e^{i(w+v)j} = O(1)$ So from (2.84)

$$\begin{aligned} H(\theta, \tau) &= \frac{n}{2}(a^2 + b^2 + \alpha^2 + \beta^2) - n\mathrm{Re}[c\gamma^*] + O(1) \\ &= \frac{n}{2}\{(a-\alpha)^2 + (b-\beta)^2\} + O(1). \end{aligned}$$

As $|a-\alpha| + |b-\beta| > \frac{1}{2}n^{-\epsilon}$, one of $|a-\alpha|, |b-\beta|$ must exceed $\frac{1}{4}n^{-\epsilon}$. Hence

$$H(\theta, \tau) \geq \frac{1}{32}n^{1-2\epsilon} + O(1), \tag{2.85}$$

where the $O(.)$ is uniform for $\theta, \tau \in A_n$.

The remaining four cases use the bound (from (2.84))

$$H(\theta, \tau) \geq \frac{n}{2}\left(|c|^2 + |\gamma|^2\right) - \left|c\gamma^*\phi(w-v)\sum_{j=1}^{n} e^{i(w-v)j}\right| - \left|c\gamma\phi(w+v)\sum_{j=1}^{n} e^{i(w+v)j}\right| + O(1). \tag{2.86}$$

For case **(ii)**, $w+v$ is again bounded away from $k2\pi$, and so $\sum_{j=1}^{n} e^{i(w+v)j} = O(1)$. Also,

$$\sup_{\frac{1}{2}n^{-1-\epsilon} < |v| < m} \left|\sum_{j=1}^{n} e^{ivj}\right| = \left|\sum_{j=1}^{n} e^{\frac{1}{2}in^{-1-\epsilon}j}\right|.$$

Using a Taylor series argument, this is bounded by $n(1 - \frac{1}{48}n^{-2\epsilon})$ for large n. So substituting this into (2.86),

$$H(\theta, \tau) \geq \frac{n}{2}\left(|c|^2 + |\gamma|^2\right) - |c||\gamma|n(1 - \tfrac{1}{48}n^{-2\epsilon}) + O(1),$$

and completion of the square for $|\gamma|$ gives

$$H(\theta, \tau) \geq \frac{1}{96}|c|^2 n^{1-2\epsilon} + O(1). \tag{2.87}$$

For case **(iii)**, $w+v$ is bounded away from $k2\pi$, so

$$\begin{aligned} H(\theta,\tau) &\geq \frac{n}{2}\left(|c|^2+|\gamma|^2-2|c||\gamma||\phi(w-v)|\right)+O(1) \\ &\geq \frac{n}{2}|c|^2\left\{1-|\phi(w-v)|^2\right\}+O(1), \end{aligned}$$

completing the square as before. By (2.72), $|\phi(w-v)|$ is uniformly bounded away from 1, and so there exists a constant C such that $H(\theta,\tau)\geq Cn+O(1)$. Case **(iv)** is done in the same way. In case **(v)** both $w+v$ and $w-v$ are bounded away from $k2\pi$ and so

$$H(\theta,\tau)=\frac{n}{2}\left(|c|^2+|\gamma|^2\right)+O(1).$$

So combining cases **(i)**–**(v)** completes the lemma. □

2.3.3 Extensions

The model in Theorem 3 can be extended by considering frequencies that are multiples of π and by allowing the model to contain several periodic components.

Frequency Multiple of π

The methods of Theorem 3 are not directly applicable for $w=k\pi$, because the limiting value of $\mathbf{\Psi}(n,\theta)$ is not continuous at these points. For these values, the information of the first n observations is described by

$$N^{-\frac{1}{2}}\mathbf{\Psi}_n^2N^{-\frac{1}{2}}=\frac{I}{2}\begin{bmatrix} 1+\phi_R(2w) & \phi_I(2w) & \frac{b}{2}(1+\phi_R(2w))-\frac{a}{2}\phi_I(2w) \\ * & 1-\phi_R(2w) & -\frac{a}{2}(1-\phi_R(2w))+\frac{b}{2}\phi_I(2w) \\ * & * & \frac{a^2}{3}(1-\phi_R(2w))+\frac{b^2}{3}(1+\phi_R(2w)) \end{bmatrix}+O(1), \tag{2.88}$$

where $\phi_R(.)$ and $\phi_I(.)$ are the real and imaginary parts of $\phi(.)$. If $\mathbf{\Psi}_n^2$ is nonsingular and the MLE is consistent, one would expect the variance of the estimates to tend towards $\mathbf{\Psi}_n^{-2}$. When $w=0$, b can be taken equal to zero without loss of generality, and the information matrix becomes

$$\mathbf{\Psi}_n^2=\frac{I}{2}\begin{bmatrix} n & 0 & 0 \\ * & 0 & 0 \\ * & * & 0 \end{bmatrix},$$

with no information on either b or w. This is not surprising for b, and for w, it could mean that the frequency estimate exhibits more rapid convergence of the type discussed in Hannan [22]. For nontrivial $w=k\pi$, condition **(2)** ensures that $\mathbf{\Psi}_n^2$ is invertible, and it would be interesting to know if the parameter estimates behave like those in Theorem 3.

Several-Frequency Case

Using the response model (2.64), results equivalent to Theorem 3 can be achieved for each of the vectors $[a_l, b_l, w_l]$. Assume conditions **(1)** and **(2)** of Theorem 3, and define the parameter vector $\theta \in \Theta$ by

$$\theta = [a_1, b_1, w_1, \ldots, a_p, b_p, w_p],$$

where the parameter space Θ is described by

$$q < a_1^2 + b_1^2,\ a_p^2 + b_p^2 < Q,\ a_{l-1}^2 + b_{l-1}^2 < a_l^2 + b_l^2 - q,\ l = 2, \ldots, p,$$

$$w_l \in (q, Q),\ w_l \pm w_m \in [2k\pi - 2q, 2k\pi + 2q],\ l, m \in 1, \ldots, p,\ k \in \mathcal{Z},$$

where $q > 0$ can be arbitrarily small and Q arbitrarily large.

Under the above conditions, the information from the first n observations is

$$\Psi_n^2 = \frac{I}{2} \begin{bmatrix} N^{\frac{1}{2}}(\Gamma_1 + O(\frac{1}{n}))N^{\frac{1}{2}} & N^{\frac{1}{2}}O(\frac{1}{n})N^{\frac{1}{2}} & \cdots & N^{\frac{1}{2}}O(\frac{1}{n})N^{\frac{1}{2}} \\ * & N^{\frac{1}{2}}(\Gamma_2 + O(\frac{1}{n}))N^{\frac{1}{2}} & \ddots & \vdots \\ * & * & \ddots & N^{\frac{1}{2}}O(\frac{1}{n})N^{\frac{1}{2}} \\ * & * & * & N^{\frac{1}{2}}(\Gamma_p + O(\frac{1}{n}))N^{\frac{1}{2}} \end{bmatrix},$$

where Γ_l is defined by

$$\Gamma_l = \begin{bmatrix} 1 & 0 & \frac{b_l}{2} \\ * & 1 & -\frac{a_l}{2} \\ * & * & \frac{a_l^2+b_l^2}{3} \end{bmatrix}.$$

Conditions **(A)**, **(B)** and **(D)** (from the proof of Theorem 3) are satisfied as before, and condition **(D)** is dependent on the following generalization of Lemma 2.

Lemma 2* *Consider the several-frequency model given above, and define for* $\epsilon \in (0, 1)$ *and* $\tau = [\alpha_1, \beta_1, v_1, \ldots, \alpha_p, \beta_p, v_p]$,

$$A_n = \left\{\theta, \tau \in \Theta :\ \sum_{l=1}^{p} \{|a_l - \alpha_l| + |b_l - \beta_l| + n|w_l - v_l|\} > n^{-\epsilon}\right\}.$$

Then there exists a finite constant C *so that*

$$\inf_{A_n} \sum_{j=1}^{n} E\left[\sum_{l=1}^{p}\{a_l \cos(w_l t_j) + b_l \sin(w_l t_j) - \alpha_l \cos(v_l t_j) - \beta_l \sin(v_l t_j)\}\right] \geq Cn^{1-2\epsilon}. \tag{2.89}$$

Defining $H(\theta, \tau)$ to be the outer summation term in (2.89), one can obtain

$$\begin{aligned} H(\theta,\tau) &= \frac{n}{2}\sum_{l=1}^{p}(|c_l|^2+|\gamma_l|^2) \\ &\quad - \sum_{l,m=1}^{p} \text{Re}\left[c_l\gamma_m^*\phi(w_l-v_m)\sum_{j=1}^{n}e^{i(w_l-v_m)j} + c_l\gamma_m\phi(w_l+v_m)\sum_{j=1}^{n}e^{i(w_l+v_m)j}\right]. \end{aligned} \tag{2.90}$$

The case $\sum_{l=1}^{p} n|w_l - v_l| < \frac{1}{2}n^{-\epsilon}$ works exactly the same as in the proof of Lemma 2. The ordering of the $\{a_l^2+b_l^2\}$ in the parameter space ensures that permuting the $\{w_l\}$ does not lead to an equivalent fit. Although there are $2p^2$ terms in the summation over l, m in (2.90), having $\theta, \tau \in \Theta$ ensures that only p of these can simultaneously grow faster than $O(1)$. [If a term involving w_l is growing with n, then all other terms involving w_l are $O(1)$; the same applies for the v_m.] Thus

$$H(\theta,\tau) \geq \frac{n}{2}\sum_{l=1}^{p}(|c_l|^2+|\gamma_l|^2) - \sum_{k=1}^{p}|c_{l_k}||\gamma_{m_k}||\phi(w_{l_k}\pm v_{m_k})|\left|\sum_{j=1}^{n}e^{i(w_{l_k}\pm v_{m_k})j}\right| + O(1),$$

where $\{l_k\}$ and $\{m_k\}$ are each permutations of $1, \ldots, p$. So case splitting as in Lemma 2, and for each pair $[w_{l_k}, v_{m_k}]$ either completing the square or using a Taylor expansion, Lemma 2* is completed.

Thus under regularity conditions, the MLE is consistent, asymptotically normal with mean θ and variance Ψ_n^{-2}, and asymptotically efficient for a suitable family of loss functions.

2.4 Simulation Study

In Sections 2.2 and 2.3, consistency results were shown for the jittered cosine model with $w \neq k\pi$, but for $w = k\pi$ the periodogram estimates were sometimes inconsistent and behavior of the MLE was not derived. In this section we present simulations using the jittered cosine model in order to investigate convergence of the estimates for frequencies at or near multiples of π, and, in cases for which the estimates are known to be consistent, in order to to compare the distribution of the estimates for finite samples with the asymptotic distribution. We also performed simulations using real observation times to see if the behavior of the estimates for these times was consistent with the behavior for the jittered sampling model.

Summary of Results

Using data generated from the jittered cosine model, we found that the finite sample distributions of the periodogram and maximum likelihood estimates were consistent with the asymptotic values for two frequencies that are not integer multiples of π and for two sample sizes. For three frequencies that are multiples of π, the MLE appeared to be consistent while the periodogram estimate was sometimes apparently inconsistent, tending towards a pseudo-alias frequency. In cases when the periodogram estimate was inconsistent, the range of true frequencies about a multiple of π for which the periodogram estimated the pseudo-alias instead of the true frequency drops in length with n at rate approximately n^{-1}.

In simulations using sample times from one of the MACHO light curves, the periodogram estimate of frequency was sometimes inconsistent for $w = k\pi$ while the MLE appeared to be consistent. For a frequency that is not a multiple of π, the maximum likelihood and periodogram estimates had similar variances, but the periodogram estimate had large bias. This suggests that the uncompetitive performance of the periodogram for random sample times comes from a bias that is caused by the unevenly-spaced times.

In general, the maximum likelihood estimate performed better than the periodogram estimate, with lower MSE and apparent consistency at the frequencies $w = k\pi$.

2.4.1 Times Sampled from the Jittered Model

Data Generation

Data were simulated from the jittered cosine model of Section 2.1, with the observational noise generated from a normal distribution with mean zero and variance σ^2; the MLE is the least squares estimate under this model. The variables $\{\delta_j\}$ were generated from a uniform distribution over $[-\frac{1}{6}, \frac{1}{6}]$; this model was chosen to mimic the approximately eight hours a day in which it is dark enough to collect astronomical data. For each given combination of parameter values $[a, b, w, \sigma^2]$, 100 datasets of size 200 observations were generated; the parameter estimates were calculated using the first 25, 50, 100 and 200 observations in turn. These sample sizes were chosen to span typical values for variable star data. The periodogram estimates were obtained by maximizing the periodogram over some range $[0, \Omega]$ as in Section 2.2. To obtain the least squares estimates, for a given frequency $v \in [0, \Omega]$ the values y_j were regressed on $\cos(vt_j), \sin(vt_j)$ to obtain estimates $\check{a}_{n,v}, \check{b}_{n,v}$; for $v = 0$ the regression was on a vector of ones and $\check{b}_{n,v}$ was taken to be zero. Then the residual sum of squares

$$\mathrm{SS}_n(v) = \sum_{j=1}^{n} \{y_j - \check{a}_{n,v}\cos(vt_j) - \check{b}_{n,v}\sin(vt_j)\}^2$$

was minimized for $v \in [0, \Omega]$, giving $\check{w}_n$, $\check{a}_n$ and $\check{b}_n$. The maximization of the periodogram and the minimization of the RSS are computationally difficult, and issues relating to this are discussed in Section 3.2. Maximization or minimization was done using the two-stage optimization scheme described in Section 3.2. In the first stage, the function is optimized over a grid of frequency values spaced $\frac{1}{20}n^{-1}$ apart on the interval $[0, \Omega]$ (where n is the number of observations used in the estimation). This gives rough estimates of the local optima of the function. The five best optima, $v_m,\ m = 1, \ldots, 5$, are further refined using grids spaced $\frac{1}{20{,}000}n^{-1}$ apart over $[v_m - \frac{1}{20}n^{-1}$, $v_m + \frac{1}{20}n^{-1}]$. These specifications correspond to the values $[a, b] = [0, \Omega]$, $r_1 = 20$, $r_2 = 20,000$, $M = 5$, and $\delta = \frac{1}{10}n^{-1}$ in the notation of Section 3.2. Optimization using a more computationally intensive method ($M = 20$, $\delta = n^{-1}$) for a limited number of examples gave identical estimates to those using the above method. This suggests that the frequency estimates using this method should be accurate within $\frac{1}{20{,}000}n^{-1}$.

Frequency not near a multiple of π

Asymptotic distributions of the parameter estimates were presented in Sections 2.2 and 2.3 for frequencies $w \neq k\pi$, but little is known about the distributions of the parameters for finite sample size. We performed simulations for a couple of frequencies to compare the observed distribution of the estimates with the asymptotic distribution.

Example 1. Frequency equal to 2.4π.

Simulations were run for a constant amplitude $a^2 + b^2 = 1$ and noise variance $\sigma^2 = 0.04$, but with five choices of phase, specified by $[a, b]$ equal to $[1, 0]$, $[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$, $[0, 1]$, $[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$ and $[-1, 0]$. Table 2.1 shows the bias divided by its standard error, variance, and MSE of the scaled estimates of frequency from maximum likelihood ($n^{3/2}\check{w}_n$) and the periodogram ($n^{3/2}\hat{w}_n$) for the five combinations of parameters. These values are displayed for estimates calculated at sample sizes of 50 and 200; the data in the "$n = 50$" row of a given model is a subset of the data in the '$n = 200$" row.

The variances of the maximum likelihood estimate agreed closely with the asymptotic variance. Using Theorem 3, the asymptotic variance of the scaled estimate $n^{3/2}\check{w}_n$ is 0.96 for each of the 5 parameter combinations. The sample variances ranged between 0.77 and 1.23. As a

Table 2.1: Bias, variance, and MSE of the scaled maximum likelihood and periodogram estimates of frequency, for 5 values of $[a, b]$ and sample sizes of 50 and 200 observations. The true frequency is $w = 2.4\pi$ and the noise variance is $\sigma^2 = 0.04$.

		Maximum Likelihood			Periodogram		
a, b	n	Bias/SE	Var.	MSE	Bias/SE	Var.	MSE
1,0	50	-0.33	1.23	1.23	0.72	5.85	5.88
	200	0.51	0.78	0.78	0.98	5.92	5.97
$\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}$	50	0.21	1.20	1.20	0.00	7.55	7.55
	200	-0.56	1.18	1.19	0.53	6.92	6.94
0,1	50	1.02	0.74	0.74	2.05	5.64	5.88
	200	0.95	1.13	1.14	0.14	6.84	6.84
$\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}$	50	-0.56	1.01	1.01	-0.95	6.37	6.42
	200	2.31	0.74	0.77	2.69	5.98	6.41
-1,0	50	1.27	1.05	1.06	0.78	7.12	7.16
	200	0.23	0.93	0.93	0.37	5.89	5.90

sample variance calculated over 100 values has approximately a $\tau^2\chi^2(99)$ distribution, where τ^2 is the unknown variance of the variates in the sample, the acceptance region of a 95%-level test of $\mathrm{H}_0 : \tau^2 = \tau_0^2$ versus $\mathrm{H}_a : \tau^2 \neq \tau_0^2$ is $[0.74\,\tau_0^2, 1.30\,\tau_0^2]$. For variance 0.96 this is [0.71,1.25], and so the 10 sample variances are consistent with the asymptotic variance for the maximum likelihood estimates. This verifies the $n^{-3/2}$ rate of convergence of the frequency estimate and suggests that the asymptotic covariance matrix can be used in the calculation of confidence intervals and tests of hypothesis. The biases in the frequency estimate were generally not large, resulting in the MSE being virtually identical to the sample variance. One of the biases was larger than ± 2 standard errors (where the standard error is calculated from the estimated variance), but as the standard error it corresponds to was unusually small, this does not seem a concern.

The asymptotic variance of the scaled periodogram estimate $n^{3/2}\hat{w}_n$ is 6.63, using Theorem 2, and the acceptance region of a 95%-level test of this variance against a two-sided alternative is [4.9,8.6]. All of the sample variances of the periodogram estimates fell in this interval, and so the observed variances are again consistent with the asymptotic value. A couple of the biases were larger than 2 standard errors, but again these correspond to unusually low variance estimates. $\square$

Table 2.2: Bias, variance, and MSE of the scaled maximum likelihood and periodogram estimates of frequency, for 2 values of $[a, b]$ and sample sizes of 50 and 200 observations. The true frequency is $w = 0.5\pi$ and the noise variance is $\sigma^2 = 0.04$.

		Maximum Likelihood			Periodogram		
a, b	n	Bias/SE	Var.	MSE	Bias/SE	Var.	MSE
1,0	50	-0.62	0.73	0.73	-0.02	1.54	1.54
	200	0.57	0.81	0.81	0.01	1.65	1.65
0,1	50	-0.77	0.89	0.89	0.39	2.24	2.24
	200	-0.46	0.66	0.66	0.16	1.67	1.67

Example 2. Frequency equal to $\frac{\pi}{2}$.

Simulations were run with $[a, b] = [1, 0]$ and $[0, 1]$ and noise variance $\sigma^2 = 0.04$, in order to compare the observed distribution with the asymptotic one when w is a multiple of $\frac{\pi}{2}$. Table 2.2 shows the bias, variance, and MSE of the scaled frequency estimates from maximum likelihood and the periodogram for the two parameter combinations and two sample sizes: 50 and 200 observations. The asymptotic variance of the scaled MLE of frequency is 0.96, and the acceptance region for testing this variance at the 95% level is [0.71,1.25]. One of the sample variances (0.66) fell outside these bounds and the other three look low on average. It is possible that the finite sample variances approach the asymptotic value more slowly at $w = \frac{\pi}{2}$, but care should be taken as this may be a spurious result of multiple tests. Nevertheless, the sample variances were of the correct order of magnitude and were close to the asymptotic values. None of the biases were large and so the MSEs were identical to the variances in this case.

The asymptotic variance of the scaled periodogram estimates is 2.00, calculated using Theorem 2. This gives acceptance region of [1.5,2.6], and all of the sample periodogram variances fell within these bounds. As in the MLE case, the biases were quite small. $\square$

Frequency Equal to a Multiple of π

Simulations were performed to demonstrate the inconsistency of the periodogram and to investigate the behavior of the least squares (MLE) estimator.

Example 3. Frequency equal to 2π.

Simulations were performed with $[a, b] = [1, 0]$ and $[0, 1]$ and noise variance $\sigma^2 =$

Table 2.3: Mean squared error of frequency estimates from maximum likelihood and the periodogram. The true frequency was $w = 2\pi$ and the MSE was calculated about the apparent limiting frequency, w_*. The figures in parentheses are the number of outlying points not included in the MSE calculation. Note: In the first column of MSE values, the omitted values were clustered about the value 0; in the third and fourth columns they were clustered about 4π and 6π respectively.

	Model with $[a, b] = [1, 0]$		Model with $[a, b] = [0, 1]$	
Observations	MLE $w_* = 2\pi$	PG $w_* = 0$	MLE $w_* = 2\pi$	PG $w_* = 4\pi$
25	1.35 e–4 (6)	0.00 e+0	3.23 e–5 (9)	3.42 e–4 (3)
50	1.61 e–5	0.00 e+0	5.66 e–6 (2)	4.12 e–5 (1)
100	1.32 e–6	0.00 e+0	6.28 e–7	4.59 e–6
200	1.90 e–7	0.00 e+0	8.56 e–8	5.62 e–7

.04. The frequency estimates were obtained by optimization of the periodogram or RSS over $v \in [0, 3.51\pi]$. Table 2.3 shows the mean squared error of the frequency estimates from the 100 datasets simulated for each combination of parameters. The MSE is calculated about the apparent limit in probability of the estimates as n becomes large, as the estimates did not always appear to converge to the true values. The frequency estimates are often clustered about a few distinct values. For example, Table 2.4 shows the sorted frequency estimates corresponding to the MSE in the third column of the first row of Table 2.3; most of the values are within 0.015 of the true frequency 2π, but nine are in the region of 4π. In all cases, outlying estimates were clustered about values that differed from the true frequency by a multiple of 2π. This is hardly surprising, as in the proof of Theorem 1 the periodogram is found to be of order n at the frequencies $\pm w \pm k\pi$ but of order $O_p(n^{1/2})$ elsewhere; equivalent results appear in the proof of Lemma 2 for the MLE case. As n gets large, these competing values must appear with decreasing frequency for the estimate to be consistent.

The MLE of frequency appears to be consistent for both values of $[a, b]$; the number of estimates at a competing frequency quickly drops to zero, and the MSE is tending to zero with order of approximately n^{-3}. Inspection of the estimates of $[a, b]$ also suggest convergence to the true values for least squares. In Section 2.3.3 it is suggested that if the least squares estimates are consistent, the limiting distribution of $N^{1/2}(\check{\theta} - \theta)$ could be normal with zero mean and variance matrix $\mathbf{\Psi}_n^{-2}$, where $\mathbf{\Psi}_n^2$ was given in (2.88). This would parallel the behavior proven for $w \neq k\pi$.

Table 2.4: 100 maximum likelihood estimates of frequency from samples of size 25, for the jittered cosine model with $w = 2\pi$ and $[a, b] = [0, 1]$.

6.270	6.270	6.272	6.272	6.272	6.272	6.273	6.273	6.274	6.274
6.274	6.275	6.275	6.275	6.276	6.276	6.277	6.277	6.277	6.277
6.278	6.278	6.278	6.278	6.278	6.279	6.279	6.280	6.280	6.280
6.280	6.281	6.281	6.281	6.281	6.281	6.281	6.281	6.281	6.282
6.282	6.282	6.282	6.282	6.282	6.283	6.283	6.283	6.284	6.284
6.284	6.284	6.284	6.284	6.284	6.284	6.284	6.284	6.284	6.284
6.285	6.285	6.285	6.285	6.285	6.285	6.285	6.286	6.286	6.287
6.287	6.287	6.287	6.287	6.287	6.287	6.287	6.288	6.288	6.288
6.288	6.288	6.289	6.289	6.289	6.290	6.290	6.292	6.292	6.294
6.296	12.535	12.538	12.547	12.553	12.553	12.555	12.568	12.582	12.594

All of the observed variances were consistent with the theoretical values from the inverse of the information matrix, and all but one of the observed means were consistent with the theoretical zero value.

The periodogram estimates did not appear to be consistent for this example. For $[a, b] = [1, 0]$, all of the frequency estimates were identically zero; this lack of variability may seem surprising, but should not as the periodogram had a local maximum exactly at zero. Why is the frequency estimate tending towards zero? From (2.63), the periodogram at zero has limiting value

$$I_n(0) = 2n|\phi(2\pi) + \phi(2\pi)|^2 + O_p(n^{1/2}),$$

so by substituting $\phi(2\pi) = 0.8270$, $\frac{1}{n}I_n(0)$ has a limiting value of 5.471. Similarly, using (2.62) and $\phi(4\pi) = 0.4135$, the limiting value of $\frac{1}{n}I_n(2\pi)$ is 3.996, which explains why the periodogram is maximized at $v = 0$ rather than $v = 2\pi$. The periodogram also seems inconsistent for $[a, b] = [0, 1]$; in this case the estimate was converging towards the value 4π. $\square$

Example 4. Frequency equal to π.

Table 2.5 shows the equivalent results for the model with $w = \pi$. For $[a, b] = [1, 0]$ the MLE of frequency converged rapidly to the true value. The variance of the observed estimates for $n = 200$ was consistent with the theoretical value given in (2.88), suggesting that the inverse information can be used as a variance estimate even when $w = k\pi$. The periodogram estimate of frequency also converged rapidly to the true value.

Table 2.5: Mean squared error of frequency estimates from maximum likelihood and the periodogram. The true frequency was $w = \pi$ and the MSE was calculated about the apparent limiting frequency, w_*. The figures in parentheses are the number of outlying points not included in the MSE calculation. Note: In the third column of MSE values, the omitted values are clustered about the values 3π and 5π, while those of the fourth column are clustered about 3π only.

	Model with $[a, b] = [1, 0]$		Model with $[a, b] = [0, 1]$	
	MLE	PG	MLE	PG
Observations	$w_* = \pi$	$w_* = \pi$	$w_* = \pi$	$w_* = 5\pi$
25	5.17 e–4	8.67 e–5	3.38 e–5 (50)	1.11 e–3 (44)
50	4.14 e–5	9.24 e–6	3.75 e–6 (42)	9.36 e–5 (45)
100	6.91 e–6	1.11 e–6	4.88 e–7 (35)	1.48 e–5 (38)
200	7.26 e–7	1.52 e–7	4.51 e–8 (20)	1.46 e–6 (38)

For $[a, b] = [0, 1]$ the MLE seems to be converging slowly to the true frequency: for $n = 200$, 20% of the estimates are clustered around 3π. The variance of the observed frequency estimates for $n = 200$ (calculated omitting the 20 outliers) was lower than that suggested by (2.88), but this could be due to bias from the omission of the outlying values. The periodogram frequency estimates are clustered mainly around $w_* = 5\pi$, but there are a group of estimates around the value 3π which do not become less common as n gets large. The expected limiting value of the periodogram at the frequencies π, 3π and 5π is $\frac{1}{n}I_n(3\pi) \to 0.342$, $\frac{1}{n}I_n(5\pi) \to 0.342$, and $\frac{1}{n}I_n(\pi) \to 0.060$, showing why the values 3π and 5π both appear with stable proportions as n gets large. □

Example 5. Zero Frequency.

In this example only $[a, b] = [1, 0]$ was used, as setting $b \neq 0$ makes no difference to the model. Both the maximum likelihood and periodogram estimates of frequency converged to the true value for $w = 0$. Two of the maximum likelihood estimates were clustered near 2π for $n = 25$, but otherwise all of the estimates of frequency were closely clustered around the true value. The MSE for the MLE was 5.67 e–4 for $n = 25$, 8.72 e–5 for $n = 50$, 1.91 e–5 for $n = 100$, and 2.80 e–6 for $n = 200$, which is approximately n^{-3} convergence. About 40% of the maximum likelihood estimates of frequency were identically zero for all sample sizes, suggesting that $\mathrm{SS}_n(v)$ often has a local minimum at $v = 0$. By contrast, the periodogram estimates of frequency were all identically zero for each sample size, illustrating rapid convergence of the type discussed in

Hannan (1971). Estimates of $[a, b]$ seemed consistent using maximum likelihood and inconsistent using the periodogram. $\square$

In summary, the least squares(maximum likelihood) estimate of frequency appeared consistent for the three frequencies used. There is no guarantee that this applies for all $w = k\pi$ and combinations of $[a, b]$, and it would be interesting to prove consistency for this case. From this limited investigation, it is plausible that least squares estimates that are consistent for $w = k\pi$ have asymptotic variance equal to the inverse of the information matrix of the observations. The periodogram was sometimes consistent and sometimes not, depending on the values of $[a, b]$, and although not illustrated here, also depending on the shape of the characteristic function $\phi(.)$.

Frequency Close to a Multiple of π

It has been shown above that the periodogram estimates are consistent for $w \neq k\pi$, and may be inconsistent for $w = k\pi$. In this section we discuss the behavior of the periodogram when the true frequency is close to $k\pi$.

The periodogram at frequency v can be expressed as

$$I_n(v) = \tfrac{2}{n}|c\mathrm{E}B_n(v + w) + c^*\mathrm{E}B_n(v - w)|^2 + O_p(n^{1/2}),$$

where the order term is uniform for $v \in [0, \Omega]$. This is shown by an argument similar to that preceding (2.26). As $I_n(v)$ is of order n at the true frequency, and the frequency estimate $\hat{w}_n$ is obtained by maximizing the periodogram, study of the maximum of

$$\begin{aligned} J_n(v) &= \tfrac{2}{n}|c\mathrm{E}B_n(v + w) + c^*\mathrm{E}B_n(v - w)|^2 \\ &= \tfrac{2}{n}\left|c\phi(v + w)\sum_{j=1}^{n} e^{i(v+w)j} + c^*\phi(v - w)e^{i(v-w)j}\right|^2 \end{aligned}$$

should give information on the behavior of $\hat{w}_n$ for large n. In Theorem 3 it was shown that the maximum likelihood estimate has uniform convergence if the true frequency is bounded away from the multiples of π. The same is true of the periodogram estimates, but a stronger result can be obtained: the estimates are uniformly consistent for a parameter space in which the frequency may approach one of the values $k\pi$ as $n \to \infty$. For a given n, consider the parameter space

$$\Theta_n = \{[a, b, w] : \ a^2 + b^2 \in (m, M), \ w \in [0, M], \ |w - k\pi| > \eta(n), \ \forall k \in \mathcal{Z}\},$$

in which $m > 0$ can be arbitrarily small, M can be arbitrarily large, and $\eta(n)$ tends to zero slower then n^{-1}, i.e. $\lim_{n\to\infty} n\eta(n) = \infty$. Inconsistency in the frequency estimate can arise when both of the terms in $J_n(v)$ are of order n; the proof of Theorem 1 shows that if only one of these terms can be of order n for a given v, then the frequency estimate is consistent. For $\theta \in \Theta_n$, if one of $v+w,\ v-w$ is closer than $\eta(n)$ from a multiple of 2π (a necessary but not sufficient condition for order n behavior), then the other is at least $\eta(n)$ away from all multiples of 2π. Without loss of generality, let $v-w$ be within $\eta(n)$ of a multiple of 2π. Then for large n,

$$\begin{aligned}\left|\sum_{j=1}^{n} e^{i(v+w)j}\right| &\leq \left|\frac{\sin(n(v+w)/2)}{n\sin((v+w)/2)}\right| \\ &\leq |\sin(\eta(n)/2)|^{-1},\end{aligned}$$

which by the definition of $\eta(n)$ is $o(n)$. So if one of the terms in $J_n(v)$ is of order n, the other is of order uniformly inferior to n. Thus the periodogram estimate of frequency is uniformly consistent for $\theta \in \Theta_n$, and uniform consistency for $[a,b]$ follows. This means that as n increases, one can estimate frequency closer and closer to multiples of π without sacrificing rate of convergence, as long as the frequency approaches $k\pi$ at a rate slower then n^{-1}.

For a given sample size and parameter values, it is informative to see how close the true frequency must be to $k\pi$ in order to cause problems with the estimation. This is illustrated by a couple of examples.

Example 6. Frequency near 2π.

In this example $[a,b] = [1,0]$ and the true frequency is $w = 2\pi(1+\eta)$, where η is small. If $\eta = 0$, the values of $\frac{1}{n}J_n(2\pi\eta)$ and $\frac{1}{n}J_n(2\pi(1+\eta))$ are 5.471 and 3.996, so we expect the periodogram frequency estimate to be in the neighborhood of zero if the variance of the periodogram is sufficiently small. If η is larger then zero then $\frac{1}{n}J_n(2\pi\eta)$ and $\frac{1}{n}J_n(2\pi(1+\eta))$ are $1.368 + O(n^{-1})$ and $2.000 + O(n^{-1})$, but the $O(.)$ term is not uniform for $\eta > 0$ and the convergence can be quite slow. If η is very close to zero, $\hat{w}_n$ should be in the vicinity of zero; if η is sufficiently large, $\hat{w}_n$ should be in the vicinity of 2π. How should we specify the change-point between the two? One way is to find a value of η for which the frequency estimate is close to $v = 2\pi$ and to $v = 0$ equally often. For the given model with sample size equal to 25, numerical computations find that the local maximum of $J_n(v)$ near $v = 0$ and the local maximum of $J_n(v)$ near $v = 2\pi$ are of equal height for $\eta = 0.01$. So for $\eta = 0.01$, one expects the frequency estimate to be in the vicinity of $v = 2\pi$ and $v = 0$ equally often, depending on the noise in $I_n(v)$. [The location of the local maximum does not have an explicit form, but as the function $B_n(v+w)$ is order n at $v = -2\pi\eta$, $B_n(v-w)$ is order n at $v = 2\pi\eta$, and $B_n(.)$ is smooth near its maxima, the local maximum of $J_n(v)$ near zero will usually fall in the interval $[-2\pi\eta, 2\pi\eta]$; similarly, the local maximum of $J_n(v)$ near $v = 2\pi$ usually falls in $[2\pi(1-\eta), 2\pi(1+\eta)]$.]

Table 2.6 shows the ratio of the local maximum of $J_n(v)$ near the true frequency $2\pi(1+\eta)$ to the local maximum of $J_n(v)$ near the competing frequency $v = 0$, for sample sizes $n = 25, 50, 100, 200$ and for four values of η. The first η value is 0.01 as mentioned above, and the rest decrease in factors of two; these were chosen to see how the ratio behaved if η changed like n^{-1} as $n \to \infty$. Along the diagonal of the table the ratios are approximately one. For larger values of η at a given sample size, the ratio is greater than one; for smaller values of η it is less than one. Thus the interval of frequencies for which the maximum of $J_n(v)$ is near 0 instead of near the true frequency decreases in length like n^{-1}. [The rate is not *exactly* n^{-1}; the values of η for which the ratio is exactly unity are 0.01 for $n = 25$, 0.051 for $n = 50$, 0.0026 for $n = 100$ and 0.013 for $n = 200$.] The values in the upper-right of the table are identical, but this is an artifact of the values of η and n chosen; for example, for $\eta = 0.01$ and $n = 75$ the values of the ratio is 1.432.

Table 2.7 shows the number of periodogram frequency estimates from 100 simulated datasets that were in the vicinity of the true frequency $w = 2\pi(1+\eta)$ instead of near zero. They were obtained for the same values of n and η used in Table 2.6; two levels of noise variance were used, $\sigma^2 = .04$ and $\sigma^2 = .16$. The combinations of n and η for which the ratio in Table 2.6 was less than one gave estimates that were all near zero (in fact, they were identically zero because the periodogram has a local maximum at $v = 0$). Similarly, the combinations for which the ratio was greater than one gave estimates that were almost all close to the true frequency. When the ratio was close to one, the frequency estimates were clustered about the true and competing frequencies. For $n = 25$ and $n = 50$, the proportion of estimates close to the true value was about one-half, but for the two larger values of n this decreased, to 44/100 for $n = 100$ and 37/100 for $n = 200$. There seem to be two explanations of this: for $n = 100, 200$, the ratio is smaller than for $n = 25, 50$; and as $n \to \infty$ the variability of the periodogram becomes small, so the maximization of the periodogram will be more sensitive to small differences in the height of $J_n(v)$ for different values of v. The results were consistent over the two levels of noise variance used. Overall, studying the maxima of $J_n(v)$ gives a clear indication of the behavior of the frequency estimate if the true model is known. □

Example 7. Frequency near π.

In this model, the true frequency is $w = \pi(1-\eta)$ for η small, and $[a, b] = [0.2, 1]$. The values of $\frac{1}{n}J_n(\pi)$ and $\frac{1}{n}J_n(3\pi)$ are 0.1634 and 0.2325, so when $\eta = 0$ the periodogram frequency estimates should converge to 3π instead of π. For η larger than zero, $\frac{1}{n}J_n(\pi(1-\eta))$ and $\frac{1}{n}J_n(3\pi - \pi\eta)$ are equal to $2.08 + O(n^{-1})$ and $1.4226 + O(n^{-1})$ respectively and the frequency

Table 2.6: Ratios of the local maximum of $J_n(v)$ near $v = 2\pi$ to the local maximum near $v = 0$, for the jittered cosine model with $w = 2\pi(1 + \eta)$, $[a, b] = [1, 0]$ and δ_1 uniformly distributed over $[-\frac{1}{6}, \frac{1}{6}]$.

	Ratios			
η	$n = 25$	$n = 50$	$n = 100$	$n = 200$
0.01	1.006	1.462	1.462	1.462
0.05	0.763	0.987	1.462	1.462
0.0025	0.737	0.762	0.978	1.462
0.00125	0.732	0.737	0.761	0.974

Table 2.7: Number of periodogram frequency estimates (out of a possible 100) in the vicinity of 2π, for data simulated from the jittered cosine model with $w = 2\pi(1 + \eta)$, $[a, b] = [1, 0]$ and δ_1 uniformly distributed over $[-\frac{1}{6}, \frac{1}{6}]$. The periodogram was maximized over $v \in [0, 3\pi]$, and two levels of observational noise variance were used.

		Number out of 100			
η	σ^2	$n = 25$	$n = 50$	$n = 100$	$n = 200$
0.01	0.04	53	99	100	100
0.05	0.04	0	50	100	100
0.0025	0.04	0	0	44	100
0.00125	0.04	0	0	0	37
0.01	0.16	55	97	100	100
0.05	0.16	0	52	99	100

estimate should converge to the true value. Table 2.8 shows the ratio of the local maximum of $J_n(v)$ near $v = \pi$ to the local maximum of $J_n(v)$ near $v = 3\pi$, for $n = 25, 50, 100, 200$ and four values of η decreasing by factors of two, chosen so that the ratios along the diagonal are close to one. As η or n increase the ratio becomes greater than one, and as they decrease it becomes less than one, but note that the increase or decrease is not monotonic in n and η. The corresponding number of frequency estimates that were in the vicinity of the true frequency are shown in Table 2.9. For each value of η, 100 datasets were generated from the model with noise variance $\sigma^2 = 0.04$, the periodogram frequency estimates were calculated over $v \in [0, 3.2\pi]$, and the number out of 100 that were close to $w = \pi(1-\eta)$ was reported. The results are consistent with those in Example (a): ratios substantially larger than one led to estimates all close to the true value; ratios less than one had all estimates near the competing frequency 3π (except for a few cases for $n = 25$, when the variability in the periodogram was larger); along the diagonal, the counts were close to 50%, and slightly larger than this when the ratio was slightly larger than one. So as in Example (a), the observed behavior of the frequency estimates from the periodogram agree with the behavior predicted by looking at $J_n(v)$. $\square$

In summary, the behavior of the periodogram estimate of frequency can be predicted for frequencies close to multiples of π if the parameters $[a, b, w]$ and the characteristic function $\phi(.)$ are known. In these two examples, the region of frequencies around a multiple of π for which the estimated frequency was not close to the true value decreased with order n^{-1}. This agrees with the theory shown above that the frequency can be uniformly estimated on a frequency space that approaches the multiples of π if the rate of approach is of order shower than n^{-1}.

2.4.2 Times from Real Life

Earlier in this chapter, consistency and asymptotic normality results were presented for the jittered cosine model which show that the periodogram can be inconsistent for $w = k\pi$ and is less efficient than maximum likelihood. However, the jittered observation model is unrealistic in several ways:

- There is exactly one observation per day, while in reality there might be multiple observations taken per night, or no data collected for a few days.
- The range of times each day in which observations are taken is constant over time, but the time period each day in which it is dark enough to collect data changes with the time of year.

Table 2.8: Ratios of the local maximum of $J_n(v)$ near $v = \pi$ to the local maximum near $v = 3\pi$, for the jittered cosine model with $w = \pi(1 - \eta)$, $[a, b] = [0.2, 1]$ and δ_1 uniformly distributed over $[-\frac{1}{6}, \frac{1}{6}]$.

	Ratio			
η	$n = 25$	$n = 50$	$n = 100$	$n = 200$
0.0062	1.051	1.721	1.400	1.421
0.0031	0.275	1.020	1.708	1.400
0.00155	0.271	0.268	1.005	1.701
0.000775	0.457	0.277	0.264	0.997

Table 2.9: Number of periodogram frequency estimates in the vicinity of π, for 100 datasets simulated from the jittered cosine model with $w = 2\pi(1 + \eta)$, $[a, b] = [1, 0]$, and $\delta_1 \sim U[-\frac{1}{6}, \frac{1}{6}]$. The noise variance was $\sigma^2 = 0.04$, and the periodogram was maximized over $v \in [0, 3.2\pi]$.

	Number out of 100			
η	$n = 25$	$n = 50$	$n = 100$	$n = 200$
0.0062	62	100	100	100
0.0031	5	65	100	100
0.00155	0	0	57	100
0.000755	7	0	0	49

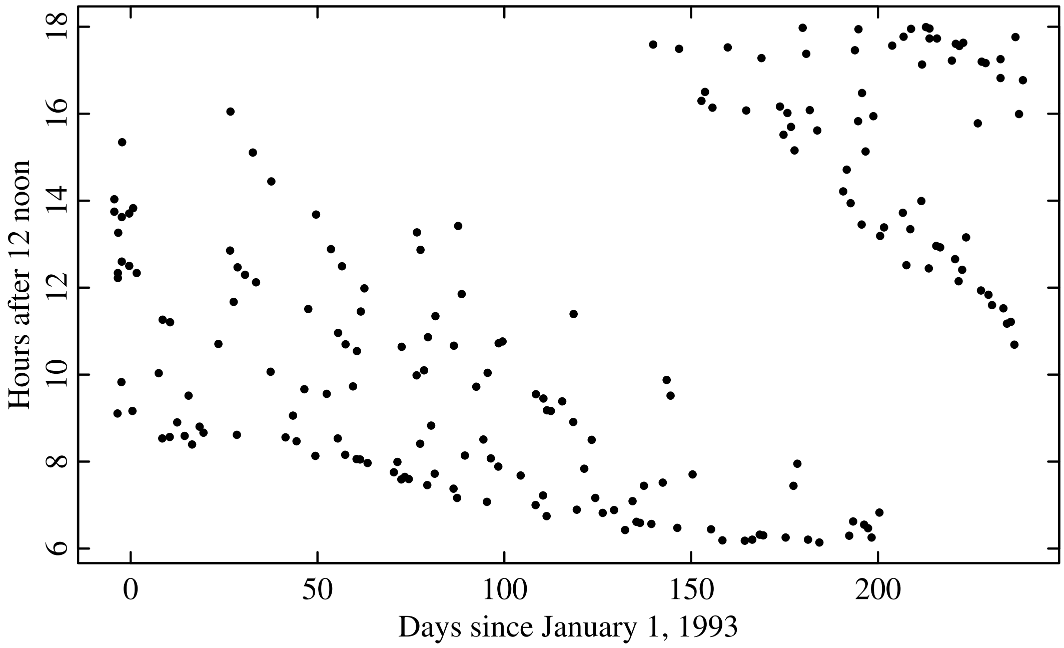

Figure 2.5: Time of night plotted against date for the times of 200 observations from the red band of star 77043:4317.

- Observation times on different days are assumed independent, while in real life there may be systematic patterns in the viewing schedule.

Thus it is possible that the results using the jittered model are not useful for real-world data. We present simulations using observation times from one of the MACHO light curves to compare the behavior of the periodogram and maximum likelihood estimates with that under the jittered model.

The observation times used are taken from the red band of star 77043:4317. In keeping with the simulations in Section 2.4.1, a sequence of 200 observations was used, spanning the 243-day period from December 27, 1992 until August 27, 1993. The time of night at which these observations were taken ranged between 6:08pm and 6:00am. Figure 2.5 shows the time of day plotted against date for these 200 times. The distribution of the times differs from the jittered model in a number of ways:

- The span of time of day observed varies with the season; near the Winter solstice (December 21) the span of times is less than eight hours, while in the middle of Summer (the Summer solstice falls near Day 170) the observation times span a 12-hour period.

- There is a substantial period of time during the night in which no measurements were taken, and this varies with the time of year; this corresponds to the times when the Large Magellanic

Cloud is either below the horizon or too close to the horizon for clear viewing.

- Over the 243-day observation period, there were 102 days (43%) on which no data was collected. A single observation was taken on 87 of the days (36%), two observations were taken on 51 days (21%) and there were three days on which three or more observations were made. This gives an average sampling rate of 0.82 observations per day.

Data Generation

Simulations were performed as described in "Data Generation" in Section 2.4.1, except that instead of generating the observation times from the jittered sampling model with $\delta_1 \sim U[-\frac{1}{6}, \frac{1}{6}]$, the observations times in Figure 2.5 were used. So while previously the set of 200 observations times varied over the 100 simulated datasets, in these simulations the observation times are fixed while the observational noise varies.

Frequency Not Near a Multiple of π.

Simulation results are presented in order to compare the bias and variance of parameter estimates from the two methods in an example in which they both converge to the true parameters.

Example 1. Frequency equal to 2.4π, continued.

Data were generated from the model (2.2) with $w = 2.4\pi$, $[a, b] = [1, 0],\ [0, 1]$, and using the observation times from Figure 2.5. The variance of the observational noise was $\sigma^2 = 0.04$. Table 2.10 displays the bias, variance, and MSE of the unscaled frequency estimates from maximum likelihood and the periodogram under this model. In all cases, all the frequency estimates were close to the true values, with the MSE converging to zero with rate approximately n^{-3}. The periodogram has a higher MSE than the MLE in almost every case, but the difference is less pronounced than in the simulations with the jittered model, when the ratio of the variance of the periodogram estimates to the variance of the MLE was about 7. The variances of the maximum likelihood and periodogram estimates are almost identical, but the bias in the periodogram estimates is much larger, falling well outside the ± 2 SE bounds in each case. This differs from the jittered sampling simulations, in which the biases were negligible but the variance differed between the maximum likelihood and periodogram estimates. In the jittered model simulations, the observation times were random, while in the simulations of this section the sampling times were fixed. The

Table 2.10: Bias, variance, and MSE of the maximum likelihood and periodogram estimates of frequency, for true frequency $w = 2.4\pi$ and 2 values of $[a, b]$. The noise variance is $\sigma^2 = 0.04$ and the observation times are those of Figure 2.5.

		Maximum Likelihood			Periodogram		
a, b	n	Bias/SE	Var.	MSE	Bias/SE	Var.	MSE
1,0	25	-0.6800	5.53 e-05	5.56 e-05	-37.60	5.41 e-05	8.17 e-04
	50	-0.0245	4.27 e-06	4.27 e-06	3.93	3.30 e-06	3.82 e-06
	100	-1.0300	5.06 e-07	5.12 e-07	-8.65	5.75 e-07	1.00 e-06
	200	-0.2040	7.71 e-08	7.71 e-08	3.92	7.22 e-08	8.32 e-08
0,1	25	0.6440	6.33 e-05	6.36 e-05	38.60	3.61 e-05	5.75 e-04
	50	0.2290	3.71 e-06	3.71 e-06	-4.20	5.11 e-06	6.01 e-06
	100	1.6800	7.30 e-07	7.50 e-07	7.41	6.07 e-07	9.40 e-07
	200	0.7190	7.37 e-08	7.41 e-08	-2.26	7.71 e-08	8.10 e-08

variances of the frequency estimate under fixed and random sampling times are related by

$$\text{Var}[w_n] = \text{E}[\text{Var}[w_n|t_1, \ldots, t_n]] + \text{Var}[\text{E}[w_n|t_1, \ldots, t_n]],$$

in which w_n denotes an estimate of w at sample size n. Since the variances of the periodogram and maximum likelihood estimates are virtually identical when conditioning on the sample times, the increased variance of the periodogram estimate under the unconditional model is due to a large value of $\text{Var}[\text{E}[\hat{w}_n|t_1, \ldots, t_n]]$. Thus the periodogram estimate has bias that depends on the random sampling times, confirming our supposition in Section 2.3.1 that the periodogram estimate is adversely affected by variability in the sampling times, while the MLE is not affected in this fashion. □

Frequency a Multiple of π.

Simulations were performed to see if the periodogram estimate appears to be inconsistent for some frequency $w = k\pi$, as was the case for the jittered model, and to compare this with the behavior of the MLE.

Example 3. Frequency Equal to 2π, continued.

Table 2.11: Mean squared error of frequency estimates from maximum likelihood and the periodogram, for frequency $w = 2\pi$, $[a, b] = [1, 0]$ and $[0, 1]$, and observation times from Figure 2.5. The MSE was calculated about the apparent limiting frequency, w_*; the figures in parentheses are the number of outlying points not included in the MSE calculation. Note: In the first and third columns of MSE values, the omitted values were clustered about 0 and 4π respectively, while the omitted values in the second and fourth columns were in the neighborhood of 2π.

	Model with $[a, b] = [1, 0]$		Model with $[a, b] = [0, 1]$	
Observations	MLE $w_* = 2\pi$	PG $w_* = 0$	MLE $w_* = 2\pi$	PG $w_* = 4\pi$
25	3.16 e–4 (1)	0.00 e+0	7.10 e–5 (3)	5.54 e–4 (25)
50	6.18 e–6	0.00 e+0	3.00 e–6	4.95 e–6
100	1.26 e–6	0.00 e+0	5.59 e–7	4.84 e–6
200	7.79 e–8	0.00 e+0 (1)	7.17 e–8	NA (100)

Table 2.11 displays the mean squared error of the maximum likelihood estimates from the model (2.2) with $w = 2\pi$, $[a, b] = [1, 0]$ and $[0, 1]$, and variance $\sigma^2 = 0.04$, using the actual observation times described above. For both values of $[a, b]$, the MLE appears to be converging to the true frequency. The periodogram estimate does not appear to converge to the true value in either case: for $[a, b] = [1, 0]$ all but one of the estimates are identically zero (similar to what was observed in Table 2.3), while for $[a, b] = [0, 1]$ the estimates were mainly clustered about the values 2π and 4π but did not seems to be converging to one of these in particular. □

Example 4. Frequency Equal to π, continued.

The results from the equivalent simulations with $w = \pi$ are presented in Table 2.12. For both values of $[a, b]$, the maximum likelihood estimate of frequency appeared to converge to the actual value, with the MSE falling with rate approximately n^{-3}. For $[a, b] = [1, 0]$, the periodogram estimate did not seem to be converging, with 71 out of 100 estimates close to 5π for $n = 100$, but all 100 estimates close to 3π for $n = 200$. Also, the MSE does not seem to be decreasing with n. For $[a, b] = [0, 1]$ the estimates seem to converge to the true value, if slowly. □

Example 4. Frequency Equal to π, continued.

Simulations were run with $w = 0$ and $[a, b] = [1, 0]$. Both the maximum likelihood and periodogram estimates converged to the true frequency. Thirteen of the maximum likelihood estimates were clustered near 2π for $n = 25$. The MSE of the maximum likelihood estimates was

Table 2.12: Mean squared error of frequency estimates from maximum likelihood and the periodogram for frequency $w = \pi$, where the MSE was calculated about the apparent limiting frequency. The observation times came from Figure 2.5, and the noise variance was $\sigma^2 = 0.04$. Note: In the first column of MSE values, the omitted values were in the neighborhood of 3π, while the omitted values were in the vicinity of π in the second column and close to zero in the third column.

		Model with $[a, b] = [1, 0]$			Model with $[a, b] = [0, 1]$	
		MLE	PG		MLE	PG
Observations		$w_* = \pi$	$w_* = 3\pi$		$w_* = \pi$	$w_* = \pi$
25		6.36 e–5 (28)	3.40 e–3 (55)		6.37 e–5 (1)	6.58 e–4
50		2.48 e–6 (8)	4.67 e–5 (19)		3.17 e–5	9.24 e–6
100		4.64 e–7 (5)	3.45 e–5 (71)		4.72 e–6	1.07 e–5
200	4.74 e–8	1.63 e–5		1.97 e–7	2.72 e–6	

7.36 e-4 for $n = 25$, 4.30 e-5 for $n = 50$, 7.75 e-6 for $n = 100$, and 1.76 e-6 for $n = 200$ (omitting the outlying values for $n = 25$), which is approximately n^{-3} convergence. Approximately 70% of the MLEs were identically zero at each sample size. The periodogram had perfect estimation, with all estimates identically zero for all sample sizes, showing fast convergence as discussed in Hannan [22]. □

The simulations using $w = 0,\ \pi,\ 2\pi$ demonstrate that the periodogram applied to actual observation times can give estimates that do not converge to the true values, as was predicted by studying the jittered model. So use of the jittered observation model for investigating theoretical properties seems to be supported by simulations using actual collection times from astronomical data. The MLE seemed to converge correctly for both frequencies, suggesting that the MLE should be used instead of the periodogram for unequally spaced observation times, if the time sampling had periodic effects and the curve of interest is a simple cosine with gaussian-like errors.

2.5 Estimation Using Semiparametric Models

The previous sections discussed the estimation of frequency when the response function is sinusoidal. Regular periodic variation in time series can have a functional form quite different from this, and the semiparametric model (2.1) is often more appropriate. In this section we first discuss complications that arise when the semiparametric model is used and introduce the concept of variance bounds for semiparametric models. We consider the variance bound of the frequency estimate for two sampling models in which the sampling times are independent. In Section 2.5.1 we derive the information bound for the frequency estimate under a model in which the range of the sampling times grows with n, and show that the bound is the same if s is known except for its phase, and in Section 2.5.2 we propose a variance bound for the frequency under the jittered sampling model.

The frequency is not identifiable under the semiparametric model unless there are additional conditions on the periodic function s. If both s and

$$s^*(u) = s(ku)$$

are in the function space $\mathcal{S}$, then the model with periodic function s^* and frequency $\frac{1}{k}w$ is identical to the model with periodic function s and frequency w. We call $\frac{1}{k}w$ a *subharmonic* of w. In order that the frequency be identifiable, we define the fundamental frequency to be the largest frequency out of the group of frequencies that give identical models. A related complication is that of the *harmonics*. If the Fourier decomposition of $s(wt)$ has a large component at the frequency kw, then the model with this frequency might explain the data almost as well as the model with frequency w. Just as the pseudo-aliases $\pm w + k$ were a complicating factor under the simple harmonic model, pseudo-aliases of the harmonics and subharmonics will play a role in the semiparametric model.

Information Bound in Parametric Models

Suppose that we have random variables $x_j,\ j = 1, \ldots, n$, and that the probability density of x_j is $f_j(x_j; \theta)$, where θ is a parameter vector from space Θ. Suppose also that Fisher information of the first n observations is again denoted by $\Psi^2(n, \theta)$ and that $r_n^{-2}\Psi^2(n, \theta)$ converges to $I(\theta)$ as $n \to \infty$, where r_n is a scaling factor. Then under regularity conditions, the variance bound for the estimators is defined by the following: For all estimators $T = \{T_n\}$ of θ for which $r_n(T_n - \theta)$ converges uniformly to a normal distribution, the asymptotic variance satisfies

$$\Sigma(\theta, T) \equiv \lim_{n\to\infty} \text{Var}[r_n(T_n - \theta)] \geq I^{-1}(\theta), \tag{2.91}$$

where the inequality is in the sense of ordering of nonnegative-definite matrices. $I^{-1}(\theta)$ is the *information bound* for estimation of θ, and an estimator T is *asymptotically efficient* if and only if equality holds in (2.91), but calculation of the information bound does not guarantee that efficient or even uniformly normal estimators exist.

Information Bound in Semiparametric Models

Suppose now that the parameters of the model are $[\theta, G]$, in which θ is a parameter vector as before and G is an infinite-dimensional parameter from space $\mathcal{G}$. Let T be an estimator of θ which converges uniformly to a normal distribution, and $\mathcal{G}_f$ be a finite-dimensional subspace of $\mathcal{G}$. Then if $I(\theta, G; \mathcal{G})$ is the limiting information of θ for this submodel under scaling r_n, under regularity conditions the asymptotic variance of $r_n(T_n - \theta)$ satisfies

$$\Sigma(\theta, G, T) \geq I^{-1}(\theta, G; \mathcal{G})$$

for parameters with $G \in \mathcal{G}$. Combining this over all such finite-dimensional subspaces $\mathcal{G}_f$, we define the information bound for θ under the semiparametric model by

$$I^{-1}(\theta, G) \equiv \sup_{\mathcal{G}_f}\{I^{-1}(\theta, G; \mathcal{G})\},$$

and under regularity conditions, estimators T of θ that converge uniformly to a normal distribution over $[\Theta, \mathcal{G}]$ have asymptotic variance bounded by $I^{-1}(\theta, G)$. Just as for the parametric model, existence of this bound does not guarantee existence of an efficient estimator or even of one that is uniformly convergent with scaling r_n.

2.5.1 Sampling Without a Day-Effect

We use this model to study the behavior of the frequency estimate when the range of the sample times grows like n, and there are no periodicities in the sampling distribution. For given n, the sample times are modeled as

$$t_j = nu_j,\ j = 1, \ldots, n, \tag{2.92}$$

where the u_j are IID from a distribution on some closed interval $[a, b]$ with density $h(.)$ that does not change with n. As the distribution of the sample times now depends on n, we introduce the triangular array notation: For each n, denote the sampling times by $t_{n1}, t_{n2}, \ldots, t_{nn}$ and the probability density of t_{nj} by $h_{nj}(t)$, which under the model (2.92) is

$$h_{nj}(t) = \tfrac{1}{n}h(\tfrac{1}{n}t). \tag{2.93}$$

The responses and observational errors are similarly denoted by y_{nj} and ϵ_{nj}, and the response model (2.1) for given n becomes

$$y_{nj} = s(wnu_j) + \epsilon_{nj},\ j = 1, \ldots, n. \tag{2.94}$$

Note that this is equivalent to a model with IID sample times $t_{nj} = u_j$, and with frequency that grows proportionally to n.

Proposition 3 *Consider the response model (2.94), and assume the following:*

(1) *The errors ϵ_{nj} are IID with density $g(\epsilon)$ independently of $\{t_{nj}\}$, and for some $\rho > 0$,*

$$\int \left|\frac{g'}{g}(\epsilon)\right|^{2+\rho} g(\epsilon)\, d\epsilon < \infty, \quad \int \left[\frac{\partial^2}{\partial\epsilon^2} g^{1/2}(\epsilon)\right]^2 d\epsilon < \infty. \tag{2.95}$$

(2) *The probability density of u_1, $h(u)$, is continuous for $u \in [a, b]$ with bounded derivative, and Var $[u_1] > 0$.*

(3) *The regression curve s comes from $\mathcal{S}$, the set of periodic functions with period one that have uniformly-bounded second derivatives, and that satisfy $\int_0^1 s'(u)^2 du > m$ for some $m > 0$.*

(4) *The parameter $\theta \equiv w$ comes from $\Theta = (m, M)$, where $m > 0$ and $M < \infty$.*

Then over the parameter space $[\Theta, \mathcal{S}]$,

(a) *If s is known except for its phase, the information bound for estimation of w is*

$$I^{-1}(w, s) = \left\{ I \, \textit{Var}\,[u_1] \int_0^1 s'(u)^2 du \right\}^{-1} \tag{2.96}$$

with scaling factor $r_n = n^{3/2}$, where I is the amount of information in the density $g(\epsilon)$.

(b) *If the regression function $s \in \mathcal{S}$ is unknown, the information bound for estimation of w is the same as in* **(a)**.

Comments:

- Although calculation of the information bound does not prove existence of asymptotically efficient estimates, the asymptotic variance of an efficient estimator of w is no larger if the shape of s is unknown than if it is known. This is surprising, as one expects to pay a penalty for not knowing the shape of the function.

- If the periodic function and its phase are known, the information bound is

$$I^{-1}(w) = \left\{ I \, \mathrm{E}[\, u_1^2 \,] \int s'(u)^2 \, du \right\}^{-1},$$

which is usually smaller than the bound for the unknown case. This bound is not invariant under shifts in the time axis, because knowing both the shape and the phase of the curve means that the curve is known at time zero, which is like having an additional observation with no noise at $t = 0$.

- If the noise is normally distributed, the information bound is inversely proportional to the square of the signal-to-noise ratio, as one would expect: if the curve s has amplitude A and the noise variance is σ^2, the information bound is proportional to σ^2/A.

- The information bound is inversely proportional to the variance of the sample times. For example, if u_1 comes from a uniform distribution over the interval $[0, C]$, the scaled variance of an efficient estimator of w is proportional to $1/C^2$. Thus to minimize the variance of the frequency estimate, observation times should be sampled over the widest possible range.

- For a given signal-to-noise ratio, the asymptotic variance of an efficient frequency estimator is smaller for regression curves with larger derivatives, through the value of $\int_0^1 s'(u)^2 du$. This means that periodic functions with jagged or unusual shapes, instead of causing problems with the estimation, should yield frequency estimates with smaller variance.

- The variance bound for subharmonics of the true frequency is equivalent to that at the true frequency. Suppose that the response function has frequency w with periodic curve $s(t)$. This is indistinguishable from the model with frequency w/k and curve $s(kt)$, where k is a positive integer, but the information bound for estimation of w is the same in each case. The frequencies w/k are known as *subharmonics*, and $[w, s]$ is not identifiable when both $s(t)$ and $s(kt)$ are in the function space $\mathcal{S}$.

Proof of Proposition 1.

(a). This situation can be modeled as

$$y_{nj} = s(wnu_j + \rho) + \epsilon_{nj},\ j = 1, \ldots, n,$$

with parameter vector $\theta = [w, \rho]$, where ρ is the unknown phase; the probability density of $x_{nj} = [y_{nj}, u_j]$ is

$$f_{nj}(x_{nj}; \theta) = g(y_{nj} - s(wnu_j + \rho))h(u_j). \tag{2.97}$$

The Fisher information of the first n observations is

$$\Psi^2(n, \theta) = I\,\mathrm{E}\left[\begin{array}{cc} ns'(wnu_j + \rho)^2 & n^2u_js'(wnu_j + \rho)^2 \\ * & n^2u_j^2s'(wnu_j + \rho)^2 \end{array}\right], \tag{2.98}$$

where I is defined previously. The expectation of $s'(wnu_1+\rho)^2$ under the sampling model (2.92) is

$$\begin{aligned}
\mathrm{E}[s'(wnu_1+\rho)^2] &= \int_a^b s'(wnu+\rho)^2 h(u)du \\
&= \sum_{k=0}^{\lfloor nw(b-a)\rfloor} \int_{a+\frac{k}{nw}}^{a+\frac{k+1}{nw}} s'(wnu+\rho)^2 h(u)du \\
&= \sum_{k=0}^{\lfloor nw(b-a)\rfloor} \left\{ h(a+\tfrac{k}{nw}) \int_{a+\frac{k}{nw}}^{a+\frac{k+1}{nw}} s'(wnu+\rho)^2 du + O((nw)^{-2}) \right\},
\end{aligned} \tag{2.99}$$

as both s' and h' are bounded [$\lfloor x \rfloor$ stands for the largest integer less than or equal to x]. Denote $\int_0^1 s'(u)^2 du$ by $||s'||^2$. Then

$$\begin{aligned}
\mathrm{E}[s'(wnu_1+\rho)^2] &= ||s'||^2 \frac{1}{nw} \sum_{k=0}^{\lfloor nw(b-a)\rfloor} h(a+\frac{k}{nw}) + O(n^{-1}) \\
&= ||s'||^2 + O(n^{-1}),
\end{aligned}$$

and similar working for the other terms in (2.98) yields

$$\mathrm{E}[s'(wnu_1)^2 u_1] = \mathrm{E}[u_1]\,||s'||^2 + O(n^{-1}),$$

$$\mathrm{E}[s'(wnu_1)^2 u_1^2] = \mathrm{E}[u_1^2]\,||s'||^2 + O(n^{-1}).$$

In all three formulae, the order term is uniform for w bounded above zero, and so the information about θ in the first n observations is

$$\Psi^2(n,\theta) = I\,||s'||^2 \begin{bmatrix} n+O(1) & n^2\mathrm{E}[u_1]+O(n) \\ * & n^3\mathrm{E}[u_1^2]+O(n^2) \end{bmatrix}.$$

This matrix is invertible for $\mathrm{Var}[u_1] > 0$ and $||s'||^2 \neq 0$, with inverse

$$\Psi^{-2}(n,\theta) = \{I\,||s'||^2\,\mathrm{Var}[u_1]\}^{-1} \begin{bmatrix} n^{-1}\mu_1' & -n^{-2}\mu_1' \\ * & n^{-3} \end{bmatrix}$$

and inverse symmetric square root

$$\Psi^{-1}(n,\theta) = \{I\,||s'||^2\,\mathrm{Var}[u_1]\}^{-1/2} \begin{bmatrix} \sqrt{\mu_2'}\,n^{-1/2} & -\frac{\mu_1'}{\sqrt{\mu_2'}} n^{-3/2} \\ * & \sqrt{1-\frac{(\mu_1')^2}{\mu_2'}}\,n^{-3/2} \end{bmatrix},$$

in which $\mu_1' = \mathrm{E}[u_1]$, $\mu_2' = \mathrm{E}[u_1^2]$, and the inferior order terms have been omitted.

Having calculated the limiting inverse information, we need to confirm that it acts as a variance bound for estimates of θ. Under conditions **(1)** – **(4)**, this model is regular in the sense of Ibragimov and Has'minskii [24]. In order to verify the information bound, the remaining property required is local asymptotic normality (LAN), defined in Ibragimov & Has'minskii, p. 120. LAN is shown by Theorem II.6.1 of Ibragimov & Has'minskii, which requires the conditions:

(I) For some $\delta > 0$ and $u \in \Re^2$,

$$\lim_{n\to\infty} \sum_{j=1}^{n} \mathrm{E} \left| \left[\tfrac{\partial}{\partial\theta} \ln f_j(x_j;\theta)\right]^T \Psi_n^{-1} u \right|^{2+\delta} = 0.$$

(II) For any $k > 0$,

$$\lim_{n\to\infty} \sup_{|u|<k} \sum_{j=1}^{n} \int \left| \Psi_n^{-1} \tfrac{\partial^2}{\partial\theta^2} f_j^{1/2}(x_j, \theta + \Psi_n^{-1} u) \Psi_n^{-1} \right|^2 dx_j = 0.$$

These are virtually the same as conditions **(B)** and **(C)** in the proof of Theorem 3, and can be shown without difficulty in the same way. Thus LAN is satisfied, and estimators T of θ for which $N^{1/2}(T_n - \theta)$ converges uniformly to a normal distribution (where $N = \mathrm{diag}[n, n^3]$) have asymptotic variance bounded by

$$\lim_{n\to\infty} N^{1/2} \Psi^{-2}(n,\theta) N^{1/2} = \{I\,\mathrm{Var}[u_1]\,||s'||^2\}^{-1} \begin{bmatrix} \mu_2' & -\mu_1' \\ * & 1 \end{bmatrix},$$

which completes the proof of **(a)**.

(b). Bickel, Klassen, Ritov, and Wellner [6], pp. 108-110, derives the information for the periodic regression model with IID sample times under normality. If we adapt their arguments to data in a triangular array, and to noise with a general probability density, the information in the first n observations for the estimation of $\theta = w$ is

$$\Psi^2(n,\theta) = In\mathrm{E}[s'(nwu_1)^2(nu_1 - \mathrm{E}[nu_1|w])^2], \tag{2.100}$$

where $\mathrm{E}[t_{nj}|w]$ is defined by

$$\mathrm{E}[t_{nj}] \equiv \frac{\sum_{k=-\infty}^{\infty} (t_{nj} + \frac{k}{w})\, h_{nj}(t_{nj} + \frac{k}{w})}{\sum_{k=-\infty}^{\infty} h_{nj}(t_{nj} + \frac{k}{w})}, \tag{2.101}$$

which is the expectation of t_{nj} over all the values of t that have the same phase with respect to frequency w. For the model $t_{nj} = nu_j$,

$$\begin{aligned} \mathrm{E}[nu_1|w] &= \frac{n\sum_{k=-\infty}^{\infty}(u_1+\frac{k}{nw})h(u_1+\frac{k}{nw})}{\sum_{k=-\infty}^{\infty}h(u_1+\frac{k}{nw})} \\ &= n\mathrm{E}[u_1]+O(w^{-2}), \end{aligned}$$

and so as w is bounded above zero,

$$\Psi^2(n,\theta) = In^3\mathrm{E}[s'(nwu_1)^2(u_1-\mathrm{E}[u_1])^2]+O(n^2).$$

The expectation term is calculated as in part **(a)**, and the scaled information $n^{-3}\Psi^2(n,\theta)$ converges uniformly to

$$I\,\mathrm{Var}[u_1]\,||s'||^2,$$

which is continuous in w and s and nonsingular. This completes the regularity conditions, and as LAN follows by Proposition II.1.2 of BKRW, **(b)** is satisfied. □

2.5.2 Sampling from the Jittered Model

In this section we present a conjecture about the information bound for the frequency under randomly jittered sampling.

Conjecture 1 *Consider the semiparametric response model (2.1) under jittered time sampling, and assume conditions* **(1)**, **(3)**, *and* **(4)** *of Proposition 1, as well as*

(2*) *The probability density of δ_1, $h(\delta)$, is continuous on $[-\Delta,\Delta]$.*

Then over the parameter space $[\Theta,\mathcal{S}]$,

(a) *If s is known except for its phase ρ, the information bound for estimation of w is*

$$\mathit{Var}[n^{3/2}w_n] \geq 12\left\{I\,||s'||_w^2\right\}^{-1}. \tag{2.102}$$

Here, w_n is a regular estimator of w and $||s'||_w^2$ is the average squared derivative of the periodic function at the sample times, defined by

$$||s'||_w^2 = \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{j=1}^{n} s'(wt_j+\rho)^2\right].$$

(b) *If the periodic function $s \in \mathcal{S}$ is unknown, the information bound for estimation of w is the same as in* **(a)**.

Comments:

- Support for this conjecture comes from consideration of the case in which s has a p-term Fourier decomposition, when $||s'||_w^2$ takes the values

$$||s'||_w^2 = \begin{cases} \int_0^1 s'(u)^2 du, & w \text{ irrational}, \\ \frac{1}{k}\sum_{j=1}^{k} \int s'\left(\frac{lj}{k} + \frac{l}{k}u + \rho\right)^2 h(u)du, & w = \frac{l}{k},\ l \in \mathcal{Z}, k \in 1, \ldots, 2p, \\ \int_0^1 s'(u)^2 du, & \text{otherwise}. \end{cases} \tag{2.103}$$

  It can be proven that the frequency bound

$$\text{Var}[n^{3/2} w_n] \geq 12\{I \int_0^1 s'(u)^2\, du\}^{-1}$$

  is valid for w bounded away from $\{\frac{l}{k},\ l \in \mathcal{Z}, k \in 1, \ldots, 2p\}$, but (2.102) cannot easily be proven at the excluded frequencies because of the discontinuities in $||s'||_w^2$. This situation is discussed in Quinn and Thomson [34] for equally-spaced times.

- For general $s \in \mathcal{S}$, $||s'||_w^2$ is possibly discontinuous at the rational frequencies, and bounding w away from all the rationals is not possible. Nevertheless, most of the discontinuities in $||s'||_w^2$ will be small and (2.102) should be an approximate lower bound on the precision of the frequency even if a rigorous proof is not possible.

- If (2.102) is valid, how do we interpret the role of $||s'||_w^2$ for the rational frequencies? $||s'||_w^2$ can be smaller or larger than $\int_0^1 s'(u)^2\, du$, depending on the sampling distribution and the shape and phase of s. Thus an estimator of frequency can be either more or less efficient at a rational frequency than at the irrational ones. This is illustrated by the jittered cosine model ((2.2),(2.4)): Under normality of the observational noise, the variance bound is

$$\text{Var}[n^{3/2} w_n] \geq \begin{cases} \frac{24\sigma^2}{(2\pi)^2}\{a^2 + b^2\}^{-1}, & w \neq \frac{k}{2}, \\ \frac{24\sigma^2}{(2\pi)^2}\left\{a^2 + b^2 + (b^2 - a^2)\phi_R(2w) - 2ab\phi_I(2w)\right\}^{-1}, & w = \frac{k}{2}, \end{cases} \tag{2.104}$$

  where σ^2 is the noise variance and $\phi_R()$ and $\phi_I()$ are the real and imaginary parts of the characteristic function of δ_1. This agrees with the asymptotic variance of w given in Theorem 3 for $w \neq k/2$, and with the inverse of the information matrix (2.88) for $w = k/2$. If the sample times are almost equally-spaced, $\phi_R(2w) \approx 1$ and $\phi_I(2w) \approx 0$, giving

$$I^{-1}(w; a, b) \approx \frac{24\sigma^2}{(2\pi)^2}\{2b^2\}^{-1},\ w = \tfrac{k}{2},$$

and so the frequency is estimable with smaller variance at these frequencies if $b^2 > a^2$, and larger variance otherwise.

- Combination of this conjecture and Proposition 3 suggests a variance bound for more general models,

$$\text{Var}[w_n] \geq n^{-1}\{I\,\text{Var}[t]\,||s'||_w^2\}^{-1}, \tag{2.105}$$

  in which Var$[t]$ is the variance of the observation times. For sampling times uniformly distributed over the interval $[0, n]$, the variance bound (2.105) becomes

$$\text{Var}[w_n] \geq \frac{12}{n^3}\{I\,||s'||_w^2\}^{-1},$$

  the same as for the jittered model in (2.102), and so the jittered model acts like uniform sampling for large n.

Chapter 3

Application

The objective of this chapter is to discuss practical methods for estimating frequency, using the MACHO light curves as an illustration. Numerous estimation methods are motivated and described in Section 3.1, and computational issues having to do with the optimization in these methods are discussed in Section 3.2. In Section 3.3, we apply these estimation methods to the examples of Chapter 1 and evaluate their performance. Section 3.4 presents the results of a simulation study which compared the precision of the estimation methods in a number of realistic situations. The final section contains a brief discussion of some interesting topics that were not able to be covered in this dissertation.

3.1 Estimation Methods

There are many different methods for estimating frequency under the semiparametric model. These methods need to be evaluated with regard to our main objective: the fast and precise estimation of the frequency which best describes the dependency in the data. As discussed in Chapter 1, these methods must satisfy this objective even when part of the curve has not been sampled, the curve is non-sinusoidal or even multimodal, or when the measurement noise is substantial. In this section we motivate and describe several frequency estimation methods, which will be evaluated in Sections 3.3 and 3.4.

We assume that the light curve of the periodic variable stars in the MACHO database satisfy the general periodic model (2.1) in which the measurement errors are distributed independently of each other and of the sampling times. Assume also that ϵ_j has mean zero and variance σ_j^2, of which we have an estimate s_j^2.

Cosine Methods

Even when the appropriate model is the general periodic regression model (2.1), we can try to estimate the frequency using the simple harmonic model (2.2). The Fourier decomposition theorem states that any function that is periodic with frequency w can be decomposed into a potentially infinite series of simple harmonic components at the frequencies $lw,\ l \in \mathcal{Z}$. If the largest Fourier component occurs at one of the harmonic frequencies, then cosine methods will tend to locate this frequency instead of the fundamental frequency. Fullerton [21] notes that the largest component in the Fourier decomposition of light curves of stars in Keplerian orbits usually occurs at the fundamental frequency, with the notable exception of eclipsing binary stars. Thus cosine methods should identify the correct frequency in the light curves for a majority of periodic variable stars.

As discussed in Sections 2.2 and 2.3, the periodogram and least-squares are the usual methods for estimating the frequency using the simple harmonic model. The behavior of the periodogram in estimating frequency is discussed in Deeming [16], and this method is often referred to as the *Deeming Periodogram* in the Astronomy literature. The advantages of the periodogram are that it is quick to calculate and that the effect of the sampling times on the estimation is completely described by the spectral window. Disadvantages of the periodogram are that the variance of the frequency estimate is larger than when using least-squares, and that the estimate can be inconsistent for certain frequencies while it appears that the least-squares estimate is consistent at these frequencies (see Section 2.4).

Both of these factors suggest that least-squares estimation be used instead of the periodogram. A version of the periodogram that is equivalent to least-squares was introduced by Lomb [29] and Scargle [36] and is known in the Astronomy literature as the Lomb-Scargle periodogram. Press and Rybicki [32] have developed an algorithm which rapidly calculates the Lomb-Scargle periodogram at an equally-spaced collection of frequencies, through the use of fast Fourier transforms. This method does not permit the use of weights and so is only appropriate when the data have variances that are not overly heterogeneous. Calculated as either a minimization of RSS or as a maximization of the Lomb-Scargle periodogram, the least-squares estimate can be calculated more rapidly than most of the other methods that we will discuss. The poor fit of the cosine to the periodic function can be expected to affect adversely the precision of the frequency estimate, and so these methods may be inappropriate for multimodal curves such as those exhibited by eclipsing binary stars.

Fourier Decomposition

Instead of modeling the periodic function by the simple harmonic model, we can model the shape of the curve more accurately by using a p-term Fourier decomposition

$$y_j = a_0 + \sum_{k=1}^{p} \{a_k \cos(2\pi w t_j) + b_k \sin(2\pi w t_j)\} + \epsilon_j. \tag{3.1}$$

For a given frequency v, one can estimate a_0, a_k, b_k by weighted linear regression, yielding fitted values $\hat{y}_{j,v}$. Define the frequency estimate from this method to be the frequency $\hat{w} \in (0, \Omega]$ that minimizes the RSS

$$\mathrm{SS}_n(v) = \sum_{j=1}^{n} \frac{1}{s_j^2}(y_j - \hat{y}_{j,v})^2. \tag{3.2}$$

This method is discussed in Quinn and Thomson [34], who derive the asymptotic properties of the estimate for equally-spaced sample times. This method has the advantages of providing an improved fit to the periodic function at a given frequency, which should improve the frequency estimate, and that the basis functions in the regression equation (3.1) are naturally periodic, which simplifies the computations. A possible difficulty with this method is that as the basis functions at the higher harmonics are highly oscillatory, there may be spurious oscillations present in the fitted light curve $\hat{y}_{j,v}$.

Cubic Splines

Another method for fitting brightness as a function of phase at a given frequency is the use of periodic regression splines or smoothing splines. We specifically consider the case of periodic cubic splines. A function $s()$ on the interval $[0, 1]$ is a periodic cubic spline with *knots* at $t_k,\ k = 1, \ldots, p$ if it satisfies the following:

1. In each interval $[t_{k-1}, t_k],\ k = 1, \ldots, p+1\ (t_0 = 0, t_{p+1} = 1)$, s is given by a cubic polynomial.

2. The function s and its derivatives of first and second order are continuous everywhere in $[0, 1]$ and satisfy the periodicity constraints

$$s^{(l)}(0) = s^{(l)}(1),\ l = 0, 1, 2.$$

Periodic smoothing splines are discussed in the context of spectral estimation in Cogburn and Davis [14] and Wahba [46]. The use of cubic regression splines in the estimation of frequency is discussed in Akerlof [1].

We propose to use cubic regression splines to estimate the frequency as follows. For each trial frequency v, calculate the phases $\rho_j = wt_j$ mod 1. Calculate fitted values for y_j by modeling brightness as a cubic spline of phase, using the B-spline basis of De Boor [15] for good numerical stability. Then define the frequency estimate to be the frequency that minimizes $\mathrm{SS}_n(v)$ over $v \in (0, \Omega]$. Some care must be taken in the choice of knots. There should be enough knots so that the function shape can be successfully approximated by the spline basis, but not so many knots that overfitting becomes a problem. Akerlof [1] uses knots that are equally-spaced along the phase interval. This can cause numerical difficulties, since the data may be unevenly scattered in phase space and one or more of the B-spline basis functions may not be supported by the data. We suggest placing the knots at the phase quantiles for equally-spaced probabilities. This puts the basis functions where there is the most information and ensures that the basis functions are supported by the data. Because a small change of frequency can lead to a change in the relative ordering of the phase values, and the knots are places at quantiles of the phase values, a tiny change in frequency can lead to a change in knot positions and a jump in the RSS. This is discussed further in Section 3.2.

Cubic splines are better able to model small features in the periodic curve, due to the restricted domain of the B-spline basis functions, but it is not certain that this will lead to a more precise estimate of frequency.

Smoothers

Another method for fitting the brightness as a function of phase at a given frequency is a smoother, such as a running mean or a running linear regression. McDonald [30] discusses estimation of frequency by use of a smoother based on split linear fits. We will be considering fitting y_j using *supersmoother*, a variable-span smoother based on running linear smooths described in Friedman [20]. Supersmoother performs three running linear smooths of the data (ρ_j, y_j) with long, medium and short span length. It then does a local cross-validation to determine what span length gives the best fit at each phase value. As the cross-validation is done on the absolute residuals instead of the squared residuals, it makes more sense to define the frequency estimate by the frequency that minimizes the sum of absolute residuals (SAR),

$$\mathrm{SAR}_n(v) = \sum_{j=1}^{n} \frac{1}{s_j} |y_j - \hat{y}_{j,v}|,$$

where the $\hat{y}_{j,v}$ are the fitted values from supersmoother at frequency v.

The supersmoother method should produce a good fitted curve for a large variety of curve shapes, since it is able to adjust the span length as required by the data and makes no explicit assumptions about the shape of the curve. As it is a smoother, the fitted curve is defined only at the observed phase values, which may be inconvenient. The SAR curve will not be continuous, since a change in the ordering of the phase values (caused by a tiny change in frequency) will change which points are in the local linear regression for a couple of phase values. The change in the SAR will be small, however, as the fitted values only change at a couple of sample points.

Measures of Dispersion

All of the above methods are based on finding the frequency for which the fitted values $\hat{y}_{j,v}$ best agree with the observed values y_j. The final class of methods do not produce fitted values for the brightnesses at a given frequency, but instead calculate a measure of dispersion of the data in the phase space and seek to minimize this over frequency.

A simple example of this is the *string-length method*, described in Dwortesky [18]. For a given frequency v, this involves producing the phase plot of the data and joining the points from left to right with line segments. The frequency estimate is the frequency that minimizes the string length

$$\mathrm{STR}_n(v) = \sum_{j=1}^{n} \{(y^*_{j+1} - y^*_j)^2 + (\rho^*_{j+1} - \rho^*_j)^2\}^{1/2}.$$

Here, y^*_j and ρ^*_j are the response and phase values sorted by phase and the values y^*_{n+1} and ρ^*_{n+1} are the same as y^*_1 and ρ^*_1. There are a number of difficulties with this method. The string-length depends on differences in phase as well as in response, and so a change of variable in either could lead to a different frequency estimate. Additionally, use of the difference in phase causes biasing in favor of periods for which the points on the phase interval are clustered together.

There are two alternative methods that seek to avoid the problems of the string-length method. One of these is due to Lafler and Kinman [28] and estimates the frequency by minimizing the quantity

$$\mathrm{LF}_n(v) = \sum_{j=1}^{n} \frac{(y^*_{j+1} - y^*_j)^2}{(s^{*2}_{j+1} + s^{*2}_j)},$$

in which s^{*2}_j is the estimated variance of y^*_j. This removes the two difficulties with the string-length method, since it does not explicitly depend on the phase differences, but is still not completely satisfactory as the difference between two points distant in phase is given the same weight as two points close in phase. The other method is due to Renson [35] and estimates the frequency by minimizing the quantity

$$\text{REN}_n(v) = \sum_{j=1}^{n} \frac{(y_{j+1}^* - y_j^*)^2}{(s_{j+1}^{*2} + s_j^{*2})((\rho_{j+1}^* - \rho_j^*)^2 + b^2)},$$

in which b is a quantity chosen so that the difference $(\rho_{j+1}^* - \rho_j^*)^2 + b^2$ should not be too small. Renson recommends taking $b^2 = \frac{s^2}{4A^2}$, where s^2 is the standard error of the measurement noise and A is the amplitude of the signal. This method gives higher weight to comparisons between brightnesses when the corresponding phases are close.

All of these methods have relatively short computation times and require no assumptions on the shape of the periodic curve, except that it be continuous. The measures $\text{LF}_n(v)$ and $\text{REN}_n(v)$ behave like the RSS for the model in which the fitted value for y_j^* is y_{j+1}^*, which has as many parameters as observations. Thus it is possible that these methods will not perform well when the measurement noise is too large.

Stellingwerf PDM

Another method based on a measure of dispersion and used by astronomers is *Phase Dispersion Minimization (PDM)*, introduced by Stellingwerf [43]. This is a modification of the Whittaker periodogram (Whittaker and Robinson [49]) in which the phase interval is divided into a number of bins, the mean response is calculated for each bin, and the frequency is chosen to minimize the RSS of the one-way analysis of variance based on these bins. The Stellingwerf method introduces additional bins, which are the original bins shifted in phase by a certain amount; the set of shifted bins are called a *cover*. For example, the Stellingwerf method with two covers of five bins yields the bin intervals [0,0.2], [0.2,0.4], [0.4,0.6], [0.6,0.8], [0.8,1], [0.1,0.3], [0.3,0.5], [0.5,0.7], [0.7,0.9], and [0,0.1]∪[0.9,1]. The overlapping bins are introduced to reduce the edge effects in the Whittaker periodogram. The PDM method should behave similarly to an estimation method based on a fixed-span smoother, and makes more assumptions about the shape of the function than the Lafler and Renson methods, through the choice of the number of bins and number of covers.

3.2 Global Optimization

All of the estimation methods discussed in the previous section calculate the frequency estimate by maximizing or minimizing a function of frequency over an interval on the real line. Let

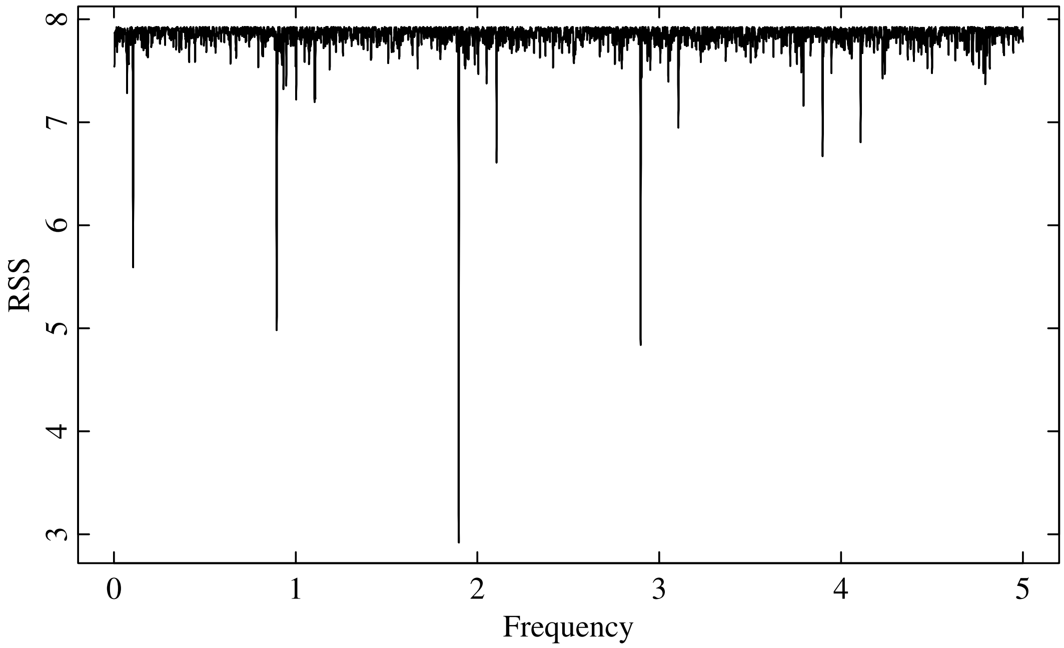

Figure 3.1: Weighted RSS of the cosine model plotted against frequency, for the blue band of star 77009:64163.

us call this function the *objective function*. In this section, we describe some of the complications that arise in this optimization and propose an optimization scheme.

Suppose that we want to minimize an objective function $M(v)$ over the frequency interval $[a, b]$ (maximization is considered by minimizing the negative of the objective function). The objective functions of the above frequency estimation methods have many local minima, making global minimization difficult. As an example, consider the least-squares estimator of the cosine model. Figure 3.1 shows the weighted RSS (3.2) for the data of Figure 1.7, plotted over the frequency range [0,5]. There is a clear global minimum at the frequency $w = 1.91$, which corresponds to the period 0.52715 shown in the phase plot in Figure 1.8. There are also a very large number of local minima, the most deep of which correspond to the pseudo-alias frequencies $\pm w + k$. More generally, there can also be deep local minima at the subharmonic frequencies w/k, the harmonic frequencies wk, and the pseudo-aliases of these as well. Together with the presence of observational noise, there is no guarantee that the deepest minimum corresponds to the fundamental frequency, and so a number of the deepest local minima should be checked for quality of fit.

Another complication is that for some frequency estimation methods, the objective

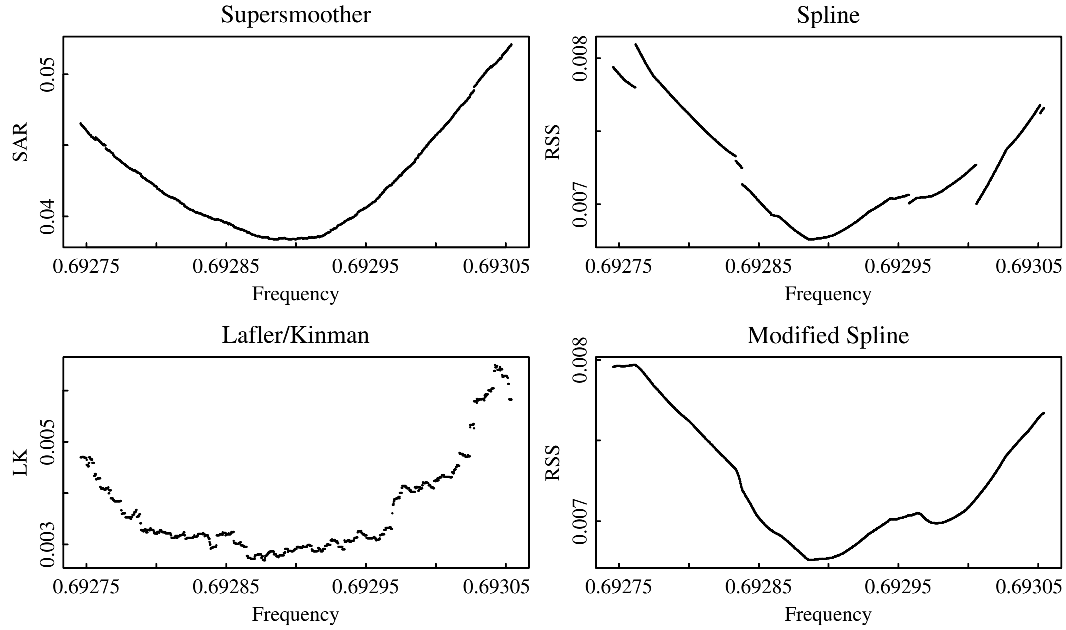

Figure 3.2: Objective functions for the supersmoother, Lafler, spline, and modified spline methods for a simulated data set on the frequency interval [0.69275,0.69305].

function may be discontinuous. The objective functions for the cosine and Fourier methods are continuous, but those of the other methods of Section 3.1 are not. Figure 3.2 shows the objective functions for four estimation methods calculated on simulated data, displayed over the frequency interval [0.69275,0.69305]. In the upper-left of the figure is the SAR from the supersmoother method; note the small discontinuities in the curve. The Lafler measure of dispersion is shown in the lower-left of the figure; the discontinuities in the objective function are much more pronounced than in the supersmoother method. How do these discontinuities come about? As the frequency changes, the phase values also change according to the relation $\rho = wt$ mod 1. For both of these methods, the discontinuities occur at a frequency at which the ordering of two phase points changes. The discontinuities in these objective functions can be seem more clearly in Figure 3.3. The upper plot shows the SAR function from the supersmoother method plotted over the frequency interval [0.69288,0.69292], and the lower plot shows the Lafler measure of dispersion over the same range. The Supersmoother SAR function is piecewise continuous, while the LK function is piecewise constant.

A different kind of discontinuity is presented by the periodic spline method. The upper-right plot of Figure 3.2 shows the RSS from the periodic spline method. A cubic spline with eight

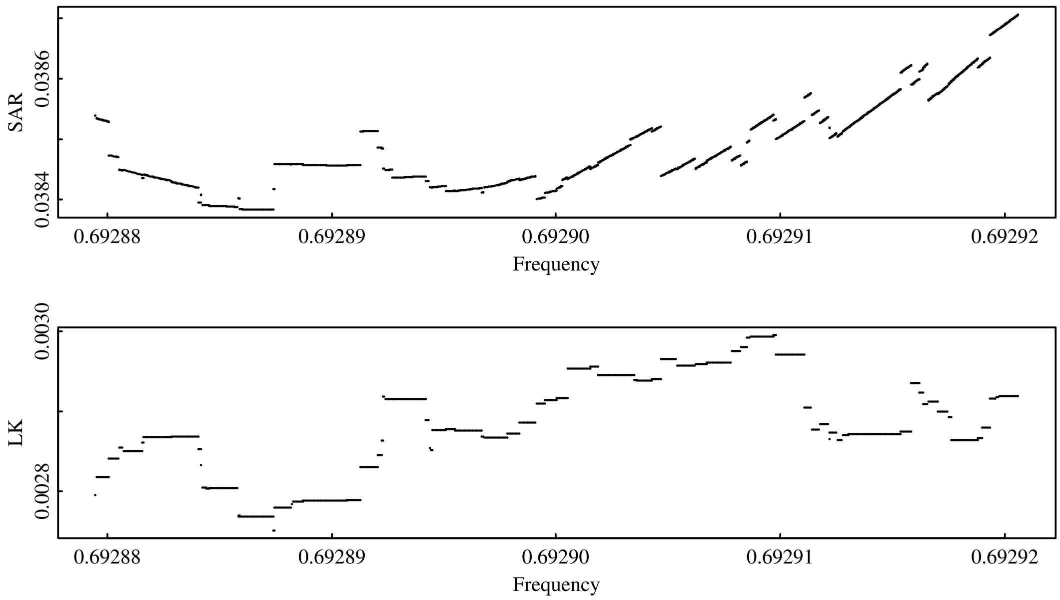

Figure 3.3: Objective functions for the supersmoother and Lafler methods for a simulated data set on the frequency interval [0.69288,0.69292].

interior knots was used, with the knots placed at the quantiles of the phase vector. Note that there are fewer discontinuities in the curve than for the supersmoother or Lafler methods, but that the size of the jumps is larger. These jumps occur when a small change in frequency causes a phase value to change from one end of the interval [0,1] to the other. Since the knots are placed at the quantiles of phase, this makes the position of the interior knots change and alters the fitted values at almost all of the points. To reduce this non-continuous behavior, we shall use a modified form of knot placement: the interior knots are placed at a weighted average of the ordered phase values near the quantile in such a way that the change in knot position is continuous in frequency. The RSS from this modified spline method is shown in the lower-right of Figure 3.2. Although the RSS curve is continuous, it is still not smooth, which may cause difficulties in the optimization.

There are many methods available for solving our optimization problem: finding the global optimum over some interval on the real line. These are divided into *deterministic methods*, such as minimization over a grid of points, and *probabilistic methods*, such as the gradient descent method with a large number of random starting values. An overview of various deterministic and probabilistic global optimization methods is given in Dixon and Szegö [17]. These methods are not guaranteed to find the global minimum without some conditions on the objective function. If a Lipschitz condition is satisfied, i.e., an upper bound exists and is known for the derivative of the objective function, then there exist optimization methods that can find all frequencies for which the objective function is within an arbitrarily small tolerance of the global minimum (Shubert [37], Mladineo [31]). A similar method can be used when there is a bound on the second derivative of the objective function (Breiman and Cutler [9]). Unfortunately, we saw above that the objective functions can be pointwise discontinuous, and so for our problem we cannot prove that the global minimum is attained.

We decided to minimize the objective function using a grid method. Although there is no bound on the derivative of the objective function, other information is available on how fine a grid is required to yield acceptable results. It was noted in Section 2.1 that the central peak in the spectral window has width of approximately $2T^{-1}$, where T is the span of the sampling times. Thus the grid spacing needs to be shorter than this in order to find the function minima. A grid spacing of T^{-1} corresponds to the Fourier frequencies, a collection of frequencies that contains all the information about a time series if it is sampled at equally-spaced times with times $1, 2, \ldots, T$. There is a limit to the amount of precision that can be attained by making the grid very fine. Consider a grid with spacing $\frac{1}{r}T^{-1}$, in which r is the *oversampling rate* compared to the Fourier frequencies. In changing from some frequency to its neighbor on the grid, the difference in phase between two points will change less than $\frac{1}{r}$. Thus for an oversampling rate of $r = 1000$, the fits at neighboring frequencies will be virtually identical, as the change in relative phase from one method to the other is less than $\frac{1}{1000}$. Nevertheless, the frequency grid should be sufficiently fine not to affect adversely the potentially high precision in the frequency estimates that was discussed in Chapter 2. If a minimum variance bound has been estimated for the problem, as in Section 2.5, the grid spacing should be smaller than the anticipated standard error of frequency.

We propose a two-stage grid minimization scheme. The objective function is first minimized over a grid of frequencies $\frac{1}{r_1}T^{-1}$ apart, with r_1 chosen so that the deepest local minima can be identified. This gives initial estimates of the best local minima of the function. Further grid minimizations are then performed in the vicinity $[v_m - \delta/2, v_m + \delta/2]$ of the M best local minima $v_m,\ m = 1, \ldots, M$ from the initial grid, with spacing $\frac{1}{r_2}T^{-1}$ fine enough not to interfere with the precision of the frequency estimate. This scheme has the advantage of using only the fine grid spacing near the minima, reducing computation time without loss of precision. The quantities that have to be chosen when using this scheme are:

a,b The range of frequencies of interest. For the variable star data, this is known from properties of the stars.

r_1, r_2 The oversampling rates of the initial and fine grids. We found from informal simulations using the MACHO data that $r_1 \geq 6$ was required in order to find the five best local minima of the full function among the best 20 local minima on the grid frequencies. The value r_2 can be chosen by considering the potential precision of the frequency estimate.

M The number of local minima of the function sampled on the initial grid that are refined by the placement of a finer grid. We typically used 20 local minima.

δ The width of the finer grid about the initial estimate of a local minimum. Since the troughs are expected to have width of approximately $2T^{-1}$, a value like T^{-1} is a good choice.

This optimization method has been found quite effective in extensive use on the MACHO data. Previous frequency estimation programs in Astronomy have used single-grid minimization, but we recommend the use of a two-stage minimization to achieve high levels of precision in the frequency estimate without much of a penalty in computation time.

3.3 Example Analyses

We applied the methods of Section 3.1 (except the PDM method) to the example data of Chapter 1, in order to compare their performance. The methods that were used are:

C1 Cosine model estimated by least-squares.

C2,C4,C6 Fourier decomposition method with 2, 4, and 6 terms in the expansion. The 2-term model (5 parameters) was chosen as the minimum model that should correctly identify an eclipsing binary, and the 4-term (9 parameters) and 6-term (13 parameters) models were chosen because they were of moderate and high complexity.

S5,S9,S13 Periodic cubic spline method with 5, 9 and 13 knots over the phase interval [0,1], using the modified knot placement described in Section 3.2. The number of knots were chosen so that the number of parameters was the same as in the Fourier decomposition models to allow direct comparison of these methods.

LF,RN Lafler and Renson measure of dispersion methods.

SM Estimation using Supersmoother.

Optimization of the objective function for each method was done using the two-stage scheme of Section 3.2 with $[a, b] = [T^{-1}, 5]$, $r_1 = 10$, $r_2 = 200$, $M = 20$, and $\delta = T^{-1}$. For a typical data set with 300 observations taken over a 400-day period, this method requires calculation of the objective function at 20,000 frequencies in the first stage and at 4000 frequencies in the second stage. Running this as optimized code on a SPARC-10 workstation gave the following computation times (CPU time in seconds):

C1	C2	C4	C6	S5	S9	S13	LF	RN	SM
100	180	370	620	290	340	360	70	80	370

For each data set and estimation method, the deepest four local minima of the objective function were inspected and the fundamental frequency identified among them.

Summary of Results

- Each of the estimation methods was able to provide an estimate of the fundamental frequency, which was usually found at the global minimum of the objective function. The multiple estimates of the fundamental frequency obtained from the various methods differed by only $0.13\,T^{-1}$, and most estimates fell in a range of length $0.05\,T^{-1}$. As the estimates from the Lafler and Renson methods were virtually identical we present only the Renson results, giving nine estimation methods in all. Since these latter two methods differ only in how they weigh points distant in phase, and these data sets were so large that large differences in phase were rare, it was not surprising that the methods behaved similarly.

- The estimation methods behaved differently with regard to the best frequency estimates that they chose. For strictly periodic and unimodal curves, the cosine method located the fundamental frequency and its pseudo-aliases, the Fourier and spline methods located the fundamental frequency and some subharmonics and aliases, and the nonparametric Renson and supersmoother methods located only the fundamental frequency and its subharmonics. For a strictly periodic but bimodal curve, the results were similar except that most methods also located the first harmonic $2w$. For a semiperiodic curve, the nonparametric models and the high-parameter spline method also located T^{-1}, which corresponds to the period that is the span of the sampling times.

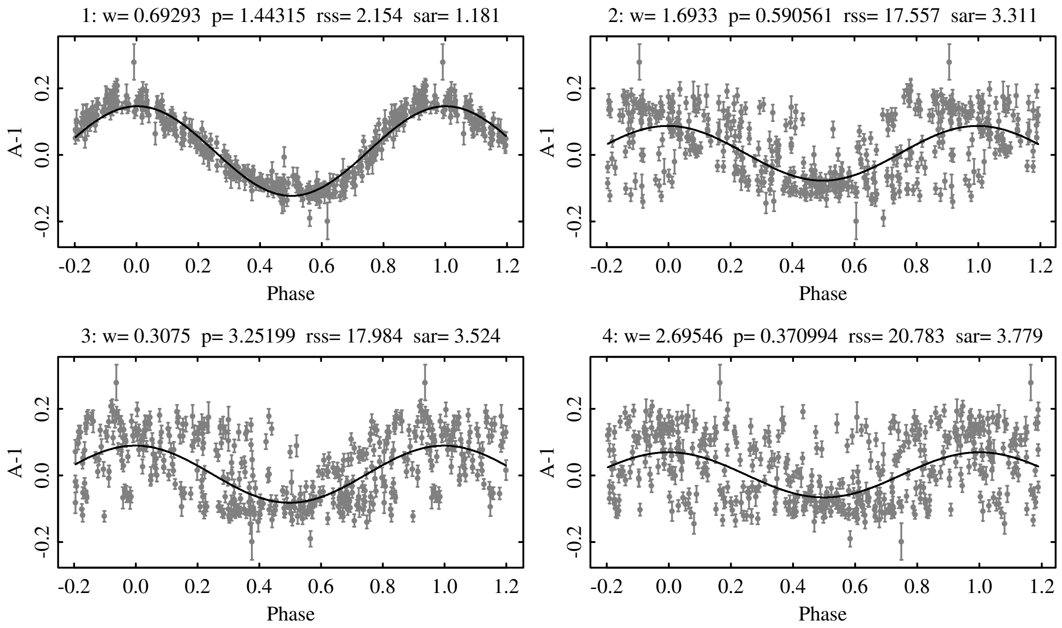

Figure 3.4: Phase plots and fitted values of the red band of star 77021:1992 for the best four frequency estimates using the cosine (C1) method.

Unimodal Periodic Curve

We performed analyses on the red band of the data displayed in Figures 1.1 and 1.2, which is from a cepheid star with an approximately sinusoidal light curve. The fundamental period is approximately 1.443 days and corresponds to frequency $w \doteq 0.693$.

Figure 3.4 shows the phase plots of these data at the best four frequencies chosen by the cosine (C1) method. The upper-left plot is that of the best estimate, the upper-right that of the second-best estimate, and so on. The frequency estimate, corresponding period estimate, weighted RSS, and weighted SAR are displayed along the top of each plot. For this method, the best frequency chosen was the fundamental frequency with estimate $\hat{w} = 0.69293$. The RSS dropped rapidly at the other estimates, identified as the pseudo-aliases $1 + w$, $1 - w$, and $2 + w$.

Table 3.1 shows the corresponding results for all nine estimation methods, namely the estimate of the fundamental frequency $\hat{w}$ and the identification of the best four estimates chosen by that method. By best, we mean the estimates corresponding to the deepest local minima of the objective function. Since the spacing in the fine frequency grid is approximately 0.000012, the frequency estimates are reported to 5 decimal places. The estimates $\hat{w}$ span $0.04\,T^{-1}$ in frequency. The more interesting part of the table is the identification of the best frequency estimates. The

Table 3.1: Estimate of the fundamental frequency and identification of the best four frequency estimates, for nine estimation methods applied to the red band data of star 77021:1992.

		Identification of Estimate			
Method	$\hat{w}$	1st	2nd	3rd	4th
C1	.69293	w	$w+1$	$-w+1$	$w+2$
C2	.69290	w	$w/2$	$w+1$	$w/2+1/2$
C4	.69290	w	$w/2$	$w/4$	$w/3$
C6	.69290	$w/2$	w	$w/3$	$w/5$
S5	.69286	w	$w/2$	$w/2+1/2$	$w+1$
S9	.69290	w	$w/2$	$w/3$	$w/4$
S13	.69290	$w/2$	w	$w/3$	$w/4$
RN	.69284	w	$w/2$	$w/3$	$w/4$
SM	.69288	w	$w/2$	$w/3$	$w/4$

2-term Fourier decomposition model (C2) found the fundamental frequency and first subharmonic, and then pseudo-aliases of these two frequencies. In comparison, the 4-term Fourier method (C4) located only the fundamental frequency and its subharmonics, since it can model curves with up to 4 maxima over the phase interval. The three spline methods behaved similarly to their Fourier counterparts. Remember that we chose the number of knots to match the number of parameters in the Fourier methods. The spline and Fourier methods with the same number of parameters usually located the same frequencies in the same order. The two nonparametric methods, the Renson and supersmoother methods, acted like the higher-parameter Fourier and spline methods in locating the fundamental frequency and its subharmonics.

Similar analyses were done on data from Figures 1.4 and 1.8, which are both periodic with unimodal curve shape. The results did not differ markedly from those above.

Bimodal Periodic Curve

It is more interesting to compare the estimation methods for a more complicated curve shape: the eclipsing binary example of Figure 1.6. Table 3.2 shows the estimates of the fundamental frequency and identification of the best four estimates for the nine estimation methods applied to these data. The estimates $\hat{w}$ span $0.07\,T^{-1}$ in frequency, with an outlying estimate of .40454

Table 3.2: Estimate of the fundamental frequency and identification of the best four frequency estimates, for nine estimation methods applied to the blue band data of star 77043:4317.

		Identification of Estimate			
Method	$\hat{w}$	1st	2nd	3rd	4th
C1	.40463	$2w$	$2w+1$	$-2w+1$	w
C2	.40462	w	$2w$	$2w+1$	$w+1/2$
C4	.40464	w	$w/2$	$2w$	$2w/3$
C6	.40465	w	$w/2$	$w/3$	$2w$
S5	.40454	w	$2w$	$2w+1$	$w+1/2$
S9	.40465	w	$w/2$	$2w$	$2w/3$
S13	.40471	w	$w/2$	$2w$	$w/3$
RN	.40467	w	$w/2$	$w/3$	$w/4$
SM	.40465	w	$w/2$	$w/3$	$2w$

given by the 5-knot spline method (S5). The cosine method chose the first harmonic $2w$ as the best estimate, followed by a couple of pseudo-aliases, with the fundamental frequency appearing only as the fourth-best estimate. The phase plots and fitted values for this method are shown in Figure 3.5. We see that as the cosine function does not provide a good fit to the data at the fundamental frequency, it chooses frequencies that by visual inspection produce a much worse description of the data. Thus the cosine model is a less useful tool for estimating frequency in data of this type.

All of the other estimation methods correctly identified the fundamental frequency as the best estimate, and chose the subharmonics and first harmonics for the other estimates. Figures 3.6 and 3.7 show the phase plots for the best four frequency estimates from the 4-term Fourier method (C4) and the 9-knot spline method (S9), both of which contain nine parameters. Both methods chose the fundamental frequency as the best estimate, and for the next-best estimates chose the first subharmonic $w/2$, the first harmonic $2w$, and the second subharmonic of the first harmonic $2w/3$. The fit to the phased light curve at the fundamental frequency is reasonably good in both cases. There is noticeable oscillation in the fitted curve which could be reduced by using a higher number of parameters, but at the expense of additional computation time. Observing the phase plots from these methods, we do not see that the spline method gives a better fit to the curve than

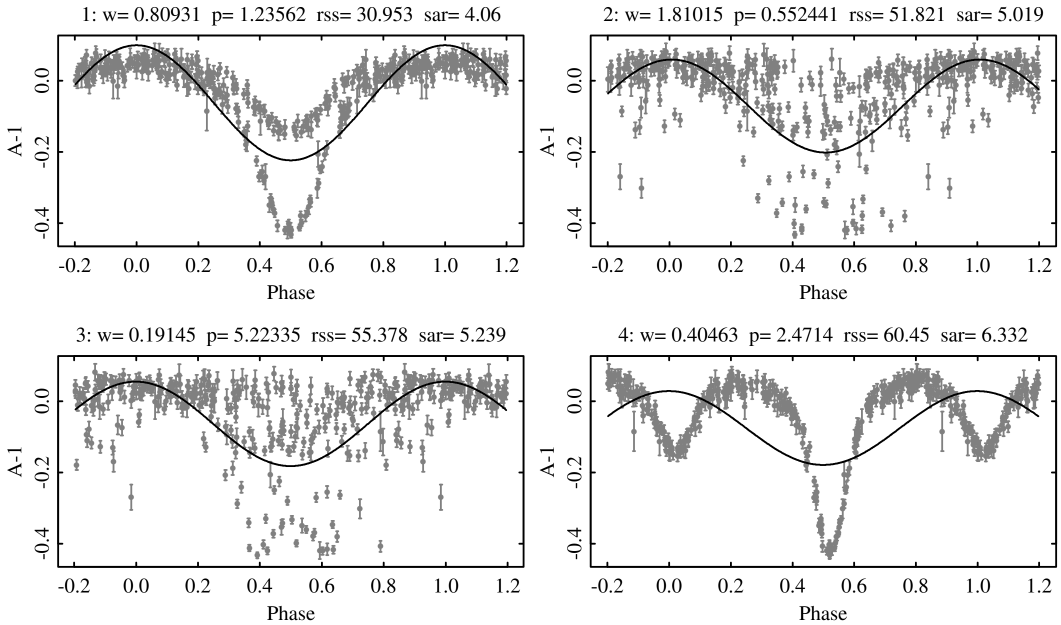

Figure 3.5: Phase plots and fitted values of the blue band of star 77043:4317 for the best four frequency estimates using the cosine (C1) method.

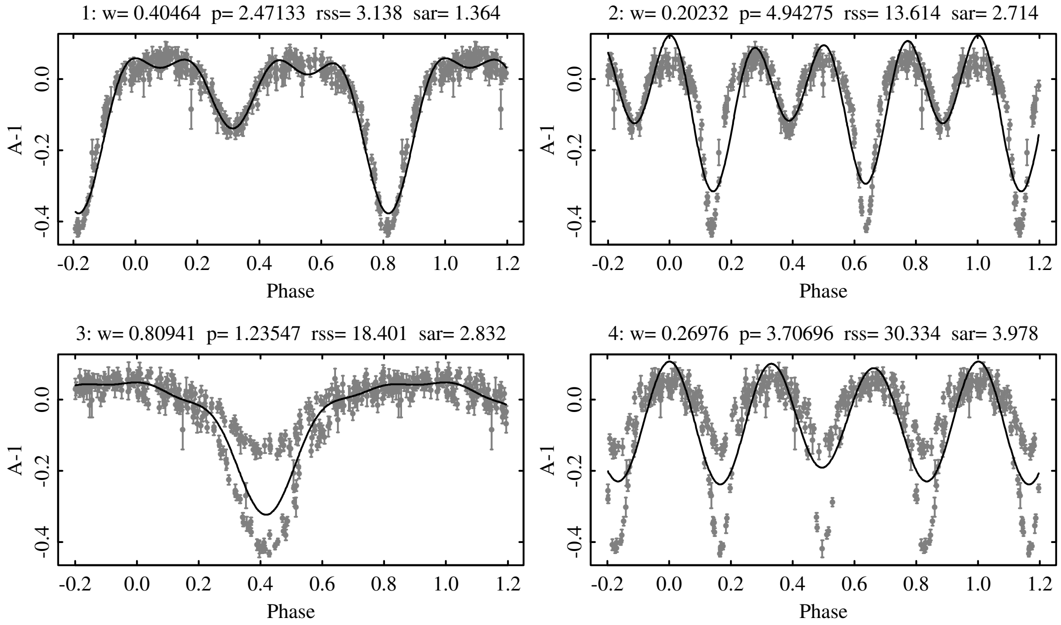

Figure 3.6: Phase plots and fitted values of the blue band of star 77043:4317 for the best four frequency estimates using the 4-term Fourier method (C4).

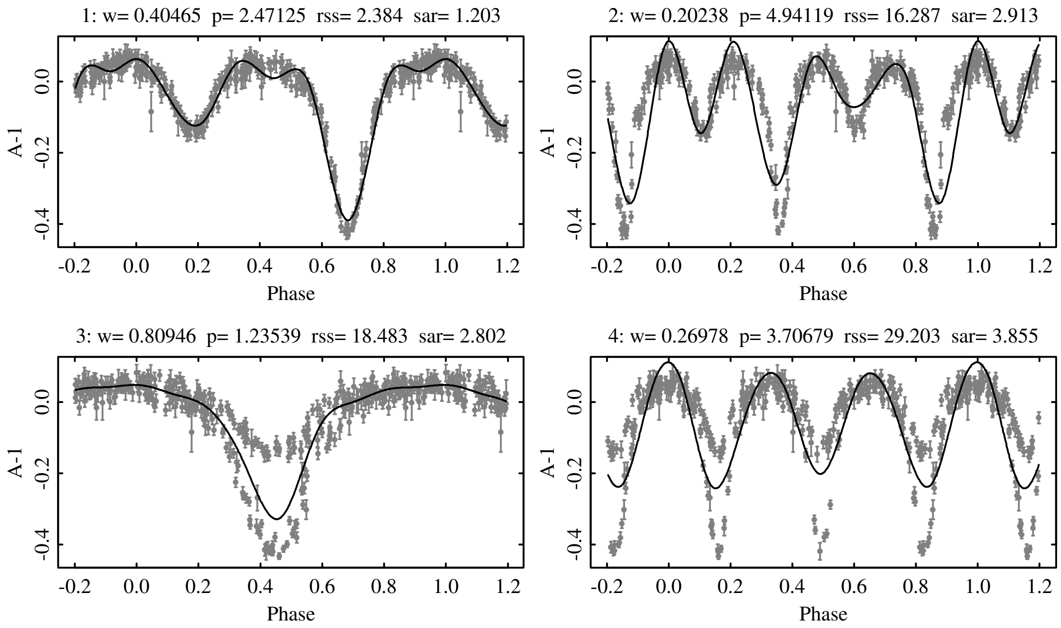

Figure 3.7: Phase plots and fitted values of the blue band of star 77043:4317 for the best four frequency estimates using the 9-knot spline method (S9).

the Fourier method with the same number of parameters as we had supposed. Nevertheless, the spline method does have the advantage of smaller computation times, due to the band-diagonal nature of the regression matrix. Figure 3.8 displays the phase plots of the best frequency estimates using supersmoother. This method has computation times similar to those of the 4-term Fourier and 9-knot spline methods, but the fitted curve is a much better description of the data at the fundamental frequency, being less oscillatory and having a lower RSS. Supersmoother gives a less satisfactory fit at the first harmonic, shown in the lower-right plot of the figure. The fitted curve is similar to that from the Fourier and spline methods, but it not as smooth at the local level. This is perhaps an unfair comparison, as no fitted curve can successfully describe the data at this frequency.

The eclipsing binary example illustrates the practical benefits of using methods that are able to successfully model curves of non-sinusoidal shapes. In section 3.4, we shall see that there is also a penalty in precision for using the cosine methods to model data with a non-sinusoidal curve.

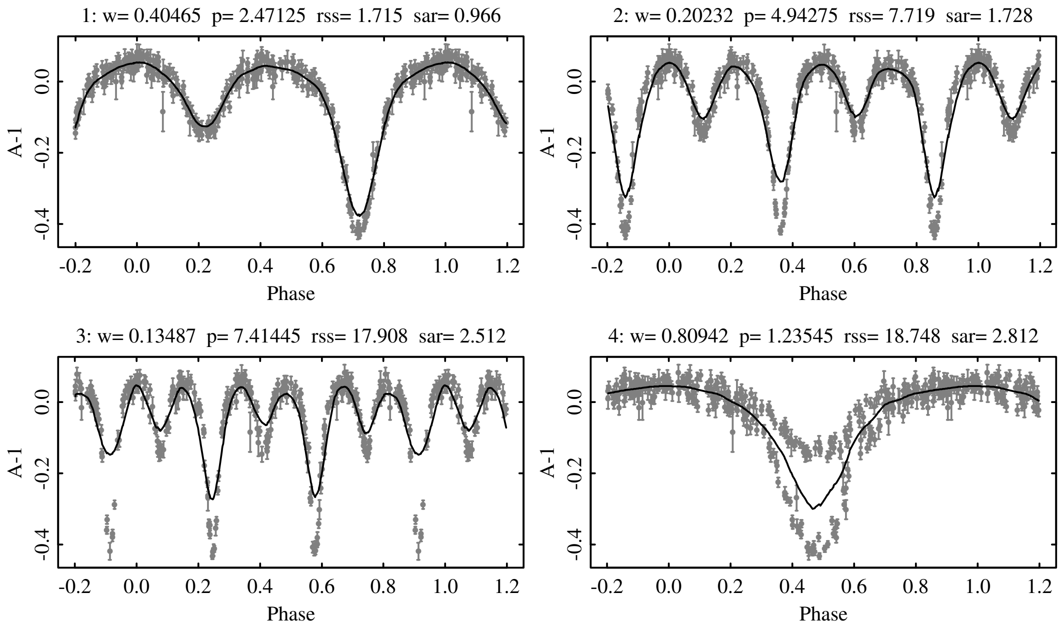

Figure 3.8: Phase plots and fitted values of the blue band of star 77043:4317 for the best four frequency estimates using the supersmoother method (SM).

Semiperiodic Curve

The final example is that of a semiperiodic long-period variable star which was shown in Figure 1.9. The methods that we are using are intended for use with strictly periodic functions, but as semiparametric data occurs commonly in practice, it is instructive to see how the methods behave in this situation. Looking at the data, we see six maxima over the 400-day observation period, suggesting an approximate period of 80 days. Estimates of this frequency and identification of the best four frequency estimates for the nine methods are shown in Table 3.3. The estimates span 0.13 T^{-1} in frequency space, with an outlying estimate of 0.01258 from the Renson method. The major difference between this example and the previous ones is that the frequency corresponding to the span of the sampling times is located by the nonparametric methods and the high-parameter spline method. Figure 3.9 shows the phase plots of the four best estimates using the supersmoother method on these data. Note that the first frequency chosen corresponds to the entire data span, the second is the first subharmonic, and only the third is the fundamental frequency. This could be considered either a strength or a weakness of these methods. If the main objective is estimation of the approximate periodicity, these methods are less suitable, as they prefer to fit a complicated multimodal curve through long stretches of the data than to fit a simple but poorly-fitting curve

Table 3.3: Estimate of the fundamental frequency and identification of the best four frequency estimates, for nine estimation methods applied to the red band data of star 78017:497. Note: There are three estimates that are unable to be identified with a harmonic or pseudo-alias of the fundamental frequency. These are $a \doteq 1.0182$, $b \doteq 0.9943$, and $c \doteq 1.0083$.

		Identification of Estimate			
Method	$\hat{w}$	1st	2nd	3rd	4th
C1	.01285	w	$-w+1$	a	b
C2	.01282	$w/2$	w	$-w+1$	$-w/2+1/2$
C4	.01278	$w/2$	$w/3$	$w/4$	w
C6	.01284	$w/4$	$w/2$	w	c
S5	.01290	$w/2$	w	$-w+1$	$-w/2+1/2$
S9	.01282	$w/4$	$w/2$	w	c
S13	.01289	T^{-1}	$w/2$	w	c
RN	.01258	T^{-1}	w	$w/2$	b
SM	.01289	T^{-1}	$w/2$	w	c

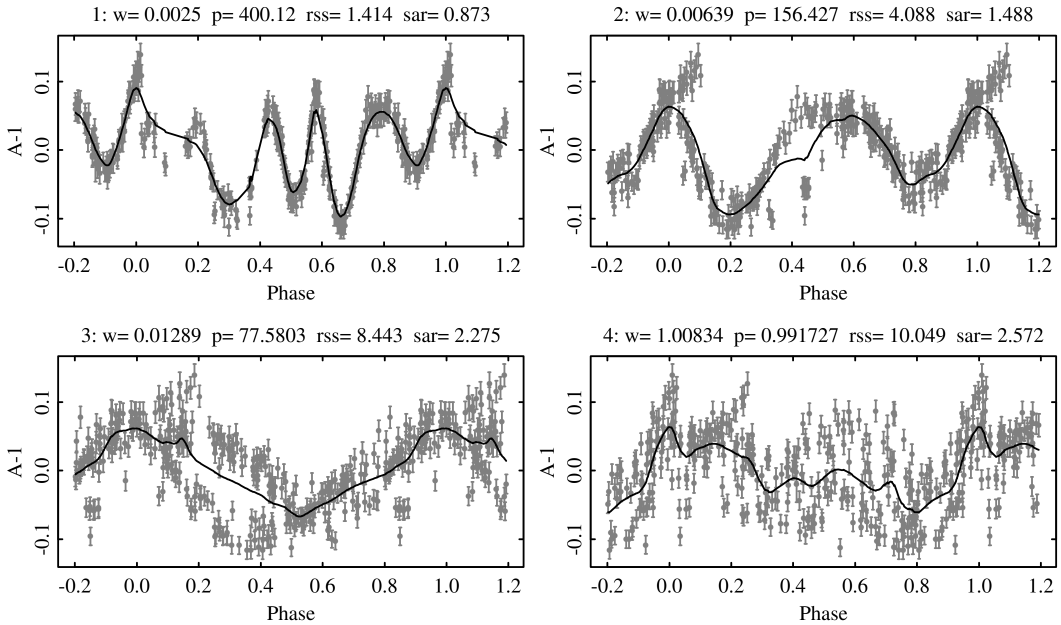

Figure 3.9: Phase plots and fitted values of the red band of star 78017:497 for the best four frequency estimates using the supersmoother method.

through a short stretch of the data. The same reason makes these methods more suitable if the objective is the most accurate but not necessarily periodic representation of the data. There are also a number of frequency estimates that were not able to be identified as relatives of the 80-day or 400-day periods, such as that of the fourth-best estimate in Figure 3.9. We suppose that these are by-products of the semi-periodicity of the curve.

Conclusions

The choice of method for estimating frequency in practice depends on the particulars of the estimation problem. The methods we used differ greatly in computation time, and that alone may determine which methods are feasible. For curves that are periodic and unimodal over the phase length, all of these methods worked well in estimating the frequency, but differed in the fitted curve that was produced (or whether they produced a fitted curve at all). Some of the higher-parameter models have the annoying property of occasionally choosing a subharmonic over the fundamental frequency, since they can model the curve at these frequencies equally well. If periodic function with multimodal curves over the phase length are expected, the cosine method is less convenient, as are the Fourier and spline methods with too few parameters to approximate the curve at the fundamental frequency. It is not clear which methods are most appropriate for semiperiodic data; the choice of method depends on the objectives of the period search.

3.4 Simulation Study

In the previous section we saw how a number of estimation methods performed on some example data with respect to location of the fundamental frequency and the quality of the fitted curve. It is also useful to know what kind of precision can be expected from these methods. We present here results of a simulation study which examines the precision of these methods on some typical datasets, and compares the variance of the frequency estimates with an estimated variance bound.

Study Description

Three models were used in the simulation: a unimodal but non-sinusoidal curve, a bimodal curve like that of an eclipsing binary star, and a curve with missing information due to uneven sampling in phase space. These models were chosen to represent common data types

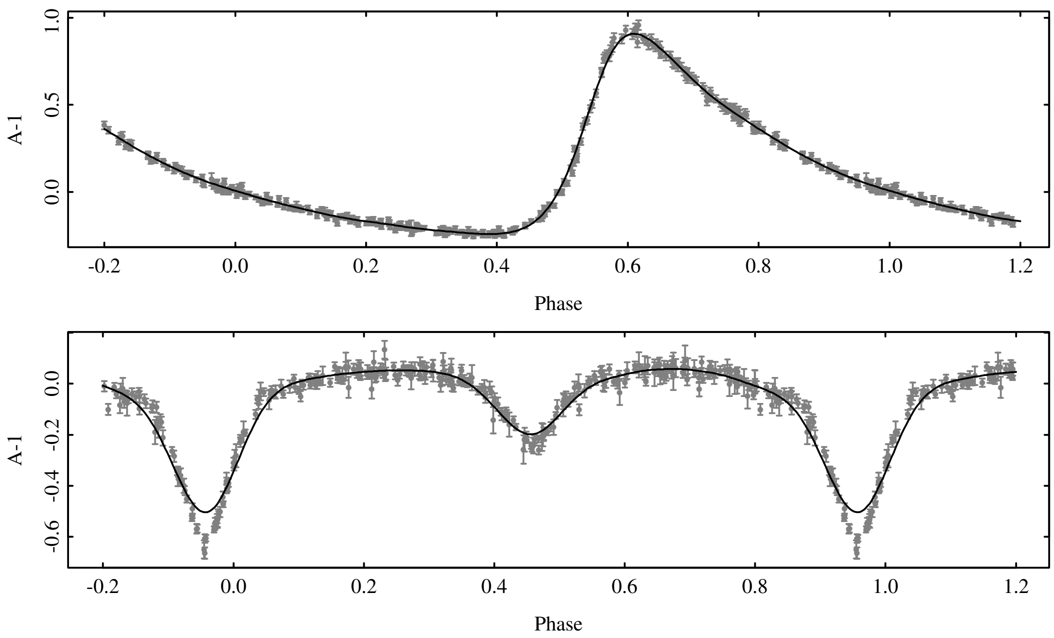

Figure 3.10: Phase plots of two of the MACHO light curves with fitted curves from a periodic smoothing spline. The upper plot is of the blue band of star 77017:379 at period 4.017 days, and the lower plot is of the red band of the star 77048:2523 at period 2.917 days.

observed in the MACHO data. Two levels of noise variance were used with each of these models to mimic typical high-noise and low-noise light curves in the MACHO data.

The curve shapes were obtained by fitting a smoothing spline to the phase plots of two of the MACHO star data sets. Smoothing splines differ from regression splines by having knots at each of the data points and by estimating the parameters through minimization of the sum of the RSS and a term which penalizes roughness in the fitted curve; see de Boor [15], Ch. 14, for more details. The raw data and fitted curves are displayed in Figure 3.10. The data in the upper plot comes from the blue band of star 77017:379, a relatively bright cepheid variable star with a fundamental period of approximately 4.017 days. The fitted curve was produced by fitting the data with a smoothing spline that had knots at all of the data points and smoothing parameter (Spar) of 6.7 e-6. This was approximately equivalent to a 35-parameter fit, as the trace of the regression projection matrix was 34.955. The data in the lower plot come from the red band of star 77048:2523, an eclipsing binary star with fundamental frequency of about 2.917 days. The fitted curve is from a smoothing spline with smoothing parameter 1.4 e-5, approximately equivalent to a 30-parameter fit. We decided to sacrifice quality of fit at the bottom of the troughs in order to maintain a smooth-shaped curve at the plateaus. The fitted curves were shifted and scaled to have

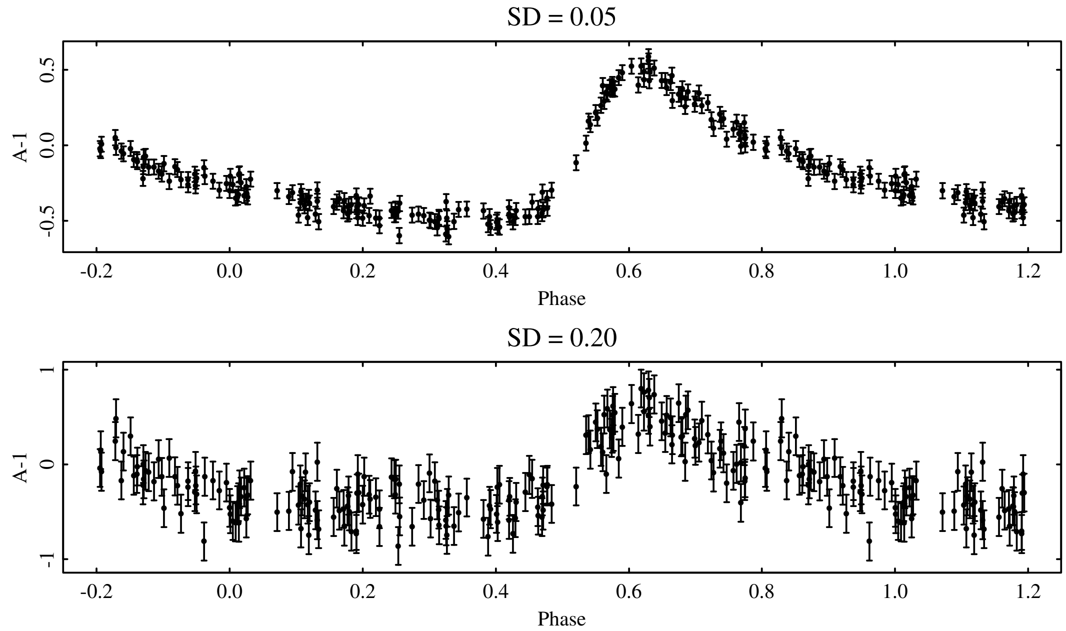

Figure 3.11: Phase plots of simulated light curves from model A at period 1.4432 days.

amplitude one, with maximum value of 0.5 and minimum value of -0.5. The same sampling times were used for all the simulated data. They were a collection of 200 times spanning 243 days taken from the red band of the data displayed in Figure 1.6.

The study used three combinations of curve shape and frequency, models A, B, and C. Model A used the fitted values from the cepheid curve at period 1.4432 days, equivalent to frequency 0.69290466. The frequency was chosen so that the observations covered the phase interval approximately uniformly. Two levels of noise variance were used: a low-noise model in which the errors were IID normal with SD=0.05, and a high-noise model with SD=0.20. Figure 3.11 shows typical simulated data from model A at the two noise levels. The upper plot is typical of many of the cepheid light curves in the MACHO data and the lower plot looks like a typical RR Lyrae light curve. Model B uses the fitted curve from the eclipsing binary data with period 1.4432 days. Typical simulated data from this model for the two levels of noise variance are shown in Figure 3.12. Model C uses the fitted curve from the cepheid data with frequency 0.5. As this is a 2-day period, data is only available over about half of the phase interval. This frequency was chosen to see how the estimation methods behaved when there were large gaps in the phased data, and to compare the precision attained in this model with that in model A, which has the same curve shape but a different frequency. Typical simulated data from model C at the two levels of noise

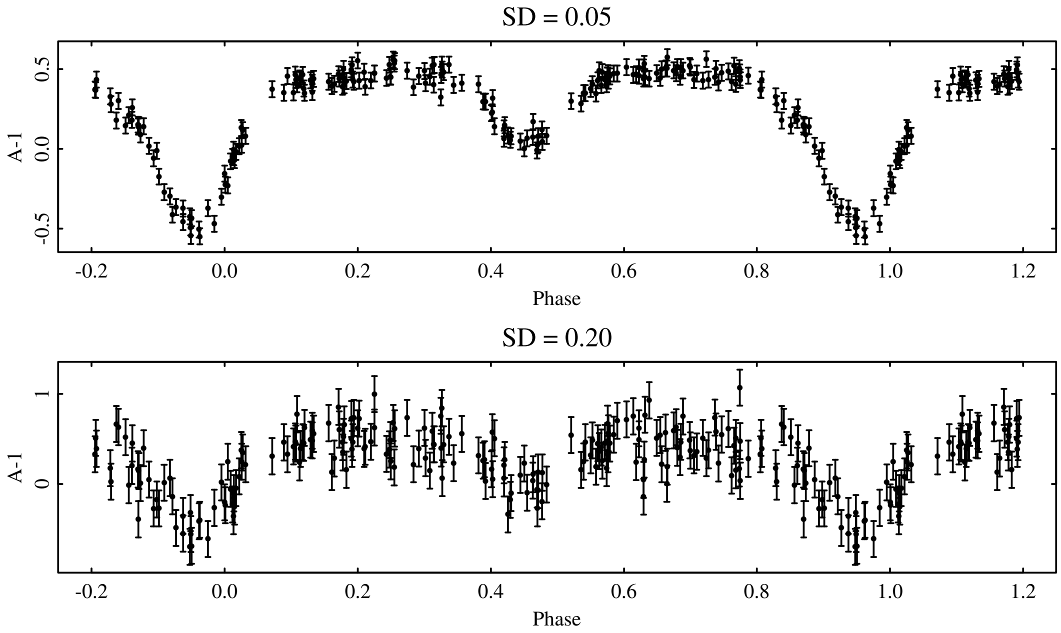

Figure 3.12: Phase plots of simulated light curves from model B at period 1.4432 days.

variance are shown in Figure 3.13.

It is useful to know what order of precision is expected for these models. Combining the ideas in Sections 2.5.1 and 2.5.2, we propose the variance bound

$$\text{Var}[\hat{w}] \geq \frac{n^{-1}\sigma^2}{\text{Var}[t]\text{E}[s'(wt)^2]}, \tag{3.3}$$

in which σ^2 is the variance of the observational noise, Var$[t]$ is the sample variance of the observation times, and E$[s'(wt)^2]$ is the average squared-derivative of the curve at the observation times. The sample variance of the 200 times in these models is 5775, and the average squared-derivative of the curve at the observations is 8.5 for model A, 16.9 for model B, and 3.1 for model C. Thus the estimated variance bounds for the three models and two noise levels are as follows:

	Model A	Model B	Model C
$\sigma = 0.05$	2.5 e-10	1.3 e-10	7.0 e-10
$\sigma = 0.20$	41 e-10	20 e-10	110 e-10

This means that the smallest standard error we expect for the frequency estimate is approximately 1 e-5.

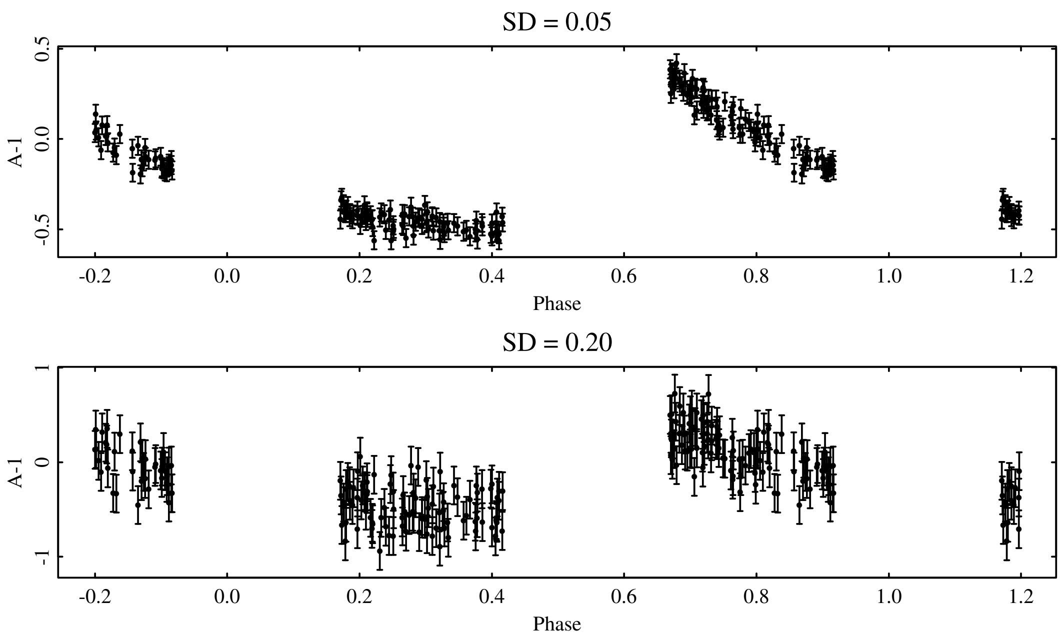

Figure 3.13: Phase plots of simulated light curves from model C at period 2 days.

The estimation methods used in this study were the same as those of Section 3.3, with the addition of the periodogram method (PG). One hundred simulated data sets were generated for each of the combinations of model and noise level. The objective function of each method was minimized over a grid of 4000 frequencies spanning the interval $[w - 0.25\,T^{-1}, w + 0.25\,T^{-1}]$, in which w is the known frequency and $T = 243.12$ is the span of the sample times. The grid spacing was chosen to be about $1/20$ of 0.00001, the smallest anticipated standard error of the frequency estimate.

Summary of Results

- Relative performance of the estimation methods depended strongly on the model and noise level. The Fourier methods consistently performed the best, as measured by the mean squared-error (MSE) of the frequency estimate. Next best were the supersmoother and spline methods, and the least-precise methods were the cosine methods and the Lafler and Renson measures of dispersion.

- The 4-term and 6-term Fourier methods had the highest precision, closely followed by the 2-term method. The MSE from these methods ranged from being approximately equal to the estimated variance bound to three times the variance bound.

- The next most successful method was supersmoother, although its performance was not consistent over the models and noise levels. Supersmoother did well when the periodic curve had a complicated shape or when the noise was small, but did poorly under high observational noise.

- The spline methods performed slightly worse than supersmoother over the simulation models, but were not as adversely affected by high noise. Each of the spline methods performed worse than the Fourier method with the same number of parameters, although the spline models were more competitive under model C. The results were not consistent for spline models with different numbers of knots; it is possible that this erratic behavior is caused by the non-smoothness in the RSS curve that was discussed in Section 3.2.

- The least-precise methods were the cosine methods (the periodogram and least-squares versions) and the Lafler and Renson measures of dispersion. The cosine methods suffered from high bias (especially the periodogram) and high variance, while the dispersion methods had low bias but high variance. The Lafler and Renson methods performed almost identically, but the Renson method was a little more accurate under model C.

Comments

- It is surprising that the Fourier methods performed significantly better than the spline methods, considering that they had the same number of parameters and showed similar fits in the examples of Section 3.3. It would be interesting to see if similar results are obtained when using smoothing splines instead of regression splines.

- There is a definite advantage to using methods that fit the periodic function more closely than the cosine methods.

- The Lafler and Renson measure of dispersion methods were not competitive, due to their high variance. In a comparative study of a number of estimation methods commonly used in astronomy, Heck, Manfroid, and Mersch [23] found that the Renson and Stellingwerf methods performed slightly better than the others. This suggests that the more precise but computationally intensive Fourier expansion method could be used to improve the quality of period estimation in astronomical data. It should be noted, however, that our simulations used data sets considerably larger than those commonly used in variable star analysis, and our conclusions might not generalize to the general case.

- As the estimated variance bound was a good guide to the best precision attained in Models A and B, and was a lower bound for the precision in Model C, it can be useful in choosing a global optimization strategy for the estimation.

Model A

Table 3.4 is a summary of the results of the frequency estimation using the model A simulated data. It presents the bias of the frequency estimate divided by its estimated SE, the variance of the frequency estimate, and the mean squared-error, for the eleven estimation methods and the two levels of observational noise. Boxplots of the frequency estimates are displayed in Figure 3.14.

For the lower level of noise variance, the methods with the lowest MSE were the three Fourier methods, the spline method with 13 knots, and the supersmoother method. The most precise of these was the 6-term Fourier method which had an MSE of 4.0 e-10, 1.6 times the estimated variance bound of 2.5 e-10. The methods with the highest MSE were the periodogram and the 5-knot and 9-knot spline methods; this was caused by large bias rather than large variance. [The 5-knot method had especially high bias compared to its low variance.] The Lafler and Renson methods also had low bias, but as they had the largest variance of any of the methods, they were not competitive. As the number of parameters increases within a given method, one would expect the bias to decrease and perhaps the variance to increase. At first glance this is not verified for the Fourier methods, but looking at the raw biases, 1.27 e-5, 1.4 e-5, 2.1 e-5, and 8.7 e-6 for the 1, 2, 4, and 6-term expansions, then there is some indication that the bias is decreasing as parameter size increases. In the spline methods, the bias seems to be decreasing and the variance seems to be increasing as the number of knots increases.

For the higher variance case, the methods with the lowest MSE were the Fourier methods, followed by the spline methods. The lowest MSE, 5.6 e-9 for the 4-term Fourier method, was 1.4 times the estimated variance bound of 4.1 e-9. Thus for both levels of noise, the estimated variance bounds seem to be a good estimate for the attainable precision under the model. The Lafler and Renson methods had the highest MSE due to their very high variance, while the spline models had the largest bias.

Table 3.4: Standardized bias, variance, and MSE of the frequency estimates for eleven estimation methods on the model A simulated data.

	SD = 0.05			SD = 0.20		
Method	Bias/SE	Variance $\times 10^{-10}$	MSE $\times 10^{-10}$	Bias/SE	Variance $\times 10^{-10}$	MSE $\times 10^{-10}$
PG	-16.7	6.4	24.1	-3.9	94	107
C1	-8.7	9.4	16.4	-1.7	129	132
C2	6.9	4.4	6.4	0.8	59	59
C4	-12.1	2.9	7.2	-3.9	50	56
C6	-4.8	3.3	4.0	-2.3	69	72
S5	68.5	1.2	59.3	11.3	33	76
S9	-27.7	4.5	39.0	-8.2	51	85
S13	-4.7	6.3	7.6	-2.5	85	89
LF	-0.7	16.6	16.5	0.0	257	254
RN	-0.7	16.6	16.5	0.0	256	254
SM	-1.9	6.3	6.4	-1.2	141	142

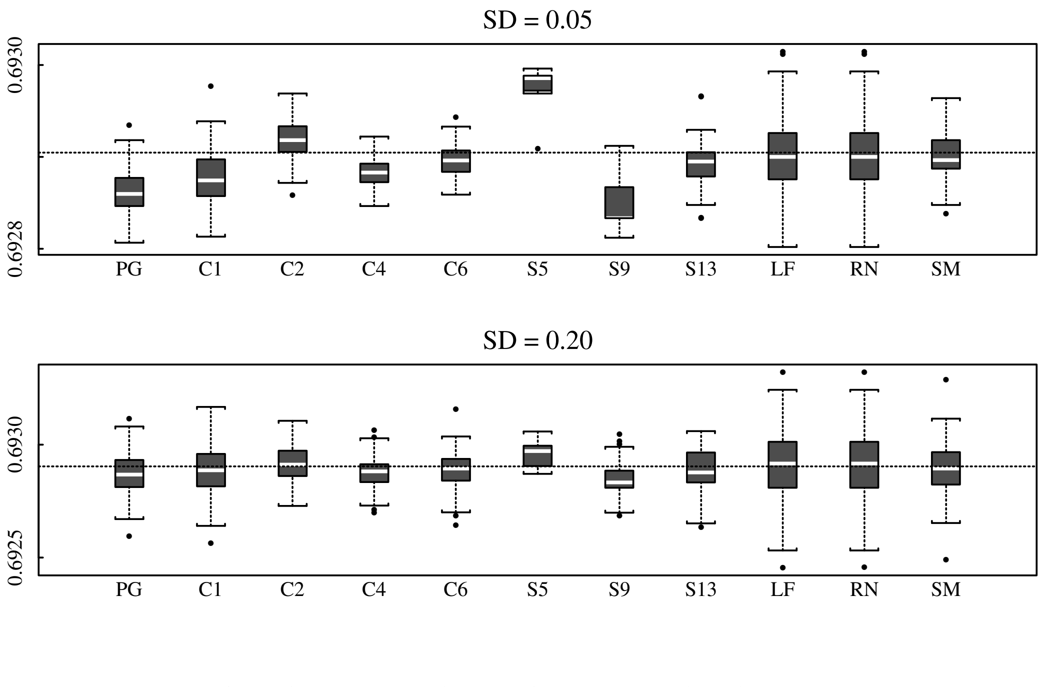

Figure 3.14: Boxplots of the frequency estimates for eleven estimation methods on the model A simulated data.

Model B

Table 3.5 summarizes the simulation results using model B, and Figure 3.15 displays the boxplots of the frequency estimates. Under the low-noise model, the methods with the lowest MSE were the 4-term Fourier, 6-term Fourier, and supersmoother methods, followed by the Lafler and Renson methods. The best Fourier method has an MSE of 1.1 e-10 which was a little smaller than the estimated variance bound of 1.3 e-10. The cosine models fared poorly for this model, probably because the cosine curve is such a poor approximation of the bimodal shape in the data. The periodogram was especially bad, with MSE 1500 times as large as under the best method. The spline methods also did not perform well, having high bias and high variance.

In the high-noise case, the Fourier, 13-knot spline, and supersmoother methods had the lowest MSE, with the lowest value of 2.5 e-9 for the 6-term Fourier method being 1.25 times as large as the estimated variance bound. Again the cosine methods performed badly: the periodogram had MSE 100 times as large as the best method, mostly due to high variance.

Model C

Table 3.6 shows the results of the model C simulations, and boxplots of the frequency estimates are displayed in Figure 3.16. For the low-noise case, the lowest-MSE methods were the 4-term and 6-term Fourier methods and the high-parameter spline methods. The 4-term Fourier method had MSE of 1.2 e-9, 1.7 times larger than the estimated variance bound. Although the higher-parameter Fourier and spline methods performed well, the 2-term Fourier and 5-knot spline methods had surprisingly large bias. The cosine, Lafler, and Renson methods all performed poorly in comparison to the best method. Note the difference in MSE between the Lafler and Renson methods. In models A and B these methods behaved identically, but in this model, the Renson method performed better. This is unsurprising, as the Renson method was intended to perform well when the phase values are unevenly distributed on [0,1], as is the case for this model.

The results for the high variance model are a little confusing. The methods with the lowest MSE were the periodogram, Fourier, and lower-parameter spline methods. None of the methods came close to the estimated variance bound of 1.1 e-8, but whether the methods are inefficient or the bound is inappropriate is unclear. For the low-noise and high-noise cases, the periodogram had very low variance but high bias. This may be related to the fast rate of convergence derived for the $w = 0.5$ case in Hannan [22]. Again the Renson method did better than the Lafler method, but neither were competitive. Supersmoother performed the worst of all these methods, mostly due to

Table 3.5: Standardized bias, variance, and MSE of the frequency estimates for eleven estimation methods on the model B simulated data.

Method	SD = 0.05			SD = 0.20		
	Bias/SE	Variance $\times 10^{-10}$	MSE $\times 10^{-10}$	Bias/SE	Variance $\times 10^{-10}$	MSE $\times 10^{-10}$
PG	36.5	111.4	1597.0	8.8	1383	2433
C1	9.0	47.8	86.2	0.8	688	685
C2	11.1	2.2	4.8	5.0	45	55
C4	9.5	1.7	3.2	3.9	38	43
C6	-0.2	1.1	1.1	2.3	24	25
S5	4.6	15.4	18.6	8.4	87	146
S9	14.5	23.4	72.2	5.8	145	193
S13	12.4	10.0	25.3	2.7	56	59
LF	-0.2	4.5	4.5	1.4	102	103
RN	-0.1	4.5	4.5	1.4	102	103
SM	-2.5	2.2	2.3	0.5	55	55

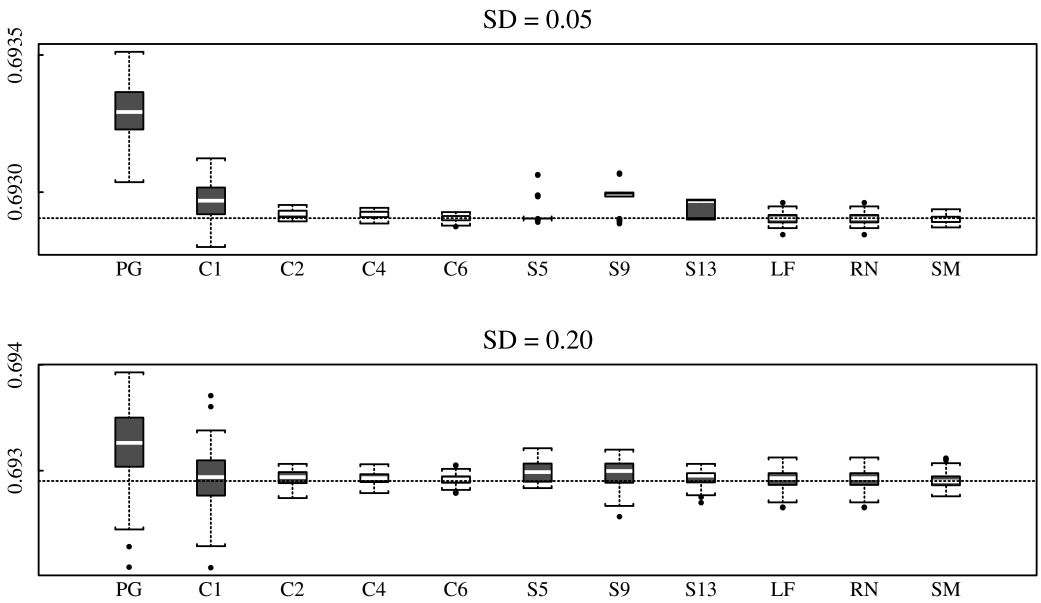

Figure 3.15: Boxplots of the frequency estimates for eleven estimation methods on the model B simulated data.

Table 3.6: Standardized bias, variance, and MSE of the frequency estimates for eleven estimation methods on the model C simulated data.

Method	SD = 0.05			SD = 0.20		
	Bias/SE	Variance $\times 10^{-10}$	MSE $\times 10^{-10}$	Bias/SE	Variance $\times 10^{-10}$	MSE $\times 10^{-10}$
PG	-117.5	2.2	305.7	-25.4	49	364
C1	-31.2	15.6	167.2	-6.1	368	501
C2	-41.8	8.6	157.6	-11.2	151	338
C4	-3.6	11.1	12.4	-1.1	367	368
C6	-0.2	15.4	15.2	-0.9	468	467
S5	-54.4	7.5	229.6	-8.0	265	432
S9	3.1	17.8	19.3	-0.3	454	450
S13	1.4	22.5	22.7	0.3	641	635
LF	0.5	153.0	151.8	-0.2	1015	1005
RN	-0.7	85.9	85.5	-0.4	848	841
SM	0.2	38.6	38.2	1.3	1004	1010

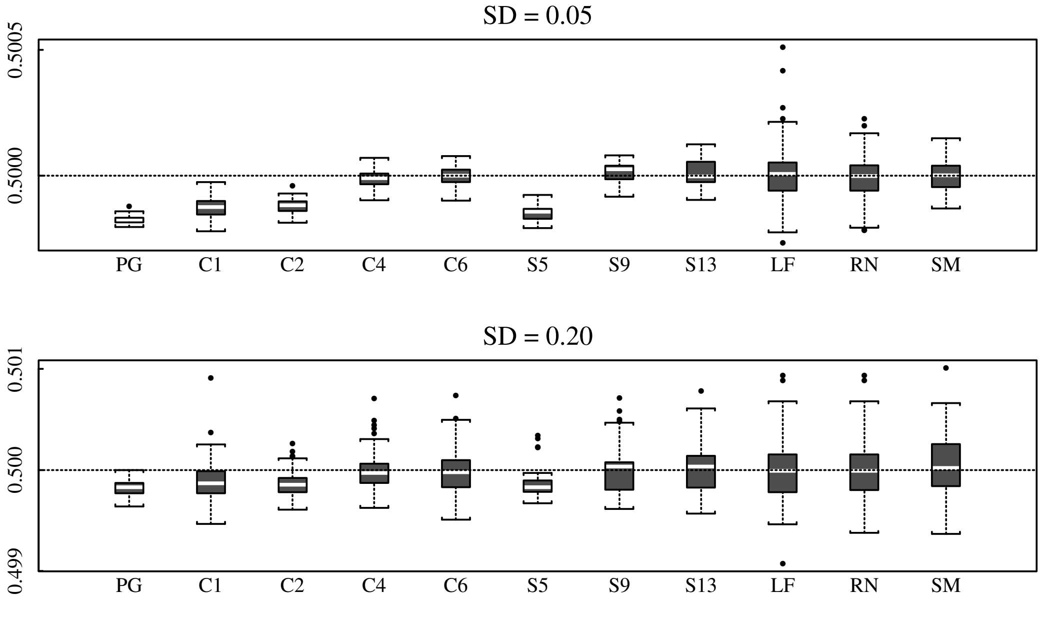

Figure 3.16: Boxplots of the frequency estimates for eleven estimation methods on the model C simulated data.

high variance.

3.5 Further Topics

Three aspects of the frequency estimation problem that we did not pursue in this research are estimation of multiple periodicities, evolution of the phase and amplitude over time, and tests of significance of the estimated frequency.

Estimation of multiple periodicities was discussed in a tangential fashion in Sections 2.2.4 and 2.3.3, but there are many other interesting issues. The estimation problem is much more complicated when several frequencies are estimated together. This can be simplified by estimating and subtracting the largest component first, then estimating and subtracting the remaining periodic components in turn. This method gives inferior estimates of the frequencies, however, and better results can be obtained by using a cyclic descent method (Bloomfield [8], pp. 20-25 , and McDonald [30], p. 673). In Fourier analysis of unequally-spaced time series, deconvolution methods can be used to improve the estimation of the secondary periodic components (Schwarz [41]).

Many phenomena have semiperiodic behavior which is not well-representable by a harmonic component at a given frequency, due to changes in the phase and amplitude over time. The long period variable star of Figure 1.9 is a good example of this. Estimation of changing phase and amplitude for equally-spaced data can be achieved by *complex demodulation* (Bloomfield [8], Tukey [45]), while MacDonald [30] proposes an estimation method for unequally-spaced times.

A pure noise time series may show large peaks in the periodogram or deep troughs in the RSS, and tests of significance are needed to check that an estimated frequency is not the spurious product of the observational noise and the spacing in the sample times. Fisher [19] derived a test of significance of the largest peak in the periodogram for equally-spaced data, and this work has been extended by many authors (Shimshoni [42], Brockwell and Davis [13], pp. 324-332.). There is little in the literature for unequally-spaced times, however. For the MACHO data, it is not relevant to test the hypothesis of a periodic component against that of a pure noise series, since the variable stars have been screened from the pure noise series on the basis of RSS about their mean. Thus all of the series contain a systematic component, though not necessarily a periodic one. It would be more useful to quantify the degree of periodicity in the data, on a scale from periodic to semiperiodic to nonperiodic, but how this could be done is uncertain.

Bibliography

[1] C. Akerlof, C. Alcock, R. Allsman, T. Axelrod, D.P. Bennett, K.H. Cook, K. Freeman, K. Griest, S. Marshall, H-S. Park, S. Perlmutter, B. Peterson, P. Quinn, J. Reimann, A. Rodgers, C.W. Stubbs, and W. Sutherland (1994) Application of cubic splines to the spectral analysis of unequally spaced data. To be published in *Astrophysical Journal*.

[2] H. Akaike (1960) Effect of timing error on the power spectrum of sampled-data. *Annals of the Institute of Statistical Mathematics* **11**, 145-165.

[3] O.E. Barndorff-Nielsen and D.R. Cox (1989) *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, London.

[4] F.J. Beutler (1970) Alias-free randomly timed sampling of stochastic processes. *IEEE Transactions on Information Theory* **16:2**, 147-152.

[5] R.N. Bhattacharya and R.R. Rao (1976) *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

[6] P.J. Bickel, C.A.J. Klaassen, Y. Ritov, and J.A. Wellner (1993) *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins, Baltimore.

[7] P. Billingsley (1986) *Probability and Measure.* Wiley, New York.

[8] P. Bloomfield (1976) *Fourier Analysis of Time Series: An Introduction.* Wiley, New York.

[9] L. Breiman and A. Cutler (1993) A deterministic algorithm for global optimization. *Mathematical Programming* **58**, 179-199.

[10] D.R. Brillinger (1975) *Time Series: Data Analysis and Theory.* Holt-Rinehart, New York.

[11] D.R. Brillinger (1986) Regression for randomly sampled spatial series: the trigonometric case. *Annals of Applied Probability* **23**, 275-289.

[12] D.R. Brillinger (1994) An application of statistics to meteorology: estimation of motion. *Festschrift for Lucien Le Cam.* (eds. D. Pollard and G. Yang) Springer, New York.

[13] P.J. Brockwell and R.A. Davis (1987) *Time Series: Theory and Methods.* Springer-Verlag, New York.

[14] R. Cogburn and H.T. Davis (1974) Periodic splines and spectral estimation. *Annals of Statistics* **2:6**, 1108-1126.

[15] C. de Boor (1978) *A Practical Guide to Splines.* Springer-Verlag, New York.

[16] T.J. Deeming (1975) Fourier analysis with unequally-spaced data. *Astrophysics and Space Science* **36**,137-158.

[17] L.C.W. Dixon and G.P. Szegö (1978) *Towards Global Optimization 2.* North-Holland, Amsterdam.

[18] M.M. Dwortesky (1983) A period-finding method for sparse randomly spaced observations or "How long is a piece of string?", *Monthly Notices of the Royal Astronomical Society* **203**, 917-924.

[19] R.A. Fisher (1929) Tests of significance in harmonic analysis. *Proceedings of the Royal Society A* **125**, 54-59.

[20] J.H. Friedman (1984) A Variable Span Smoother. Technical Report No. 5, Laboratory for Computational Statistics, Department of Statistics, Stanford University.

[21] A.W. Fullerton (1986) Searching for periodicity in astronomical data. *The Study of Variable Stars Using Small Telescopes.* ed. J.R. Percy, Cambridge University Press, Cambridge.

[22] E.J. Hannan (1973) The estimation of frequency. *Journal of Applied Probability* **10**, 510-519.

[23] A. Heck, J. Manfroid, and G. Mersch (1985) On period determination methods. *Astronomy and Astrophysics Supplement Series* **59**, 63-72.

[24] I.A. Ibragimov and R.Z. Has'minskii (1981) *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York.

[25] Y. Isokawa (1983) Estimation of frequency by random sampling. *Annals of the Institute of Statistical Mathematics A* **35**,201-213.

[26] A.V. Ivanov (1979) A solution of the problem of detecting hidden periodicities. *Theory of Probability and Mathematical Statistics* **20**, 51-68.

[27] Y.A. Kutoyants (1984) *Parameter Estimation for Stochastic Processes.* Heldermann Verlag, Berlin.

[28] J. Lafler and T.D. Kinman (1965) An RR Lyrae survey with the Lick 20-inch astrograph. II. The calculation of RR Lyrae periods by electronic computer. *Astrophysical Journal Supplement Series* **11**, 216-222.

[29] N.R. Lomb (1976) Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science* **39**, 447-462.

[30] J.A. McDonald (1986) Periodic smoothing of time series. *SIAM Journal on Scientific and Statistical Computing* **7:2**,665-688.

[31] R.H. Mladineo (1986) An algorithm for finding the global maximum of a multimodal, multivariate function. *Mathematical Programming* **34**, 188-200.

[32] W.H. Press and G.B. Rybicki (1989) Fast algorithm for spectral analysis of unevenly spaced data. *Astrophysical Journal* **338**, 277-280.

[33] B.G. Quinn (1989) Estimating the number of terms in a sinusoidal regression. *J. Time Series Analysis* **10**, 12-31.

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

[35] P. Renson (1978) Méthode de recherche des périodes des étoiles variables. *Astronomy and Astrophysics* **63**, 125-129.

[36] J.D. Scargle (1982) Studies in astronomical times series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *Astrophysical Journal* **263**, 835-853.

[37] B.O. Shubert (1972) A sequential method seeking the global maximum of a function. *SIAM Journal of Numerical Analysis* **9(3)**, 379-388.

[38] A. Schuster (1897) On the investigation of hidden periodocities with application to a supposed 26 day period of meterological phenomena. *Terr. Magn.* **3**, 13-41.

[39] A. Schuster (1900) The periodogram of magnetic decliniation as obtained from the records of the Greenwich Observatory during the years 1871-1895. *Cambridge Phil. Trans.* **18**, 107-135.

[40] A. Schuster (1906) On the periodocities of sunspots. *Phil. Trans. R. Soc. A* **206**,69-100.

[41] U.J. Schwarz (1978) Mathematical-statistical description of the iterative beam removing technique (Method CLEAN). *Astronomy and Astrophysics* **65**, 345-356.

[42] M. Shimshoni (1971) On Fisher's test of significance in harmonic analysis. *Geophys. J. R. Astron. Soc.* **23**, 373-377.

[43] R.F. Stellingwerf (1978) Period determination using phase dispersion minimization. *Astrophysical Journal* **224**, 953-960.

[44] A.D. Thrall (1979) *Spectral Estimation for a Randomly Sampled Time Series.* Ph.D. dissertation, Dept. of Statistics, U.C. Berkeley.

[45] J.W. Tukey (1961) Discussion, emphasizing the connection between analysis of variance and spectrum analysis. *Technometrics* **3**, 1-29.

[46] G. Wahba (1980) Automatic smoothing of the log periodogram. *Journal of the American Statistical Association* **75**, 122-132.

[47] X. Wang (1991) *On the Estimation of Trigonometric and Related Signals.* Ph.D. dissertation, Dept. of Statistics, U.C. Berkeley.

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

[49] E. Whittaker and G. Robinson (1944) *The Calculus of Observations.* Blackie, London.

[50] P. Whittle (1952) The simultaneous estimation of a time series' harmonic components and covariance structure. *Trabajos de Estadistica* **3**, 43-57.

[51] P. Whittle (1959) Sur la distribution du maximum d'un polynome trigonométrique à coefficients aléatoires. *Colloques Internationaux du Centre National de la Recherche Scientifique* **87**, 173-184.