Mon. Not. R. Astron. Soc. **419,** 1197–1218 (2012) doi:10.1111/j.1365-2966.2011.19778.x

Estimation of the coherence time of stochastic oscillations from modest samples

Chris Koen★

Department of Statistics, University of the Western Cape, Private Bag X17, Bellville 7535 Cape, South Africa

Accepted 2011 September 5. Received 2011 September 4; in original form 2011 May 4

ABSTRACT

'Quasi-periodic' or 'solar-like' oscillations can be described by three parameters – a characteristic frequency, a coherence time (or 'quality factor') and the variance of the random driving process. This paper is concerned with the estimation of these quantities, particularly the coherence time, from modest sample sizes (observations covering of the order of a hundred or fewer oscillation periods). Under these circumstances, finite sample properties of the periodogram (bias and covariance) formally invalidate the commonly used maximum-likelihood procedure. It is shown that it none the less gives reasonable results, although an appropriate covariance matrix should be used for the standard errors of the estimates. Tailoring the frequency interval used, and oversampling the periodogram, can substantially improve parameter estimation. Maximum-likelihood estimation in the time-domain has simpler statistical properties, and generally performs better for the parameter values considered in this paper. The effects of added measurement errors are also studied. An example analysis of pulsating star data is given.

Key words: methods: statistical – stars: oscillations.

1 INTRODUCTION

Periodicities in the radiation from astronomical objects are common. In many cases, the cyclical variations are not deterministic – there are unpredictable systematic changes in amplitudes and frequencies/phases of the periodicities. (Mathematically, frequency and phase variations cannot be distinguished from one another, although there may be physical reasons for preferring one description to the other.) Such variability is often termed 'quasi-periodic': short-time-scale quasi-periodic oscillations (QPOs) have been observed in, for example, binary stars with compact components (white dwarfs, neutron stars or black holes – e.g. Finger 1998; Mauche 2002), and in flares from magnetars (e.g. Watts & Strohmayer 2007) and the Sun (e.g. Jakimiec & Tomczak 2010). The term 'solar-like oscillations' (SLOs) is preferred for the stochastically excited pulsations seen in some late-type stars, notably the Sun (see Bedding & Kjeldsen 2003, for a review).

A fundamental property of QPOs and SLOs is their coherence times (referred to as 'mode lifetimes' in the SLO literature). In the astronomy literature, this and other properties of QPOs and SLOs are almost invariably studied in the frequency-domain, a standard reference being Anderson, Duvall & Jefferies (1990). [See Fukumura et al. (2010) for a rare time-domain study of QPOs.] The frequency content of a set $\{y_1, y_2, \ldots, y_N\}$ of N observations, assumed here to be at regularly spaced time-points, can be investigated by means of the periodogram

$$I(\nu) \equiv \frac{1}{\pi N}\left|\sum_{j=1}^{N}(y_j - \overline{y})\mathrm{e}^{-\mathrm{i}2\pi\nu j}\right|^2, \qquad 0 < \nu < 0.5. \tag{1}$$

The presence of a periodicity in the data gives rise to a periodogram peak at the corresponding frequency. If the cyclical variation has a random element, due to variable amplitude, frequency or phase, then the periodogram peak will be broader than if the periodicity were stable: the peak width provides a measure of the coherence of the periodicity. The most common approach to the estimation of coherence times is based on fitting a Lorenzian form

$$S(\nu) = \frac{A}{(\nu - \nu_0)^2 + \gamma^2} \tag{2}$$

to the periodogram peak.

★E-mail: ckoen@uwc.ac.za

Anderson et al. (1990) describe a maximum-likelihood (ML) method for fitting the Lorenzian shape, from which estimates of the mode frequency and lifetime, as well as background noise level, can be made. In order to apply the method, the periodogram is calculated in frequencies $\nu_1, \nu_2, \ldots, \nu_M$ chosen so that $I(\nu_1), I(\nu_2), \ldots, I(\nu_M)$ are statistically independent. The advantage of such a choice of frequencies is that the likelihood function, or joint probability density of the data, is a simple product of the probability terms for each periodogram ordinate. Using the fact that, in general, for large samples, $I(\nu_j)$ is exponentially distributed with the mean equal to $S(\nu_j)$, the likelihood function is

$$L = \prod_{j=1}^{M} [S(\nu_j)]^{-1} \exp\{-[I(\nu_j)/S(\nu_j)]\}. \tag{3}$$

Equation (3) is no longer valid if covariance between the different $I(\nu_j)$ is non-zero. Quantitatively, the values of $I(\nu_j)$ are uncorrelated in the Fourier frequencies

$$\nu_j = j/N, \quad j = 1, 2, \ldots, [(N-1)/2], \tag{4}$$

where the notation indicates that $(N-1)/2$ should be rounded down to the nearest integer. For very long data sets, the frequency resolution is very fine, and the Fourier frequencies are quite closely spaced. For short data sets, it may be advantageous to use a denser set of frequencies, in order to sample the shape of the spectral peak. Since the ML method is then no longer formally correct in such situations, it is worth investigating potential competitors. This paper compares the performance of the ML frequency-domain method with two alternatives, namely least squares (LS) in the frequency-domain and ML applied in the time-domain.

Throughout, it is assumed that observations are regularly spaced in time, with no gaps.

The referee of this paper has made the important point that there is another class of processes which are superficially similar to those dealt with in this paper, but which are none the less fundamentally different. This is the case in which periodicities are intrinsically deterministic, but amplitudes or phases may be variable. An example is the unpredictable amplitude variability, at fixed frequencies, of the pulsating hot subdwarf star Feige 48 (Reed et al. 2004). Statistical tests based on this model were discussed by Koen (2009). This paper is concerned only with models in which periodicities are of a purely stochastic nature, describable by a second-order differential equation with a white-noise-driving term.

2 THE SPECTRUM AND THE AUTOCOVARIANCE FUNCTION

The equation describing a damped harmonic oscillator driven by noise $\epsilon(t)$ is

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + \alpha_1 \frac{\mathrm{d}y}{\mathrm{d}t} + \alpha_0 y(t) = \frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + \frac{\omega_0}{Q}\frac{\mathrm{d}y}{\mathrm{d}t} + \omega_0^2 y(t) = \epsilon(t)\ . \tag{5}$$

In equation (5), Q is the damping constant and $\omega_0 = 2\pi\nu_0$ is the angular frequency of the system under zero damping ($Q \to \infty$). The variance of the driving noise, assumed to be Gaussian and white, will be denoted by σ_ϵ^2. The damping constant, or 'quality', is related to the mode lifetime, as will be made clear below. In mathematical terms, the stochastic differential equation (SDE) (5) describes a second-order continuous time autoregressive [CAR (2)] process. Equation (5) implies a particular form of the spectral density (in the frequency-domain) and the autocovariance function (in the time-domain).

The spectrum associated with equation (5) is

$$S(\omega) = \frac{1}{\pi} \frac{\sigma_\epsilon^2}{\left|\sum_{k=0}^{2} \alpha_k (\mathrm{i}\omega)^k\right|^2}, \tag{6}$$

where $\alpha_2 = 1$ (e.g. Brockwell 2001). This is equivalent to

$$S(\omega) = \frac{1}{\pi} \frac{\sigma_\epsilon^2}{(\alpha_0 - \omega^2)^2 + (\alpha_1\omega)^2} = \frac{1}{\pi} \frac{\sigma_\epsilon^2}{\left(\omega_0^2 - \omega^2\right)^2 + (\omega_0\omega/Q)^2}. \tag{7}$$

It is interesting that the frequency of maximum power is not at ω_0, but at

$$\omega_* = \omega_0 \sqrt{1 - 0.25/Q^2}\ . \tag{8}$$

The approximation (2) is derived from equation (7) under the assumption that ω is not very different from ω_0:

$$\begin{aligned} S(\omega) &\approx \frac{1}{\pi} \frac{\sigma_\epsilon^2}{4\omega_0^2(\omega_0 - \omega)^2 + \omega_0^4/Q^2} \\ &= \frac{1}{\pi} \frac{\sigma_\epsilon^2/\left(16\pi^2\omega_0^2\right)}{(\nu_0 - \nu)^2 + (\nu_0/2Q)^2}. \end{aligned} \tag{9}$$

Comparing equation (2) with equation (9) shows that

$$\begin{aligned} A &= \sigma_\epsilon^2/\left(16\pi^2\omega_0^2\right), \\ \gamma &= \nu_0/(2Q). \end{aligned} \tag{10}$$

It is well known (and easy to show) that the full width at half-maximum (FWHM) of the spectral peak described by equation (2) is 2γ, that is, from the last equation above, ν_0/Q. The spacing of the Fourier frequencies is $\Delta\nu_\mathrm{F} = 1/N$ (see equation 4); hence, there are

$$M = \mathrm{FWHM}/\Delta\nu_\mathrm{F} = N\nu_0/Q = n_\mathrm{cycles}/Q$$

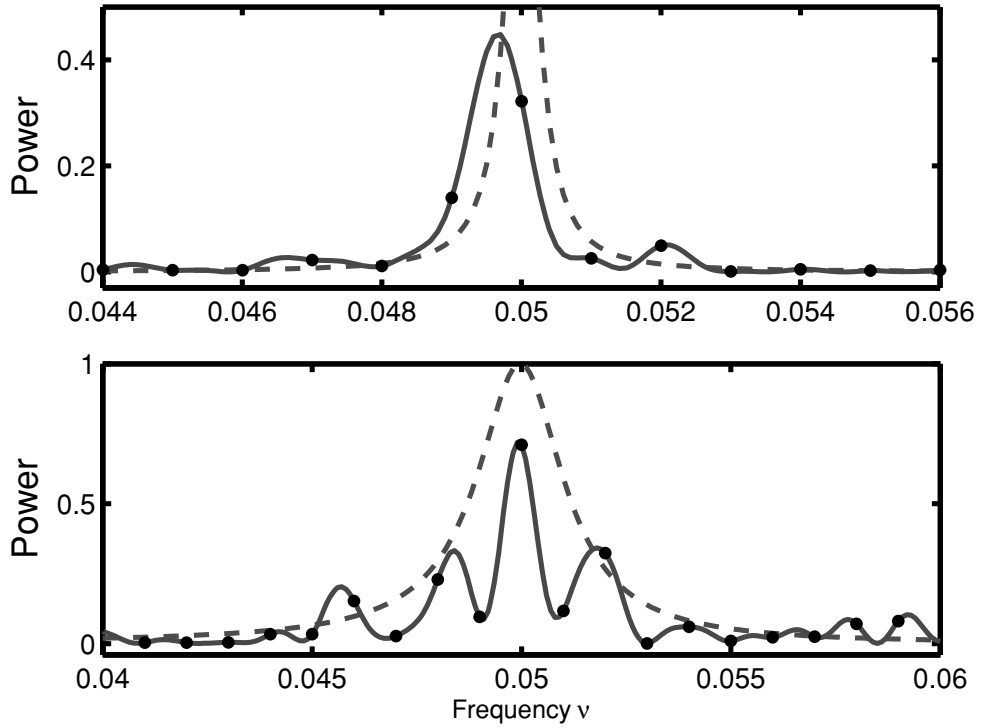

Figure 1. Theoretical (broken lines) and estimated (solid lines) spectra for $\nu_0 = 0.05$ and $Q = 100$ (top panel) and $Q = 20$ (bottom panel). The spectra have been normalized such that the maximum power of the theoretical spectra is unity. The dots mark the Fourier frequencies implied by the length $N = 1000$ of the time-series. Note that the scales on the two panels are different.

'independent frequencies' in the core of the spectral peak, where $n_{\rm cycles}$ is the number of oscillation cycles (as defined by ν_0) in the time-interval covered by the N observations. Clearly, small $n_{\rm cycles}$ will give a very small number M of terms in equation (3). This implies that it may be useful to oversample, that is, $\Delta\nu \ll 1/N$, particularly for large Q.

The last point is illustrated in Fig. 1, for a series length $N = 1000$. Two realizations with $\nu_0 = 0.05$ are shown, one with $Q = 100$ (top panel) and the other with $Q = 20$ (bottom panel). Both the theoretical spectra (7) and the periodograms (1) are plotted. The dots show the values of $I(\nu)$ in the Fourier frequencies (4).

Two other facets of Fig. 1 are worth remarking on. The first is that the periodogram peaks are smaller than the maxima of $S(\nu)$, particularly for $Q = 100$. This is no accident – the periodogram is, in fact, a biased estimator of S, a point looked at in more detail in Section 4. The second aspect, which is of practical importance, is that it is not clear what frequency range should be included in estimation functions such as the likelihood in equation (3). It seems likely that it will be counterproductive to use the entire frequency interval (0, 0.5), since frequencies far from the spectral peak do not carry much useful information, and may only add to the uncertainty of the estimation. In the present context, this point is particularly pertinent: the periodogram values are exponentially distributed; hence, small values of the spectrum are weighted more heavily than larger values – a circumstance which seems perverse, given that the spectral peak is of prime importance. For remarks on this problem, encountered in a different context, see Koen & Lombard (2004).

The autocovariance function of $y(t)$ follows from equations (20) in Koen (2005) as

$$\begin{aligned} C(\ell) &= \frac{\omega_0 C(0)}{\omega_*} \mathrm{e}^{-\ell/\tau_{\rm L}} \cos(\omega_* \ell + \phi), \\ \phi &= \tan^{-1}(-\omega_0/2Q\omega_*), \\ C(0) &= \frac{Q\sigma_\epsilon^2}{2\omega_0^3}, \\ \tau_{\rm L} &= 2Q/\omega_0. \end{aligned} \tag{11}$$

Note that $\tau_{\rm L}$ is the approximate e-folding time of the autocovariance, that is, it is a measure of the mode lifetime. The last equation of equations (11) can also be written as $\tau_{\rm L} = (Q/\pi)P_0$, where P_0 is the period corresponding to ω_0: this shows that the mode lifetime is roughly $Q/3$ periods.

The autocorrelation function is

$$\rho(\ell) = C(\ell)/C(0) = \frac{\omega_0}{\omega_*} \mathrm{e}^{-\ell/\tau_{\rm L}} \cos(\omega_* \ell + \phi). \tag{12}$$

For $Q > 10$ or so,

$$C(\ell) \approx C(0)\mathrm{e}^{-\ell/\tau_{\mathrm{L}}}\cos(\omega_*\ell),$$
$$\rho(\ell) \approx \mathrm{e}^{-\ell/\tau_{\mathrm{L}}}\cos(\omega_*\ell). \tag{13}$$

3 DATA SIMULATION

Testing estimation methods, or even just visualizing data, requires the facility to generate artificial data with known characteristics. One way of proceeding is to base simulations on discretization of the SDE (5) (for the relevant theory, see e.g. Török 1994; Gilsing & Shardlow 2007). The simple scheme

$$Y_{n+1} = Y_n + hY'_n,$$
$$Y'_{n+1} = -hY_n + (1 - h/Q)Y'_n + \eta_{n+1}\ , \tag{14}$$

will produce simulated $Y(t)$ spaced at (small) time-intervals h. In equations (14), the time is measured in units of $P_0/2\pi$. The random variable η has mean zero, and variance $h\omega_0^2\sigma_\epsilon^2$.

In order for equations (14) to be accurate, the time-step h has to be very small. This will make the method impractical if data are to be simulated over a long time-base. An alternative which could be used to produce a set of regularly spaced data satisfying equation (5) relies on the correspondence between CAR series and discrete time autoregressive moving average (ARMA) time-series. In particular, ARMA(2,1) series, that is, series of the form

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} - \theta_1\xi_{t-1} + \xi_t \tag{15}$$

(where ϕ_1, ϕ_2 and θ_1 are constants, and ξ_t is white noise), can be tailored to have the same small-lag autocovariance structure as Y_t (e.g. Phadke & Wu 1974). The required coefficients in equation (15) are given by

$$\phi_1 = 2\mathrm{e}^{-1/\tau_{\mathrm{L}}}\cos\omega_*\,,$$
$$\phi_2 = -\mathrm{e}^{-2/\tau_{\mathrm{L}}}, \tag{16}$$

and the solutions θ_1 and σ_ξ^2 of

$$A = \phi_1 + \frac{1 - \phi_1\rho(1) - \phi_2\rho(2)}{\phi_1 + (\phi_2 - 1)\rho(1)},$$
$$\theta_1^2 - A\theta_1 + 1 = 0, \quad |\theta_1| < 1,$$
$$\sigma_\xi^2 = \begin{cases} C(0)[\phi_1 + (\phi_2 - 1)\rho(1)]/\theta_1 & \theta_1 \neq 0 \\ C(0)[1 - \phi_1\rho(1) - \phi_2\rho(2)] & \theta_1 = 0\,. \end{cases} \tag{17}$$

[Note that there is an error in equation (2.8) of Phadke & Wu (1974).]

The autocorrelation function values $\rho(1)$ and $\rho(2)$ can be calculated from equation (12), and $C(0)$ is given by equations (11). Note though that Z_t in equation (15) is only an approximation of the continuous time-process y in the time-points t: the spectrum of Z_t is given by

$$S(\omega) = \frac{\sigma_\eta^2}{\pi}\frac{|1 - \theta_1\mathrm{e}^{-\mathrm{i}\omega}|^2}{|1 - \phi_1\mathrm{e}^{-\mathrm{i}\omega} - \phi_2\mathrm{e}^{-2\mathrm{i}\omega}|^2}$$

(e.g. Wei 1990) which is generally different from equation (6).

If the length N of the data set to be simulated is modest (at most some hundreds), then the following general strategy can be used to generate y_t with white Gaussian driving noise:

(i) Equations (11) can be used to calculate the covariances of all the values of a time-series with specified parameter values and length N: the results form the entries in the covariance matrix $\mathbf{\Sigma}$ of y_t.

(ii) Since $\mathbf{\Sigma}$ is a covariance matrix, it is positive definite and can be decomposed as

$$\mathbf{\Sigma} = \mathbf{R}^t\mathbf{R}, \tag{18}$$

where $\mathbf{R}$ is an upper triangular matrix (e.g. Healy 1986).

(iii) Generate a vector of N independent standard Gaussians $\boldsymbol{z}$; then,

$$\boldsymbol{y} = \mathbf{R}^t\boldsymbol{z}$$

conforms to the SDE (5).

It is noteworthy that the scheme outlined depends on the time-points of observation only through the covariances (11): this implies that it can be used regardless of the time-spacing, that is, it applies also to irregularly observed time-series.

4 FINITE SAMPLE PROPERTIES OF THE PERIODOGRAM

A point which is sometimes overlooked is that the periodogram is only an *asymptotically* unbiased estimator of the spectrum: for finite samples, it is biased. The small sample bias is more pronounced for large values of Q, as shown in Fig. 1. Bias also increases with decreasing sample size – see Fig. 2.

The bias is the difference between the spectrum and the expected value (ensemble average) of the periodogram. In general,

$$\mathrm{E}I(\omega) = \frac{1}{2\pi N}\int_0^{\pi}\left[\frac{\sin N(\omega - x)/2}{\sin(\omega - x)/2}\right]^2 S(x)\,\mathrm{d}x \tag{19}$$

(e.g. Brillinger 1981). The function multiplying $S(x)$ in the integrand is known as Fejér's kernel. Alternatively,

$$\begin{aligned}\mathrm{E}I(\omega) &= \frac{1}{N}\mathrm{E}\left|\sum_{j=1}^{N} y_j \mathrm{e}^{-\mathrm{i}2\pi\nu j}\right|^2,\\ &= \frac{1}{N}\mathrm{E}\left(\sum_{j=1}^{N} y_j \mathrm{e}^{-\mathrm{i}2\pi\nu j}\right)\left(\sum_{k=1}^{N} y_k \mathrm{e}^{\mathrm{i}2\pi\nu k}\right),\\ &= \frac{1}{N}\sum_{j,k=1}^{N} C(|j-k|)\mathrm{e}^{-\mathrm{i}\omega(j-k)},\\ &= C(0) + \frac{2}{N}\sum_{\ell=1}^{N}(N-\ell)C(\ell)\cos(\ell\omega),\end{aligned} \tag{20}$$

with $C(\ell)$ given by equations (11).

The implication is that, in practice, $S(\nu)$ should be replaced by $\mathrm{E}I(\nu)$ in estimating equations. Anderson et al. (1990) wrote 'The final model used for comparison with the data is a convolution of the model ... with the power spectrum of the observed window function of the data set. ... This takes into account the redistribution of power caused by gaps in the data'. Fig. 2 shows that, in small samples, this also applies in cases where there are no gaps in the data.

For Gaussian data, it is known (e.g. Anderson 1971; Krogstad 1982) that the interrelation between periodogram ordinates at arbitrary frequencies ν and λ is described by

$$\mathrm{cov}[I(\nu), I(\lambda)] = \frac{1}{N^2}\sum_{j,k,\ell,m=1}^{N}[C(j-\ell)C(k-m) + C(j-m)C(k-\ell)]\mathrm{e}^{-2\pi\mathrm{i}[(j-k)\lambda+(\ell-m)\nu]}, \tag{21}$$

$C(\ell)$ being the autocovariance function at lag ℓ of the time-series y_j. Summation over the four indices is computationally extremely expensive, even for moderately short time-series. Note though that

$$\begin{aligned}&\sum_{j,k,\ell,m} C(j-\ell)C(k-m)\mathrm{e}^{-2\pi\mathrm{i}[(j-k)\lambda+(\ell-m)\nu]}\\ &\quad= \left[\sum_{j,\ell} C(j-\ell)\mathrm{e}^{-2\pi\mathrm{i}(j\lambda+\ell\nu)}\right]\left[\sum_{k,m} C(k-m)\mathrm{e}^{2\pi\mathrm{i}(k\lambda+m\nu)}\right]\\ &\quad= \left|\left[\sum_{j,\ell} C(j-\ell)\mathrm{e}^{-2\pi\mathrm{i}(j\lambda+\ell\nu)}\right]\right|^2.\end{aligned}$$

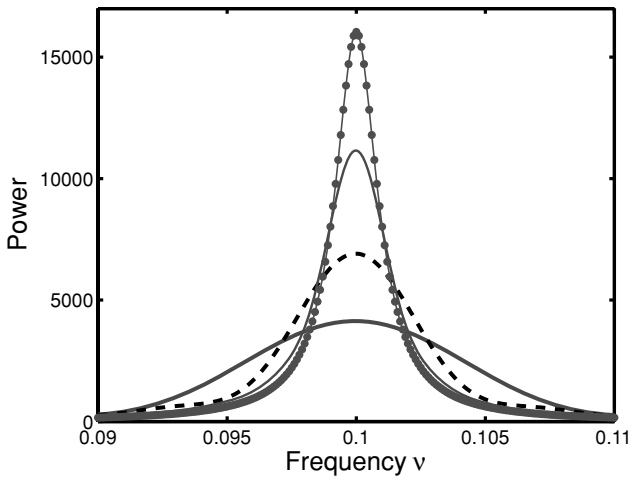

Figure 2. The spectral density (connected dots) for $\nu_0 = 0.1$ and $Q = 50$. Expected values of the periodogram are shown by the thick solid line ($N = 100$), broken line ($N = 200$) and thin solid line ($N = 500$). The bias is obvious.

The terms in equation (21) containing $C(j-m)C(k-\ell)$ can be treated similarly, with the result that

$$\mathrm{cov}[I(\nu), I(\lambda)] = \frac{1}{N^2}\left[\left|\sum_{j,\ell} C(j-\ell)\mathrm{e}^{-2\pi\mathrm{i}(j\lambda+\ell\nu)}\right|^2 + \left|\sum_{j,\ell} C(j-\ell)\mathrm{e}^{-2\pi\mathrm{i}(j\lambda-\ell\nu)}\right|^2\right] \tag{22}$$

which can calculated much more quickly.

For data sets with lengths of a few thousands, even equation (22) is very computer-intensive. It is therefore worthwhile to work through some lengthy but easy algebra in order to derive formulae for faster computation. Consider

$$S_1 = \sum_{j,\ell=1}^{N} C(j-\ell)\mathrm{e}^{-2\pi\mathrm{i}(j\lambda+\ell\nu)}, \qquad S_2 = \sum_{j,\ell=1}^{N} C(j-\ell)\mathrm{e}^{-2\pi\mathrm{i}(j\lambda-\ell\nu)} \tag{23}$$

so that

$$\mathrm{cov}[I(\nu), I(\lambda)] = \frac{1}{N^2}\left(S_1 S_1^* + S_2 S_2^*\right), \tag{24}$$

where the superscript $*$ indicates complex conjugation. By summing over terms with the same values of $C(j-\ell)$, it can be shown that, provided $\nu \neq \lambda$,

$$S_1 = \left\{C(0)\left[1-\mathrm{e}^{-2\pi\mathrm{i}N(\lambda+\nu)}\right] + \sum_{\ell=1}^{N-1} C(\ell)\left(\mathrm{e}^{-2\pi\mathrm{i}\ell\nu} + \mathrm{e}^{-2\pi\mathrm{i}\ell\lambda}\right)\left[1-\mathrm{e}^{-2\pi\mathrm{i}(N-\ell)(\lambda+\nu)}\right]\right\} \Big/ \left[1-\mathrm{e}^{-2\pi\mathrm{i}(\lambda+\nu)}\right] \tag{25}$$

and similarly for S_2, with $-\nu$ replacing ν. It is noteworthy that these formulae are valid for any stationary time-series.

A little further rearrangement gives

$$\begin{aligned} S_1 &= \left[1-\mathrm{e}^{2\pi\mathrm{i}(\lambda+\nu)}\right]\left\{[C(0)+h(\nu)+h(\lambda)] - \mathrm{e}^{-2\pi\mathrm{i}N(\lambda+\nu)}[C(0)+h^*(\nu)+h^*(\lambda)]\right\}/2\{1-\cos[2\pi(\lambda+\nu)]\}, \\ S_2 &= \left[1-\mathrm{e}^{2\pi\mathrm{i}(\lambda-\nu)}\right]\left\{[C(0)+h^*(\nu)+h(\lambda)] - \mathrm{e}^{-2\pi\mathrm{i}N(\lambda-\nu)}[C(0)+h(\nu)+h^*(\lambda)]\right\}/2\{1-\cos[2\pi(\lambda-\nu)]\}, \end{aligned} \tag{26}$$

where

$$h(\nu) = \sum_{\ell=1}^{N-1} C(\ell)\mathrm{e}^{-2\pi\mathrm{i}\ell\nu} \tag{27}$$

and similarly for $h(\lambda)$. For the specific covariance function in equations (11), it can be shown that

$$\begin{aligned} h(\nu) &= \frac{Q\sigma_\epsilon^2}{4\omega_* \omega_0^2}\left(\mathrm{e}^{\mathrm{i}\phi}\frac{\mathrm{e}^{a_1}-\mathrm{e}^{Na_1}}{1-\mathrm{e}^{a_1}} + \mathrm{e}^{-\mathrm{i}\phi}\frac{\mathrm{e}^{a_2}-\mathrm{e}^{Na_2}}{1-\mathrm{e}^{a_2}}\right), \\ a_1 &= \mathrm{i}(\omega_* - 2\pi\nu) - 1/\tau, \quad a_2 = -\mathrm{i}(\omega_* + 2\pi\nu) - 1/\tau\,. \end{aligned}$$

For the special case $\lambda = \nu$,

$$\begin{aligned} S_1 &= (1-\mathrm{e}^{4\pi\mathrm{i}\nu})\left\{[C(0)+2h(\nu)] - \mathrm{e}^{-4\pi\mathrm{i}N\nu}[C(0)+2h^*(\nu)]\right\}/2[1-\cos(4\pi\nu)], \\ S_2 &= NC(0) + 2\sum_{\ell=1}^{N-1}(N-\ell)C(\ell)\cos(2\pi\ell\nu)\,. \end{aligned} \tag{28}$$

From equations (28) and (20), it follows that $S_2 = N\mathrm{E}[I(\nu)]$, and hence

$$\mathrm{var}[I(\nu)] = \mathrm{cov}[I(\nu), I(\nu)] = \frac{1}{N^2}S_1(\nu)S_1^*(\nu) + [\mathrm{E}I(\nu)]^2\,. \tag{29}$$

If $\nu = j/N$ and $\lambda = k/N$ are distinct Fourier frequencies (see equation 4),

$$\begin{aligned} S_1 &= \frac{-2\mathrm{i}}{1-\mathrm{e}^{-2\pi\mathrm{i}(k+j)/N}}\sum_{\ell=1}^{N-1} C(\ell)\left(\sin\frac{2\pi\ell k}{N} + \sin\frac{2\pi\ell j}{N}\right), \\ S_2 &= \frac{-2\mathrm{i}}{1-\mathrm{e}^{-2\pi\mathrm{i}(k-j)/N}}\sum_{\ell=1}^{N-1} C(\ell)\left(\sin\frac{2\pi\ell k}{N} - \sin\frac{2\pi\ell j}{N}\right). \end{aligned} \tag{30}$$

If $k = j$,

$$\begin{aligned} S_1 &= \frac{-4\mathrm{i}}{1-\mathrm{e}^{-4\pi\mathrm{i}j/N}}\sum_{\ell=1}^{N-1} C(\ell)\sin\frac{2\pi\ell j}{N}, \\ S_2 &= NC(0) + 2\sum_{\ell=1}^{N-1}(N-\ell)C(\ell)\cos\frac{2\pi\ell j}{N} = N\mathrm{E}[I(\nu)]. \end{aligned} \tag{31}$$

An important point following from equations (29)–(31) is that the periodogram ordinates are neither independent nor exponentially distributed: these attributes only apply asymptotically, as $N \to \infty$. Figs 3 and 4 show colour-coded correlation matrices

$$R(\nu_j, \nu_k) = \mathrm{cov}[I(\nu_j), I(\nu_k)] / \left\{\mathrm{cov}[I(\nu_j), I(\nu_j)]\mathrm{cov}[I(\nu_k), I(\nu_k)]\right\}^{1/2} \tag{32}$$

for $N = 500$ and 1000, respectively, for the parameters $\nu_0 = 0.05$ and $Q = 50$. Clearly, periodogram values at frequencies larger than ν_0 can be very strongly correlated, particularly for smaller N. The correlation increases with increasing Q.

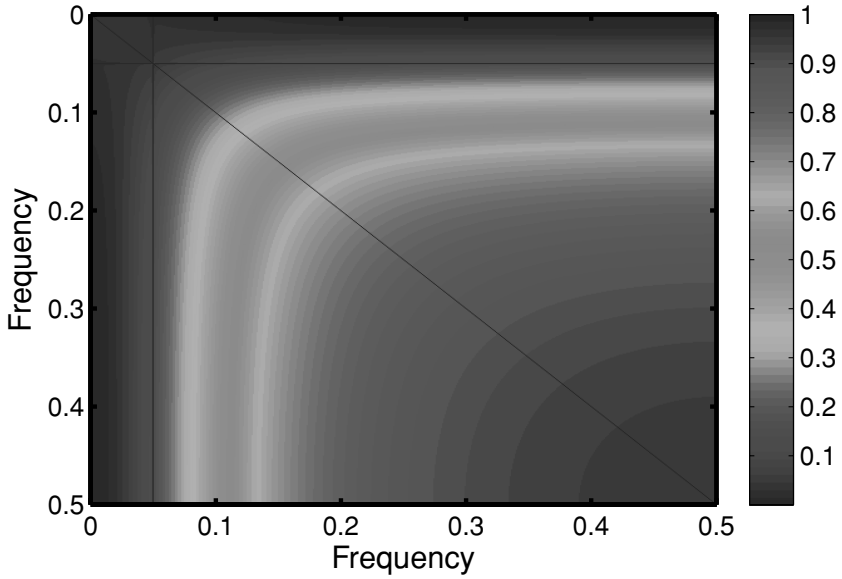

Figure 3. The correlation matrix of the periodiogram ordinates, calculated in the Fourier frequencies (4), for parameter values $\nu_0 = 0.05$, $Q = 50$ and $N = 1000$. (Colour in the online version)

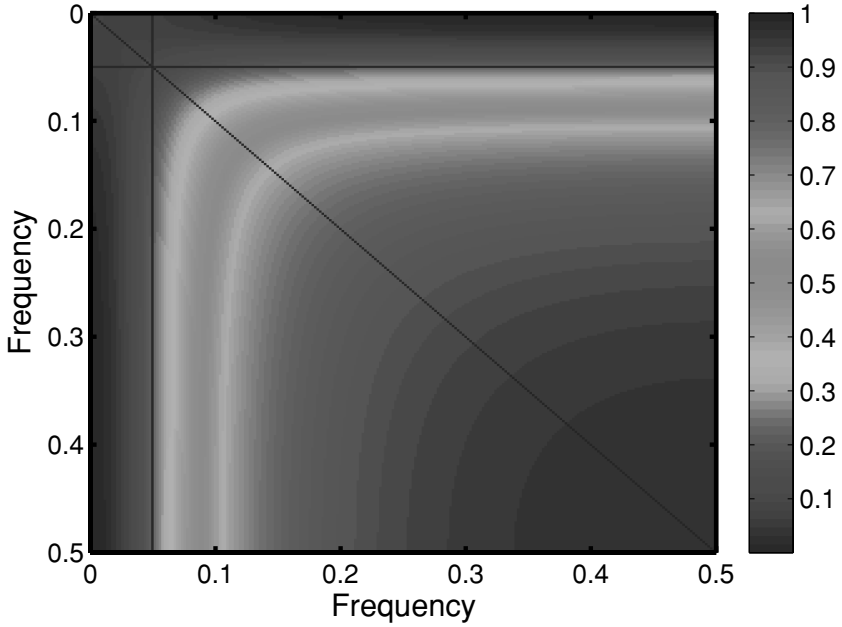

Figure 4. As for Fig. 3, but for a sample size $N = 500$.

The exponential distribution is characterized by the fact that the standard deviation and expected value are the same. Fig. 5 shows the ratio of these two statistics for the periodogram, for a few parameter combinations. The ratio has a minimum value of unity in the frequency ν_0; otherwise, the standard deviation exceeds the mean. The implication is that $I(\nu)$ does not in general have an exponential distribution: the spread of values is larger than for an exponential distribution with the same mean.

5 THE FREQUENCY-DOMAIN ESTIMATORS: STANDARD ERRORS

The upshot of the results covered in Section 4 is that for 'small' samples equation (3) is not, strictly speaking, a likelihood function. Use of it for estimation is more accurately referred to as 'quasi-maximum-likelihood estimation' (QMLE). Furthermore, since the optimal properties of MLE will no longer necessarily hold, it is worthwhile considering also the alternative of LS estimation (LSE).

It is convenient to define the vector

$$\boldsymbol{\theta} = \begin{bmatrix} \omega_0 \\ Q \\ \sigma_\epsilon^2 \end{bmatrix}$$

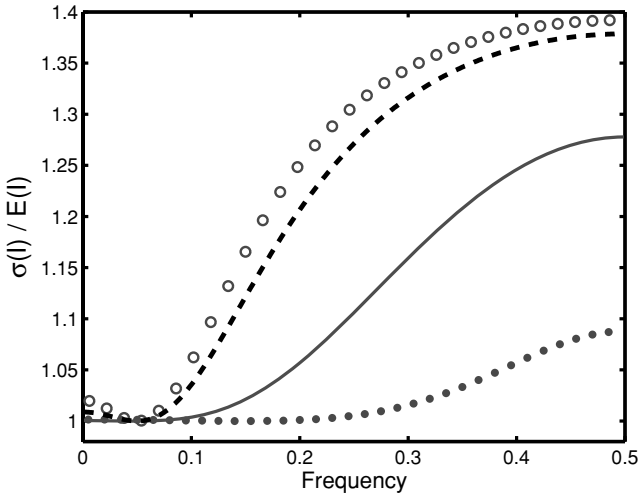

Figure 5. The ratio of the periodogram standard deviation to the expected value given by equations (20). Open circles: ($\nu_0 = 0.05$, $Q = 50$, $N = 500$); broken line: ($\nu_0 = 0.05$, $Q = 50$, $N = 1000$); solid line: ($\nu_0 = 0.05$, $Q = 10$, $N = 1000$); dots: ($\nu_0 = 0.15$, $Q = 50$, $N = 1000$).

of parameters to be estimated. QMLE and LSE of the components of $\boldsymbol{\theta}$ can be cast in *estimating equation* form by introducing the objective function $\Psi(\boldsymbol{y}, \boldsymbol{\theta})$, given by either

$$\Psi(\boldsymbol{I}, \boldsymbol{\theta}) = -\sum_{j=1}^{M} \left[\frac{I(\nu_j)}{S_*(\nu_j, \boldsymbol{\theta})} + \log S_*(\nu_j, \boldsymbol{\theta}) \right] \tag{33}$$

(QMLE) or

$$\Psi(\boldsymbol{I}, \boldsymbol{\theta}) = \sum_{j=1}^{M} [I(\nu_j) - S_*(\nu_j, \boldsymbol{\theta})]^2 \tag{34}$$

(LSE). In accordance with the results of the previous section, the spectrum S has been replaced by the expected values of the periodogram: $S_*(\omega) \equiv \mathrm{E}I(\omega)$.

In what follows, the frequencies will not necessarily be given by equation (4), since oversampling may be profitable.

The estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained by maximizing Ψ with respect to the components θ_k of $\boldsymbol{\theta}$. This leads to a set of K equations of the form

$$\psi_k = \frac{\partial \Psi}{\partial \theta_k} = 0, \qquad k = 1, 2, \ldots, K. \tag{35}$$

A note: although $\boldsymbol{\theta}$ has three components, in practice, it is only necessary to optimize Ψ with respect to ω_0 and Q. This is because it is possible to deduce explicit estimators for σ_ϵ^2 in terms of the periodogram, $\widehat{\omega_0}$ and $\widehat{Q}$:

$$\begin{aligned} \widehat{\sigma_\epsilon^2} &= \sum_j \left[I(\omega_j) S_{**}(\omega_j) \Big/ \sum_j S_{**}^2(\omega_j) \right] \qquad \text{(LSE)}, \\ \widehat{\sigma_\epsilon^2} &= \frac{1}{N} \sum_j \left[I(\omega_j) / S_{**}(\omega_j) \right] \qquad \text{(QMLE)}, \end{aligned} \tag{36}$$

where

$$S_{**}(\omega_j) \equiv \mathrm{E}I / \sigma_\epsilon^2 \tag{37}$$

and the right-hand sides of the equations are calculated from equations (20) using the estimated values of Q and ω_0.

A benefit of the general form (35) is that there is then also a general covariance matrix for the components of $\widehat{\boldsymbol{\theta}}$, the so-called sandwich estimator,

$$\mathbf{C}(\widehat{\boldsymbol{\theta}}) = \mathbf{A}^{-1} \mathbf{B} (\mathbf{A}^{-1})' \tag{38}$$

(e.g. Cameron & Trivedi 2005). The components of the matrices $\mathbf{A}$ and $\mathbf{B}$ are given by

$$\begin{aligned} \mathbf{A}_{rs} &= -\mathrm{E} \left(\frac{\partial \psi_r(\boldsymbol{I}, \boldsymbol{t})}{\partial t_s} \right)_{\boldsymbol{t}=\boldsymbol{\theta}} \\ \mathbf{B}_{rs} &= \mathrm{cov}(\psi_r, \psi_s) \, . \end{aligned} \tag{39}$$

In the case of the LS estimators derived from equation (34),

$$\psi_k = -2 \sum_{j=1}^{N} [I(\nu_j) - S_*(\nu_j, \boldsymbol{t})] \frac{\partial S_*}{\partial t_k} \tag{40}$$

and therefore

$$\mathbf{A}_{rs} = -2\sum_{j=1}^{M}\left[\frac{\partial S_*(\nu_j,\boldsymbol{t})}{\partial t_r}\frac{\partial S_*(\nu_j,\boldsymbol{t})}{\partial t_s}\right]_{\boldsymbol{t}=\boldsymbol{\theta}}, \tag{41}$$

$$\mathbf{B}_{rs} = 4\sum_{i,j=1}^{M}\left[\frac{\partial S_*(\nu_i,\boldsymbol{\theta})}{\partial \theta_r}\frac{\partial S_*(\nu_j,\boldsymbol{\theta})}{\partial \theta_s}\right]\mathrm{cov}(I_i,\,I_j). \tag{42}$$

In the case of QMLE,

$$\mathbf{A}_{rs} = \sum_{j=1}^{M}\frac{\partial \log S_*(\omega_j,\boldsymbol{\theta})}{\partial \theta_r}\frac{\partial \log S_*(\omega_j,\boldsymbol{\theta})}{\partial \theta_s}, \tag{43}$$

which is identical to the Fisher information matrix. The components of $\mathbf{B}$ are

$$\mathbf{B}_{rs} = \sum_{i,j=1}^{M}\left[\frac{\partial \log S_*(\nu_i,\boldsymbol{\theta})}{\partial \theta_r}\frac{\partial \log S_*(\nu_j,\boldsymbol{\theta})}{\partial \theta_s}\right]\frac{\mathrm{cov}(I_i,\,I_j)}{S_*(\nu_i,\boldsymbol{\theta})S_*(\nu_j,\boldsymbol{\theta})}. \tag{44}$$

Note that if, as is usually assumed,

$$\mathrm{cov}(I_i,\,I_j) = S_*(\nu_i)S_*(\nu_j)\delta_{ij}\ ,$$

where ν_i and ν_j are Fourier frequencies, and δ_{ij} is the Kronecker delta, then $\mathbf{B} = \mathbf{A}$ and equation (38) reduces to

$$\mathbf{C}(\widehat{\boldsymbol{\theta}}) = \mathbf{A}^{-1}\ , \tag{45}$$

that is, the covariance matrix of the estimates equals the inverse Fisher information matrix.

The derivatives in equation (42) are

$$\begin{aligned}
\frac{\partial S_*(\omega)}{\partial \omega_0} &= -3\frac{S_*}{\omega_0} - 2\frac{C(0)}{N}\sum_{\ell=1}^{N-1}\ell(N-\ell)\mathrm{e}^{-\ell/\tau_\mathrm{L}}\cos(\ell\omega)\left[\sin(\ell\omega_* + \phi) + \frac{\cos(\ell\omega_* + \phi)}{\sqrt{4Q^2-1}}\right],\\
\frac{\partial S_*(\omega)}{\partial Q} &= \frac{S_*(\omega)}{Q} - \frac{C(0)\omega^2}{4Q^3\omega_*^2}[S_* - C(0)]\\
&\quad + \frac{C(0))\omega_0^2}{NQ^2\omega_*^2}\sum_{\ell=1}^{N-1}(N-\ell)\mathrm{e}^{-\ell/\tau_\mathrm{L}}\cos(\ell\omega)\left[\ell\omega_*\cos(\ell\omega_* + \phi) + \left(\frac{\ell\omega_0^3}{2Q\omega_*^2} - 1\right)\sin(\ell\omega_* + \phi)\right],\\
\frac{\partial S_*(\omega)}{\partial \sigma_\epsilon^2} &= \frac{S_*}{\sigma_\epsilon^2}.
\end{aligned} \tag{46}$$

Standard errors (SEs) are given by the square roots of the diagonals of the sandwich matrices.

6 THE FREQUENCY-DOMAIN ESTIMATORS: RESULTS

The results of some simulation experiments are summarized in Table 1. For each of the four parameter combinations, 1000 sample series were generated, and QMLE and LSE performed. The means and SEs of the estimated parameters are given. For comparison, SEs calculated from the sandwich estimators, and, in the case of QMLE, the inverse Fisher matrices, are also tabled. Throughout, it is assumed that $\sigma_\epsilon = 1$. Typical forms of the distributions of the QML and LS estimates are illustrated in Figs 6 and 7, respectively.

The following remarks can be offered:

(i) When maximizing the quasi-likelihood, or minimizing the sum of squares, an upper limit on Q of 999 was imposed. The reason was that solutions sometimes 'diverged' to $Q \to \infty$. Inspection of the data showed that this happened in cases where the generated time-series strongly resembled deterministic cycles, with little sign of damping. As expected, the tendency to find $Q \to \infty$ increases with increasing Q, with decreasing N, and with decreasing ν_0 (i.e. fewer cycles covered). The effect is much worse for LSE than for QMLE.

(ii) Estimated frequencies are essentially unbiased, while estimates of Q and σ_ϵ^2 are biased (upwards and downwards, respectively).

(iii) Despite the misspecification of the likelihood function, QMLE is substantially superior to LSE, both in terms of smaller bias and smaller SEs of the estimates.

(iv) The LS SEs predicted by the sandwich matrix (38) are generally very far off the mark. SEs based on the Fisher information matrix (SE 1 in Table 1) are accurate predictors of the QMLE errors in $\widehat{\omega}_0$, but underestimate the true SEs in $\widehat{Q}$ and $\widehat{\sigma_\epsilon^2}$. The QMLE sandwich estimators are much more realistic. It is speculated that the poor performance of the LS sandwich estimators is due to very large sample sizes required for this mode of estimation to approach the asymptotic regime – see, for example, the distribution of $\widehat{\sigma_\epsilon^2}$. For small Q, the QMLE error estimates SE 1 and SE 2 converge.

(v) SEs decrease with increasing sample size, and errors in estimated Q and σ_ϵ^2 decrease with increasing frequency (more cycles covered by the same number of observations). Errors in $\widehat{\omega}_0$ decrease with increasing Q (periodicities are better defined), but errors in estimated Q and σ_ϵ^2 increase.

Table 1. A summary of the results of some simulation experiments to test the efficacy of frequency-domain estimation of model parameters. For each of the five models, the parameter values are given in the second column of the table. Mean LS parameter estimates are in the third column, with standard deviations in brackets. The fourth column contains LSE SEs, according to the sandwich estimator (38). The remaining three columns give the QMLE results: mean estimates (followed by standard deviations in brackets); SEs calculated from the inverse Fisher matrix (SE 1); and SEs according to the sandwich matrix (SE 2). The last line of information for each of the five models shows the sample size, and the fraction of finite-Q solutions for both LSE and QMLE. Estimates derived from divergent solutions have not been included in the reported simulation results.

Parameter	True value	LS estimate	SE	QML estimate	SE 1	SE 2
ω_0	0.628 32	0.6282 (0.0052)	0.0044	0.6282 (0.0029)	0.0027	0.0026
Q	50	146 (141)	55	67 (84)	9.7	69
σ_ϵ^2	1	0.65 (0.71)	1.28	0.90 (0.16)	0.09	0.15
$N = 1000$	$p(\mathrm{LS}) = 0.952$	$p(\mathrm{QMLE}) = 0.996$				
ω_0	0.628 32	0.6283 (0.0037)	0.0028	0.6282 (0.0022)	0.0020	0.0019
Q	100	213 (185)	117	116 (149)	21.1	184
σ_ϵ^2	1	0.81 (0.88)	1.20	0.90 (0.25)	0.11	0.30
$N = 1000$	$p(\mathrm{LS}) = 0.854$	$p(\mathrm{QMLE}) = 0.973$				
ω_0	0.628 32	0.6282 (0.0077)	0.0055	0.6282 (0.0043)	0.0041	0.0039
Q	50	114 (130)	59	63 (87)	14.6	85
σ_ϵ^2	1	0.84 (0.87)	1.20	0.87 (0.25)	0.15	0.28
$N = 500$	$p(\mathrm{LS}) = 0.873$	$p(\mathrm{QMLE}) = 0.976$				
ω_0	0.314 16	0.3143 (0.0036)	0.0028	0.3144 (0.0022)	0.0020	0.0019
Q	50	120 (127)	59	65 (108)	9.9	92
σ_ϵ^2	1	0.78 (0.78)	1.20	0.86 (0.27)	0.13	0.30
$N = 1000$	$p(\mathrm{LS}) = 0.886$	$p(\mathrm{QMLE}) = 0.975$				
ω_0	0.628 32	0.6278 (0.015)	0.015	0.6287 (0.0086)	0.0083	0.0088
Q	5	13.7 (42.1)	1.5	5.02 (0.75)	0.73	0.75
σ_ϵ^2	1	0.86 (0.39)	0.058	1.01 (0.14)	0.058	0.12
$N = 1000$	$p(\mathrm{LS}) = 1.0$	$p(\mathrm{QMLE}) = 1.0$				

(vi) For $Q = 50$ and 100, estimation of ω_0 is largely decoupled from estimation of the other two parameters (correlation coefficients typically less than 0.1 in absolute value), while correlations between $\widehat{Q}$ and $\widehat{\sigma_\epsilon^2}$ are in the range $(-0.52, -0.44)$, for both LSE and QMLE. [Even stronger (anti)correlations are predicted the sandwich matrices: $(-0.93, -0.89)$ for LSE and $(-0.71, -0.44)$ for QMLE.] For smaller Q, $\widehat{\omega}_0$ and $\widehat{\sigma_\epsilon^2}$ are more strongly correlated ($r > 0.3$ for both LSE and QMLE, when $Q = 5$).

It is important to note that the means and SEs in Table 1 are calculated from the finite-Q estimates only; hence, these simulation SEs are really lower limits. The same practice is followed in *all* results reported below.

7 THE FREQUENCY-DOMAIN ESTIMATOR: POSSIBLE VARIATIONS

Frequency-domain estimation of Q, and to a lesser extent σ_ϵ, has limited success for parameter values (in particular sample sizes) considered in this paper. Modifications of the basic methodologies were therefore also considered.

7.1 Variants of QMLE

Table 2 reports the results of a simulation study aimed at studying the effects of (1) oversampling the periodogram; and (2) restricting the frequency interval of the periodogram which is used for estimation (subsampling). The second column of the table lists the oversampling factor β, defined by

$$\nu_j = j/(\beta N), \quad j = 1, 2, \ldots, [\beta(N-1)/2] \tag{47}$$

which can be compared with equation (4).

Inspection of Table 2 shows the following:

(i) Oversampling of the periodogram alleviates the problem of divergent estimates of Q. On the other hand, it does not seem to affect the range of values of $\widehat{Q}$ when divergence is *not* a problem (see the results in Table 2 for $Q = 10$).

(ii) The mean estimated ω_0 is impervious to both oversampling and subsampling of the periodogram. Oversampling can apparently increase both bias and SEs of the estimates of Q and σ_ϵ^2 – but it should be borne in mind that bias and SEs are less understated if there are fewer divergent estimates.

(iii) Judicious choices of subsampling intervals can lead to considerable improvements in QML estimates, both in bias and SEs – compare the two blocks of results for the $Q = 10$ model in Table 2.

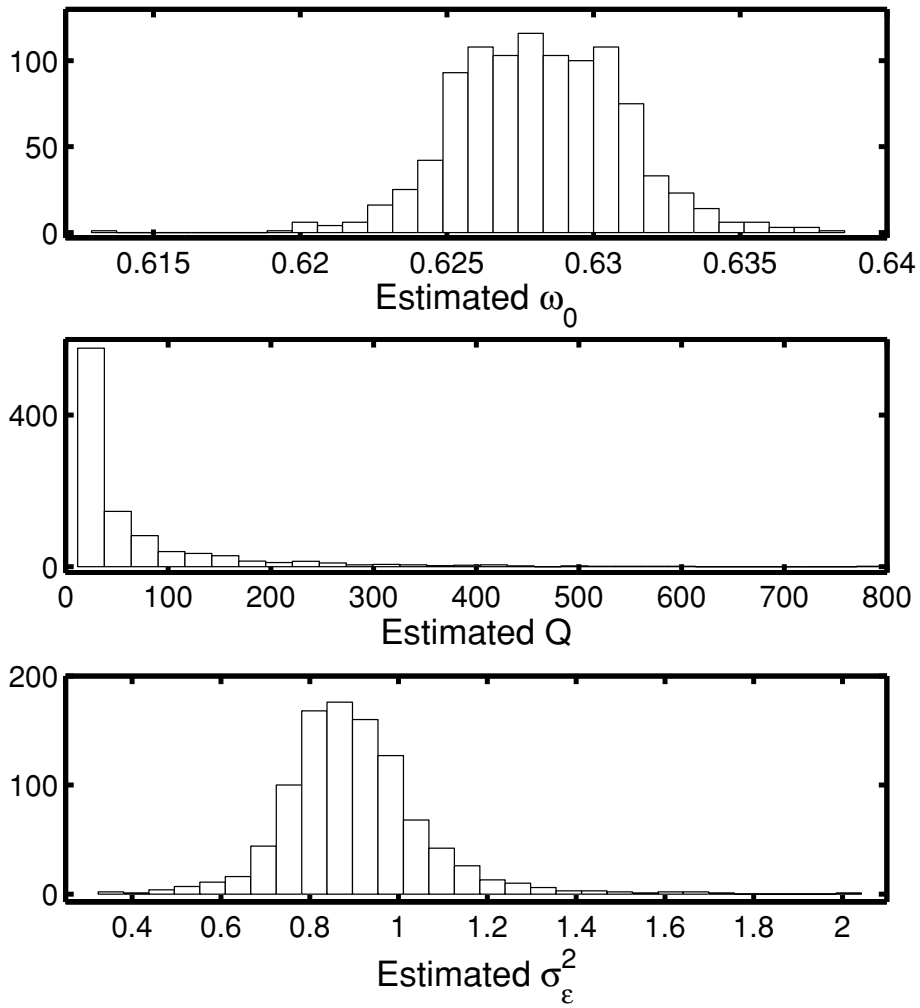

Figure 6. Histograms showing the relative frequencies of the estimates obtained by the frequency-domain QMLE method. These results were obtained from 1000 simulations, of which the estimated Q were finite for 996. The parameters used to generate the artificial data were $\omega_0 = 0.628\,32$, $Q = 50$, $\sigma_\epsilon^2 = 1$ and $N = 1000$.

(iv) There is very good agreement between the simulation results and the sandwich estimators in the last block of results in the table, but sandwich error estimates for $\widehat{Q}$ are usually substantially too small for model 1.

Fig. 8 illustrates the effects of oversampling and subsampling on the sandwich SEs of $\widehat{Q}$, for the two models in Table 2. The subsampling is done by selecting intervals of various widths $\Delta\nu$, symmetrically placed around the true frequency ν_0. For the second model ($Q = 10$, $N = 500$), results are insensitive to $\Delta\nu$, provided it is not too small, but for the model with $Q = 100$ and $N = 1000$, there are optimal frequency intervals, which depend on the degree of oversampling.

7.2 A variant of LSE

The objective function in equation (34) is based on the difference between the observables (periodogram values) and the model (expected periodogram values). If the observed values have different variances and/or are interrelated, then weighting of the terms in the sum (34) should, in theory, improve estimation. The weighted LS objective function is

$$\Psi(\boldsymbol{I}, \boldsymbol{\theta}) = \boldsymbol{z}' \boldsymbol{\Sigma}^{-1} \boldsymbol{z}, \tag{48}$$

where $\boldsymbol{z}$ is a column vector with components

$$z_j = I(\nu_j) - S_*(\nu_j, \boldsymbol{\theta})$$

and $\boldsymbol{\Sigma}_{ij} = \mathrm{cov}[I(\nu_i), I(\nu_j)]$.

Analogous to equations (36), the estimator

$$\widehat{\sigma_\epsilon^2} = \boldsymbol{I}' \boldsymbol{\Sigma}_*^{-1} \boldsymbol{I} / \boldsymbol{I}' \boldsymbol{\Sigma}_*^{-1} \boldsymbol{S}_{**} \tag{49}$$

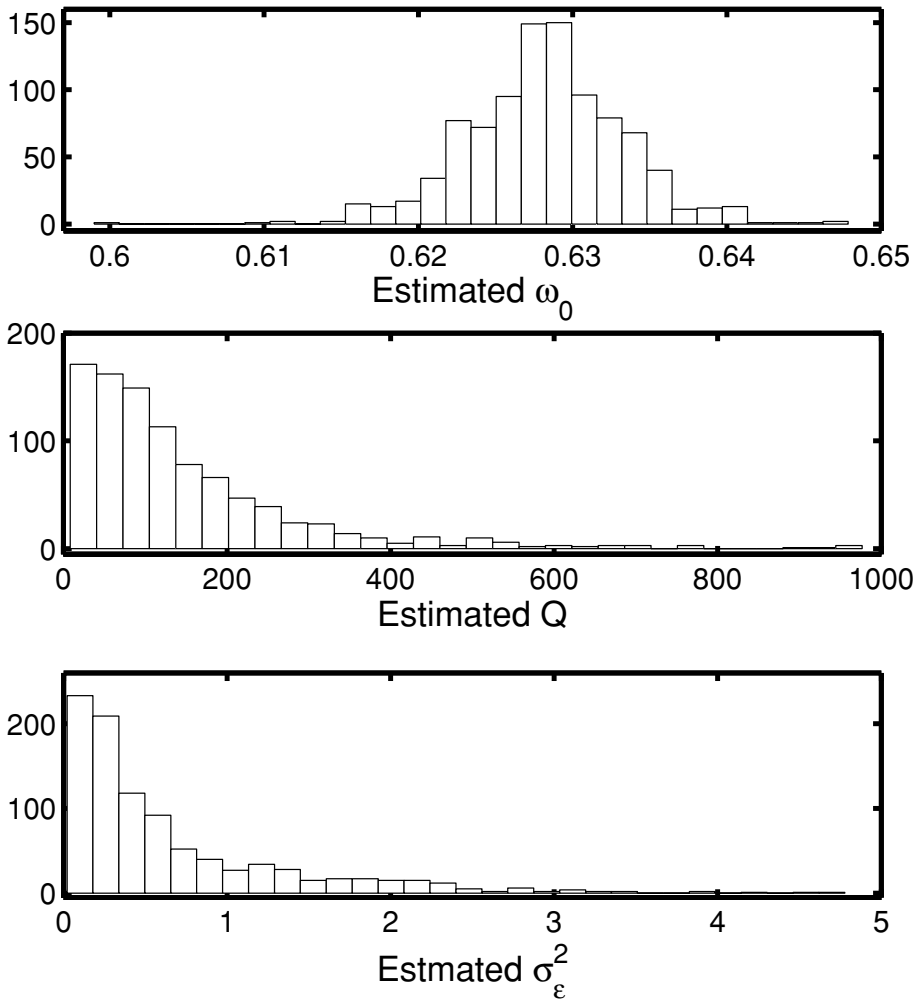

Figure 7. As for Fig. 6, but showing results of LSE. The results for 48 (out of the total 1000) simulated data sets for which $\widehat{Q} \rightarrow \infty$ are excluded from the histograms.

follows from equation (48). In equation (49), $\boldsymbol{I}$ and $\boldsymbol{S}_{**}$ are column vectors with components $I(\nu_j)$ and $S_{**}(\nu_j)$, respectively, and

$\boldsymbol{\Sigma}_* = \boldsymbol{\Sigma}/\sigma_\epsilon^2$.

Simulation experiments show that the estimates obtained from weighted LS are far worse than those obtained from unweighted, ordinary LS. This result seems, at first glance, to be counter-intuitive: incorporation of extra information (in the form of the periodogram covariances) should lead to superior estimates. This overlooks the fact that the periodogram covariance matrix is itself subject to considerable uncertainty. The objective function (48) can be written as

$$\Psi(\boldsymbol{I}, \boldsymbol{\theta}) = \left(\sigma_\epsilon^{-2}\boldsymbol{I} - \boldsymbol{S}_{**}\right)' \boldsymbol{\Sigma}_*^{-1} \left(\sigma_\epsilon^{-2}\boldsymbol{I} - \boldsymbol{S}_{**}\right) \tag{50}$$

which demonstrates a strong dependence on σ_ϵ^2. However, equation (49) shows that $\widehat{\sigma_\epsilon^2}$ is a much more sensitive function of the inconsistent spectral estimator $I(\nu)$ than its QMLE or LSE counterparts. In fact, simulation results show that the estimator (49) is badly biased, and also has large variance.

8 THE TIME-DOMAIN ESTIMATOR

The SDE (5) can also be fitted to the data without transformation to the frequency-domain, by making use of the autocovariance function $C(\ell)$ in a ML procedure. One advantage is that the method can be applied without modification to irregularly sampled time-series. The common approach proceeds recursively through the data, using Kalman filtering (e.g. Jones 1981; Brockwell 2001). For 'small' data sets, it is feasible, and conceptually simpler, to work with the full covariance matrix $\boldsymbol{\Sigma}$ of the entire data set. The Gaussian log-likelihood function is

$$\mathcal{L} = -\frac{1}{2}\left[N \log 2\pi + \log|\boldsymbol{\Sigma}| + (\boldsymbol{y} - \mu\boldsymbol{e})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mu\boldsymbol{e})\right], \tag{51}$$

where $\boldsymbol{e}$ is a column of N elements equal to unity.

Table 2. A summary of the results of some simulation experiments designed to study the effects on QMLE of oversampling the periodogram, and of restricting the frequency interval used for estimation. The second column gives the degree of oversampling, and the following three columns give the mean values of the estimated parameters. The standard deviations of the estimates are given as the first number in brackets; the second number is the SE calculated from the sandwich estimators. The last column lists the number of solutions for which $\widehat{Q} \to \infty$ (first model), or the maximum value of $\widehat{Q}$ (second model). Each line in the table reports the results of 1000 simulations from the particular model.

Frequency range	β	ω_0	Q	σ_ϵ^2	#($Q > 999$)/max($\widehat{Q}$)
Model 1: $\nu = 0.628\,32$, $Q = 100$, $\sigma_\epsilon^2 = 1$, $N = 1000$					
(0, 0.5)	1	0.6282 (0.0022, 0.0019)	116 (149, 184)	0.90 (0.25, 0.30)	27
(0, 0.5)	2	0.6283 (0.0028, 0.0018)	199 (211, 127)	0.80 (0.28, 0.42)	15
(0, 0.5)	3	0.6283 (0.0021, 0.0024)	185 (206, 118)	0.83 (0.24, 0.30)	18
(0, 0.2]	1	0.6284 (0.0021, 0.0020)	123 (119, 82)	0.94 (0.26, 0.25)	34
(0, 0.2]	2	0.6285 (0.0022, 0.0018)	179 (153, 79)	0.87 (0.21, 0.22)	5
(0, 0.2]	3	0.6283 (0.0023, 0.0019)	181 (156, 73)	0.88 (0.22, 0.20)	0
[0.09, 0.11]	1	0.6282 (0.0021, 0.0020)	116 (102, 70)	1.03 (0.43, 0.42)	20
[0.09, 0.11]	2	0.6282 (0.0021, 0.0019)	140 (135, 67)	0.99 (0.38, 0.34)	8
[0.09, 0.11]	3	0.6283 (0.0021, 0.0019)	127 (123, 67)	1.01 (0.37, 0.35)	1
[0.09, 0.11]	4	0.6284 (0.0021, 0.0019)	136 (130, 67)	0.99 (0.37, 0.35)	1
Model 2: $\nu = 0.628\,32$, $Q = 10$, $\sigma_\epsilon^2 = 1$, $N = 500$					
(0, 0.5)	1	0.6277 (0.0083, 0.0088)	14.6 (15.8, 4.9)	0.91 (0.14, 0.21)	186
(0, 0.5)	2	0.6284 (0.0088, 0.0084)	16.2 (25.1, 3.9)	0.92 (0.16, 0.16)	276
(0, 0.5)	3	0.6278 (0.0081, 0.0084)	13.4 (14.1, 3.9)	0.93 (0.13, 0.16)	192
[0.05, 0.15]	1	0.6284 (0.0093, 0.0087)	10.4 (3.2, 3.1)	1.00 (0.19, 0.19)	27
[0.05, 0.15]	2	0.6286 (0.0092, 0.0085)	10.6 (3.2, 3.1)	0.99 (0.18, 0.18)	25
[0.05, 0.15]	3	0.6284 (0.0088, 0.0085)	10.4 (3.2, 3.1)	1.01 (0.19, 0.18)	26

It is convenient to define

$$\boldsymbol{\Sigma}_* \equiv \boldsymbol{\Sigma}/\sigma_\epsilon^2 \,,$$

so that

$$\mathcal{L} = -\frac{1}{2}\left[N \log 2\pi\sigma_\epsilon^2 + \log|\boldsymbol{\Sigma}_*| + (\boldsymbol{y} - \mu\boldsymbol{e})'\boldsymbol{\Sigma}_*^{-1}(\boldsymbol{y} - \mu\boldsymbol{e})/\sigma_\epsilon^2\right]. \tag{52}$$

This allows easy derivation of the estimator

$$\widehat{\sigma_\epsilon^2} = \frac{1}{N}(\boldsymbol{y} - \mu\boldsymbol{e})'\boldsymbol{\Sigma}_*^{-1}(\boldsymbol{y} - \mu\boldsymbol{e}) \tag{53}$$

for the variance σ_ϵ^2. An explicit estimator for the mean is also readily written down:

$$\widehat{\mu} = \boldsymbol{e}'\boldsymbol{\Sigma}_*^{-1}\boldsymbol{y}/\boldsymbol{e}'\boldsymbol{\Sigma}_*^{-1}\boldsymbol{e} \,. \tag{54}$$

This leaves the two unknowns ν_0 and Q to be found, by maximization of

$$\mathcal{L} = -\frac{1}{2}\left[N\left(1 + \log 2\pi\widehat{\sigma_\epsilon^2}\right) + \log|\boldsymbol{\Sigma}_*|\right]. \tag{55}$$

In principle, maximization of equation (55) is straightforward, but in practice both the computational burden and the memory requirements posed by large data sets are formidable, since the number of entries in $\boldsymbol{\Sigma}_*$ is N^2. Fortunately, the matrix is of Toeplitz form:

$$\boldsymbol{\Sigma} = \begin{bmatrix} C(0) & C(1) & C(2) & \cdots & C(N-2) & C(N-1) \\ C(1) & C(0) & C(1) & \cdots & C(N-3) & C(N-2) \\ C(2) & C(1) & C(0) & \cdots & C(N-4) & C(N-3) \\ \vdots & & & & & \vdots \\ C(N-1) & C(N-2) & C(N-3) & \cdots & C(1) & C(0) \end{bmatrix} \tag{56}$$

which allows use of fast, low-storage algorithms. Determination of $\boldsymbol{\Sigma}_*^{-1}$ is not necessary: note, for example, that $\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$ is the solution $\boldsymbol{x}$ of the system of equations

$$\boldsymbol{\Sigma}\boldsymbol{x} = \boldsymbol{y} \,. \tag{57}$$

The solution of equation (57) can be be found by use of the well-known Levinson (1947) algorithm (see e.g. Golub & van Loan 1983). A recipe for the rapid evaluation of the determinants of Toeplitz matrices is given by Zhang, Leithead & Leith (2005).

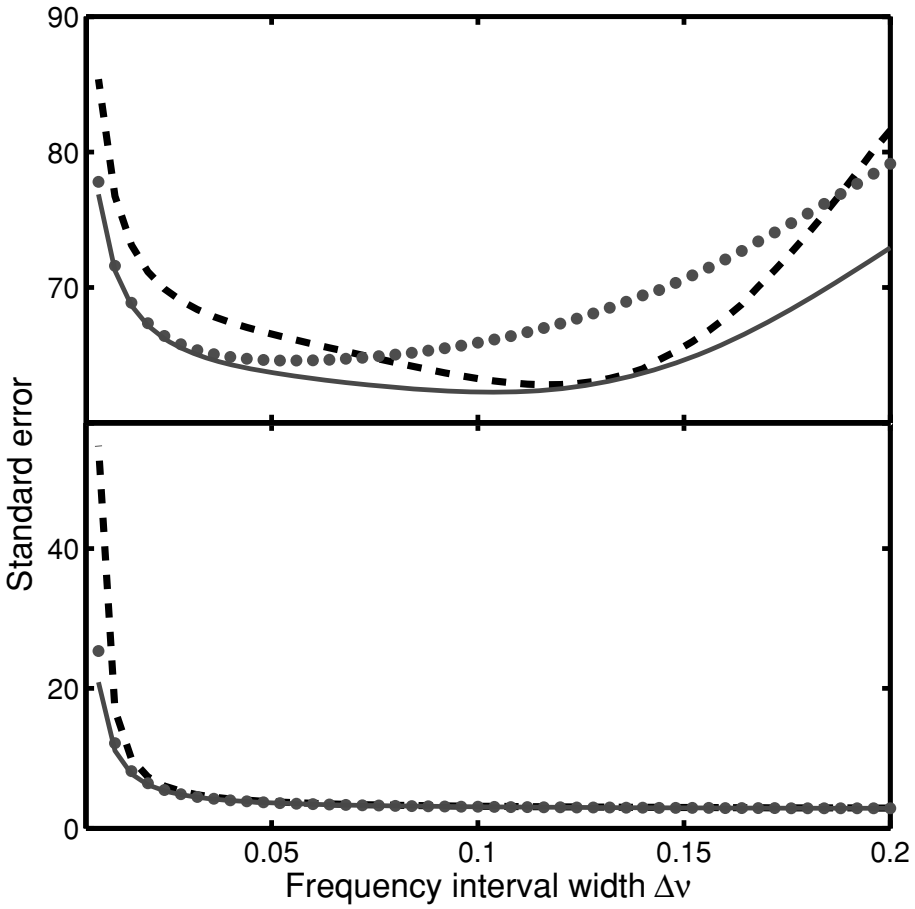

Figure 8. Sandwich SEs of $\widehat{Q}$ for QMLE over a restricted frequency interval of width $\Delta\nu$, centred on $\nu_0 = 0.1$. Top panel: ($Q = 100$, $N = 1000$). Bottom panel: ($Q = 10$, $N = 500$). The different line styles show the effects of different degrees of oversampling of the periodogram. Broken lines – no oversampling; dots – $\beta = 2$; solid line – $\beta = 5$. On these scales, results for $\beta = 3$ and 8 are indistinguishable from those for $\beta = 5$.

Asymptotically, the covariance matrix of the ML estimates is given by the inverse of the Fisher information matrix $\mathbf{F}$. The entries in the latter are (Koen 2005, equation A3):

$$F(i,j) = \frac{\partial\mu}{\partial\theta_i}\boldsymbol{e}'\boldsymbol{\Sigma}^{-1}\boldsymbol{e}\frac{\partial\mu}{\partial\theta_j} + \frac{1}{2}\mathrm{trace}\left[\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\right]. \tag{58}$$

Letting

$$\boldsymbol{\theta} = \left[\omega_0 \;\; Q \;\; \sigma_\epsilon^2 \;\; \mu\right], \tag{59}$$

it follows that

$$\begin{aligned}
F(1,1) &= \frac{1}{2}\mathrm{trace}\left[\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\omega_0}\right]^2,\\
F(1,2) = F(2,1) &= \frac{1}{2}\mathrm{trace}\left[\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\omega_0}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial Q}\right],\\
F(1,3) = F(3,1) &= \frac{1}{2\sigma_\varepsilon^2}\mathrm{trace}\left[\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\omega_0}\right],\\
F(2,2) &= \frac{1}{2}\mathrm{trace}\left[\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial Q}\right]^2\\
F(2,3) = F(3,2) &= \frac{1}{2\sigma_\varepsilon^2}\mathrm{trace}\left[\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial Q}\right],\\
F(3,3) &= \frac{N}{2\sigma_\epsilon^2},\\
F(4,4) &= \boldsymbol{e}'\boldsymbol{\Sigma}^{-1}\boldsymbol{e},\\
F(4,k) = F(k,4) &= 0\,, \qquad k = 1,2,3.
\end{aligned} \tag{60}$$

The derivatives in equations (60) are not difficult to calculate:

$$\frac{\partial C(\ell)}{\partial \omega_0} = -C(\ell)\left(\frac{3}{\omega_0} + \frac{\ell}{2Q}\right) - \ell C(0)\mathrm{e}^{-\ell/\tau}\sin(\omega_*\ell + \phi),$$
$$\frac{\partial C(\ell)}{\partial Q} = \frac{C(\ell)}{Q}\left[1 - \left(\frac{\omega_0}{2\omega_* Q}\right)^2 + \frac{\ell\omega_0}{2Q}\right] - \frac{\sigma_\epsilon^2}{4Q\omega_0\omega_*^2}\left(1 + \frac{\ell\omega_0}{2Q}\right)\mathrm{e}^{-\ell/\tau}\sin(\omega_*\ell + \phi),$$
$$\frac{\partial C(\ell)}{\partial \sigma_\epsilon^2} = \frac{C(\ell)}{\sigma_\epsilon^2}. \tag{61}$$

9 THE TIME-DOMAIN ESTIMATOR: RESULTS

Illustrative results are shown in Fig. 9 and Table 3, which may be compared to the corresponding frequency-domain results in Fig. 6 and Table 1. It is clear that

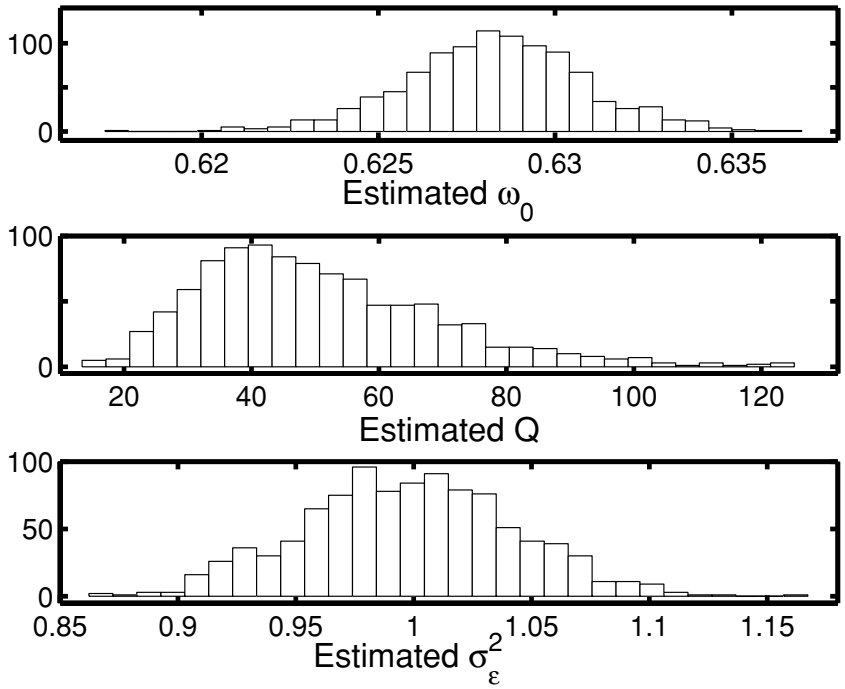

Figure 9. As for Fig. 6, but for time-domain estimation. All 1000 estimates of Q were finite.

Table 3. The results of some simulation experiments to test the efficacy of time-domain estimation of model parameters. For each of the five models, the first line states the assumed parameter values, while the second summarizes the results of 1000 simulations. Mean estimated parameters are followed by standard deviations (first number in brackets) and SEs from the inverse Fisher information matrix (second number in brackets). All models used $\mu = 0$ and $\sigma_\epsilon^2 = 1$.

ω_0	Q	σ_ϵ^2	μ
$\omega = 0.628\,32$, $Q = 50$, $N = 1000$			
0.6283 (0.0025, 0.0025)	50.6 (18.7, 18.7)	0.998 (0.045, 0.045)	0.000 (0.079,0.080)
$\omega = 0.628\,32$, $Q = 100$, $N = 1000$			
0.6284 (0.0018, 0.0019)	105.2 (53.2, 49.4)	1.001 (0.046, 0.044)	0.000 (0.079,0.080)
$\omega = 0.628\,32$, $Q = 50$, $N = 500$			
0.6283 (0.0040, 0.0036)	50.0 (26.0, 24.8)	0.997 (0.067, 0.064)	0.0019 (0.11,0.11)
$\omega = 0.314\,16$, $Q = 50$, $N = 1000$			
0.3143 (0.0020, 0.0018)	50.5 (26.1, 24.7)	1.000 (0.046, 0.045)	−0.001 (0.318,0.321)
$\omega = 0.628\,32$, $Q = 5$, $N = 1000$			
0.6286 (0.0080, 0.0079)	5.03 (0.66, 0.67)	0.998 (0.048, 0.048)	0.003 (0.076,0.080)

(i) estimation of the quality factor, and the driving noise variance, is much more efficient in the time-domain than in the frequency-domain, for all parameter combinations considered. Not only are the SEs smaller, but time-domain estimates are also unbiased. SEs of $\widehat{\omega}_0$ are marginally smaller than the corresponding frequency-domain estimates;

(ii) the problem of diverging $\widehat{Q}$ encountered in the frequency-domain, for large Q, is not seen in the time-domain. Comparison of Figs 6 and 9 shows that the distribution of time-domain-estimated Q does not have the extended tail seen in the frequency-domain distribution;

(iii) there is very good agreement between the time-domain asymptotic error estimates, and the SEs determined by simulation. This contrasts sharply with the frequency-domain results;

(iv) SEs as calculated from inverse Fisher matrices are lower limits attainable by *any* unbiased estimator (the Cramér–Rao lower bounds – e.g. Cox & Hinkley 1974). Comparison of Tables 1 and 3 shows that the frequency-domain Cramér–Rao bounds on $\widehat{Q}$ are smaller than their time-domain counterparts, except for the model with $Q = 5$. Time-domain estimators of ω_0 have slightly lower bounds, and time-domain estimators of σ_ϵ^2 substantially smaller bounds, than their frequency-domain counterparts.

Point (iii) implies that accurate SEs can be simply calculated from $\mathbf{F}^{-1}$, where the entries in the matrix are given by equation (59).

10 THE EFFECTS OF MEASUREMENT ERRORS

The general time-series form of the measurement error process $e(t)$ may be known (e.g. red noise), such that its spectrum and covariance matrix can be written down. The error spectrum or covariance matrix may contain a number of unknown parameter values, which specify the particular properties of the measurement noise within the general form (e.g. an autoregressive parameter giving the specific red noise form). Provided noise and signal are independent, the noise spectrum can simply be added to $S(\omega)$ in equation (7), or the noise covariance can be added to $\mathbf{\Sigma}$ in equation (56):

$$\begin{aligned} S_c(\omega) &= S(\omega) + S_e(\omega), \\ \mathbf{\Sigma}_c &= \mathbf{\Sigma} + \mathbf{\Sigma}_e, \end{aligned} \tag{62}$$

where $S(\omega)$ and $\mathbf{\Sigma}$ are given by equations (9) and (56). The subscript c denotes the spectrum/covariance matrix of the noise-contaminated observations

$$y_c(t) = y(t) + e(t).$$

The notation

$$C_e(\ell) = \mathrm{E}[e(t)e(t+\ell)]$$

will be used below for the autocovariance of the measurement noise.

10.1 Frequency-domain estimation

In general

$$\mathrm{E}I_e(\omega) \neq S_e(\omega)$$

will hold, and QMLE should be based on

$$\mathrm{E}I_c(\omega) = \mathrm{E}I(\omega) + \mathrm{E}I_e(\omega)\ . \tag{63}$$

The last term in equation (63) can be calculated from equation (19) or (20), using the appropriate forms of $S_e(\omega)$ or $C_e(\omega)$. Note that it is no longer possible to obtain an explicit estimator such as equations (36) for σ_ϵ^2: the quasi-likelihood needs to be maximized with respect to all three parameters in equations (9), as well as any unspecified parameters in the measurement error spectrum.

Expressions (24) and (26)–(29) for the covariances of the periodogram remain valid, with $C_c(\ell)$ replacing $C(\ell)$. Contributions of the error $e(t)$ and signal $y(t)$ to the periodogram covariances are not easily separable in these equations. A much simpler approach follows from equations (62), subject to the independence of signal and noise:

$$\mathrm{cov}[S_c(\nu), S_c(\lambda)] = \mathrm{cov}[S(\nu), S(\lambda)] + \mathrm{cov}[S_e(\nu), S_e(\lambda)]\ .$$

In the case of white noise, the last term reduces to

$$\mathrm{cov}[S_e(\nu), S_e(\lambda)] = \sigma_e^4 \delta(\nu, \lambda)\ .$$

The forms (38), (43) and (44), for the sandwich estimator of the QMLE covariances, still hold, but with S_{c*} replacing S_* everywhere. The left-hand sides of equations (46) are replaced by derivatives of S_{c*}, rather than S_*; the right-hand sides remain unchanged. Equations (46) are furthermore supplemented by derivatives of S_{c*} with respect to unspecified parameters of the measurement noise process. For example, if $e(t)$ is white noise with unknown variance $S_e(\omega) = \sigma_e^4$,

$$\frac{\partial S_{c*}(\omega)}{\partial \sigma_e^2} = 1. \tag{64}$$

10.2 Time-domain estimation

The log-likelihood $\mathcal{L}$ is given by equation (51), with $\mathbf{\Sigma}_c$ replacing $\mathbf{\Sigma}$. As in the case of frequency-domain estimation, it is not possible to derive an explicit estimator for σ_ϵ^2. The entries in the covariance matrix (56) are replaced by $C_c(\ell) = C(\ell) + C_e(\ell)$. In the case of white noise, only the diagonals are affected: the $C(0)$ are replaced by $C(0) + \sigma_e^2$.

If the measurement error process can be characterized by a single parameter θ_5 which needs to be estimated along with the 'signal' parameters, define

$$\boldsymbol{\theta} = \left[\omega_0 \;\; Q \;\; \sigma_\epsilon^2 \;\; \mu \;\; \theta_5\right]' .$$

Equations (59) for the information matrix are then replaced by

$$\begin{aligned}
F(1,1) &= \frac{1}{2}\mathrm{trace}\left[\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\omega_0}\right]^2,\\
F(1,2) = F(2,1) &= \frac{1}{2}\mathrm{trace}\left[\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\omega_0}\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}}{\partial Q}\right],\\
F(1,3) = F(3,1) &= \frac{1}{2\sigma_\epsilon^2}\mathrm{trace}\left[\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\omega_0}\mathbf{\Sigma}_c^{-1}\mathbf{\Sigma}\right],\\
F(2,2) &= \frac{1}{2}\mathrm{trace}\left[\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}}{\partial Q}\right]^2,\\
F(2,3) = F(3,2) &= \frac{1}{2\sigma_\epsilon^2}\mathrm{trace}\left[\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}}{\partial Q}\mathbf{\Sigma}_c^{-1}\mathbf{\Sigma}\right],\\
F(3,3) &= \frac{1}{2\sigma_\epsilon^4}\mathrm{trace}\left[\mathbf{\Sigma}_c^{-1}\mathbf{\Sigma}\right]^2,\\
F(4,4) &= \boldsymbol{e}'\mathbf{\Sigma}_c^{-1}\boldsymbol{e},\\
F(4,k) = F(k,4) &= 0\,, \qquad k = 1,2,3,5,\\
F(1,5) = F(5,1) &= \frac{1}{2}\mathrm{trace}\left[\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\omega_0}\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}_e}{\partial\theta_5}\right],\\
F(2,5) = F(5,2) &= \frac{1}{2}\mathrm{trace}\left[\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}}{\partial Q}\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}_e}{\partial\theta_5}\right],\\
F(3,5) = F(5,3) &= \frac{1}{2\sigma_\epsilon^2}\mathrm{trace}\left[\mathbf{\Sigma}_c^{-1}\mathbf{\Sigma}\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}_e}{\partial\theta_5}\right],\\
F(5,5) &= \frac{1}{2}\mathrm{trace}\left[\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}_e}{\partial\theta_5}\right]^2 .
\end{aligned} \tag{65}$$

As an example, for white noise with unknown variance σ_e^2,

$$\frac{\partial\mathbf{\Sigma}_e}{\partial\theta_5} = \boldsymbol{I}.$$

Note the simplification

$$\begin{aligned}
\mathbf{\Sigma}_c^{-1}\mathbf{\Sigma} &= \mathbf{\Sigma}_c^{-1}(\mathbf{\Sigma}_c - \mathbf{\Sigma}_e)\\
&= \boldsymbol{I} - \sigma_e^2\mathbf{\Sigma}_c^{-1}
\end{aligned}$$

so that

$$\begin{aligned}
F(3,1) &= \frac{1}{2\sigma_\epsilon^2}\mathrm{trace}\left[\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\omega_0}\right] - \frac{\sigma_e^2}{2\sigma_\epsilon^2}\mathrm{trace}\left[\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\omega_0}\mathbf{\Sigma}_c^{-1}\right],\\
F(3,2) &= \frac{1}{2\sigma_\epsilon^2}\mathrm{trace}\left[\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}}{\partial Q}\right] - \frac{\sigma_e^2}{2\sigma_\epsilon^2}\mathrm{trace}\left[\mathbf{\Sigma}_c^{-1}\frac{\partial\mathbf{\Sigma}}{\partial Q}\mathbf{\Sigma}_c^{-1}\right],\\
F(3,3) &= \frac{1}{2\sigma_\epsilon^2}\left[N - 2\sigma_e^2\mathrm{trace}\left(\mathbf{\Sigma}_c^{-1}\right) + \sigma_e^4\mathrm{trace}\left(\mathbf{\Sigma}_c^{-2}\right)\right],\\
F(3,5) &= \frac{1}{2\sigma_\epsilon^2}\left[\mathrm{trace}\left(\mathbf{\Sigma}_c^{-1}\right) - \sigma_e^2\mathrm{trace}\left(\mathbf{\Sigma}_c^{-2}\right)\right].
\end{aligned} \tag{66}$$

If the measurement noise process is fully specified, with no unknown parameters, the vector $\boldsymbol{\theta}$ only has four components (as in Section 8), and the last four equations in equations (65) are ignored.

Equations (61) are unaffected.

10.3 Some illustrative simulation results

Only the simplest case of white noise with variance σ_e^2 is considered here: the noise spectrum is then

$$S_e(\omega) = \sigma_e^2 \tag{67}$$

and the covariance matrix is

$$\mathbf{C}_e = \sigma_e^2 \mathbf{I}, \tag{68}$$

where $\mathbf{I}$ is the identity matrix. There are two possibilities: either σ_e^2 can be assumed known, or it needs to be estimated as a fifth unknown parameter value.

Simulation results are summarized in Table 4. Entries may be compared to those for noise-free models in Tables 1 and 3. Values of σ_e were chosen such that the 'signal-to-noise' ratio $\sqrt{C(0)}/\sigma_e = 2$ for a given set of parameter values.

(i) Results are only reported for the case of unspecified σ_e, because parameter estimates are very similar if σ_e is assumed known.

(ii) For those models in Table 1 for which divergent estimates of Q were encountered, the problem is generally slightly worse in Table 4. Similarly to Table 3, time-domain estimates do not diverge if $\sigma_e \neq 0$.

(iii) Frequency-domain estimates of Q have much smaller bias and SEs in the presence of measurement error. Although SEs of $\widehat{\sigma_\epsilon^2}$ are increased, the estimates are unbiased if σ_e^2 is non-zero. Estimates of ω_0 and σ_e^2 are also unbiased.

(iv) In the time-domain, all parameter estimates remain essentially unbiased. SEs of $\widehat{\omega}_0$ and $\widehat{Q}$ are slightly inflated, while those of $\widehat{\sigma_\epsilon^2}$ are typically increased by a factor of ∼4 over Table 3 values.

(v) Frequency- and time-domain estimates are much more similar if there are measurement errors, but SEs of $\widehat{Q}$ estimated in the time-domain remain smaller (except for small Q).

(vi) In the time-domain, the asymptotic SEs are only slightly smaller than the values determined by simulation. The same applies to the frequency-domain, with the exception of the parameter Q.

Table 4. A summary of the results of some simulation experiments illustrating the effects of measurement errors. For each of the five models, the parameter values are given in the second column of the table. Frequency-domain QML parameter estimates are given in the third column, with standard deviations in brackets. The fourth column contains QMLE SEs, according to the sandwich estimator (38). The fifth and sixth columns give time-domain estimation results: estimates (followed by standard deviations in brackets) and SEs calculated from the inverse Fisher matrix (SE 2). The last line of information for each of the five models shows the sample size and the fraction of finite-Q solutions for QMLE. Estimates derived from divergent solutions have not been included in the reported simulation results.

Parameter	True value	QML estimate	SE 1	Time-domain	S.E.2
ω_0	0.628 32	0.6284 (0.0030)	0.0027	0.6284 (0.0029)	0.0027
Q	50	53 (27)	23	52 (22)	20.4
σ_ϵ^2	1	1.01 (0.24)	0.22	1.00 (0.17)	0.17
σ_e^2	25	25.01 (1.38)	1.28	24.98 (1.34)	1.30
$N = 1000$	$p(\text{QMLE}) = 0.996$				
ω_0	0.628 32	0.6284 (0.0023)	0.0020	0.6283 (0.0021)	0.0019
Q	100	111 (79)	66	104 (57)	53
σ_ϵ^2	1	1.00 (0.35)	0.34	1.00 (0.21)	0.20
σ_e^2	50.41	50.41 (2.45)	2.50	50.34 (2.49)	2.53
$N = 1000$	$p(\text{QMLE}) = 0.951$				
ω_0	0.628 32	0.6284 (0.0047)	0.0040	0.6284 (0.0047)	0.0038
Q	50	59 (57)	33	53 (32)	27
σ_ϵ^2	1	1.02 (0.40)	0.36	1.00 (0.25)	0.25
σ_e^2	25	25.06 (1.91)	1.84	24.96 (1.84)	1.84
$N = 500$	$p(\text{QMLE}) = 0.958$				
ω_0	0.314 16	0.3142 (0.0024)	0.0020	0.3142 (0.0022)	0.0019
Q	50	60 (71)	32	52 (31)	27
σ_ϵ^2	1	1.01 (0.38)	0.33	1.00 (0.21)	0.20
σ_e^2	201.64	201.47 (9.70)	9.92	201.40 (9.90)	9.83
$N = 1000$	$p(\text{QMLE}) = 0.961$				
ω_0	0.628 32	0.0.6283 (0.0102)	0.0089	0.6287 (0.0097)	0.0096
Q	5	5.06 (0.83)	0.80	5.06 (0.86)	0.83
σ_ϵ^2	1	1.00 (0.13)	0.11	1.00 (0.13)	0.13
σ_e^2	2.56	2.56 (0.15)	0.15	2.56 (0.16)	0.15
$N = 1000$	$p(\text{QMLE}) = 1.0$				

11 MISSING VALUES

In the case of time-domain estimation, the basic theory in Sections 8 and 10.2 is unaffected. The only change is that each missing value results in the removal of one row, and one column, from the covariance matrix (56). This has the important practical consequence that the matrix no longer has the Toeplitz form – instead, it is block Toeplitz.

QML estimates in the frequency-domain can still be obtained from equation (3), with the spectrum S replaced by the expected periodogram $\mathrm{E}I(\nu)$. The periodogram is calculated as in equation (1), where the summation extends over the available observations. The expected periodogram is given by the penultimate line in equations (20), again with the summation taken over the available values. Evaluation of $\mathrm{E}I(\omega)$ is computationally expensive, and if there are only a few missing values, then the alternative

$$\begin{aligned}\mathrm{E}I(\omega) &= \frac{1}{N}\sum_{j\in O}\sum_{k\in O} C(|j-k|)\mathrm{e}^{-\mathrm{i}\omega(j-k)} \\ &= \frac{N_O}{N}C(0) + \frac{2}{N}\sum_{\ell=1}^{N}(N-\ell)C(\ell) - \frac{2}{N}\sum_{j\in M}\sum_{k\in O} C(|j-k|)\cos(j-k)\omega \\ &\quad -\frac{2}{N}\sum_{j\in M}\sum_{k\in M, k>j} C(k-j)\cos(j-k)\omega \end{aligned} \tag{69}$$

can be evaluated more quickly. In equations (69), O and M are, respectively, the sets of time-points of observations and missing values, and N_O is the number of observed values (the cardinality of O). If there are measurement errors, then the expected noise periodogram should also be revised: for example, for white noise with variance σ_e^2,

$$\mathrm{E}I_e(\omega) = \frac{N_O}{N}\sigma_e^2 \,. \tag{70}$$

The summation in the covariance formula (21) proceeds over indices in the set O only. Equation (24) still applies, with the summation in equations (23) again restricted to the set of observed values. In general, further simplification as in equations (25)– (31) is no longer possible.

12 AN EXAMPLE

Fig. 10 (top panel) shows the periodogram of a 2.1-hour photoelectric photometric run on the pulsating hot subdwarf star PG 0048+091 (Koen et al. 2004). Observations were obtained at 10 s intervals: for simplicity, this is adopted as the unit of time. There are $N_O = 694$ measurements over $N = 739$ time-points, with 45 missing values occurring in five distinct blocks. The data have been pre-whitened by a polynomial, in order to remove the low frequency effects of atmospheric transparency changes.

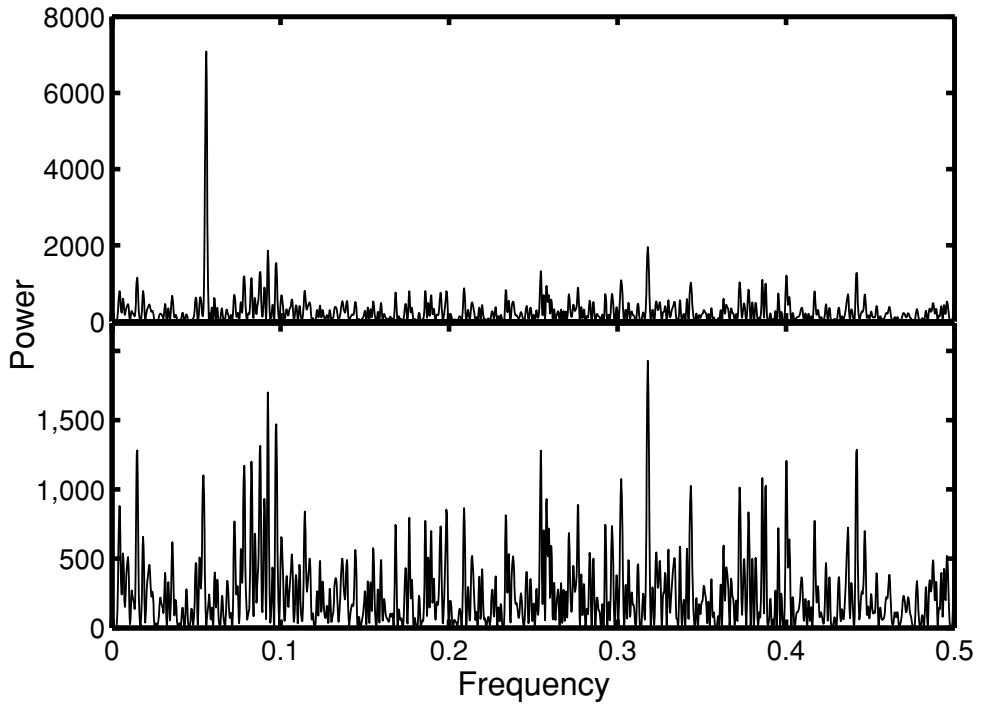

Figure 10. Top panel: the periodogram of 2.1 hours of high-speed photometry of the pulsating star PG 0048+091. Bottom panel: the periodogram of the residuals after subtracting a LS-fitted sinusoid with frequency 5.6 mHz. Note the different vertical scales of the two panels of the plot.

Table 5. Estimated parameters of the second-order SDE fitted to observations of the pulsating subdwarf star PG 0048+091. Allowance has been made for the presence of measurement errors (last column).

Method	ν_0 (mHz)	Q	σ_ϵ (mmag)	σ_e (mmag)
Frequency-domain	5.59	316	0.069	15.9
Time-domain	5.59	184	0.094	15.8

The periodogram of the observations is dominated by the peak at $\nu = 0.056$ cycles $(10\,\mathrm{s})^{-1}$ (i.e. 5.6 mHz). The bottom panel shows the residual spectrum after pre-whitening the data by a fitted deterministic sinusoid.

Table 5 summarizes the results of fitting equation (5) to the observations, allowing for white noise measurement errors. The results suggest that the oscillations may be stochastic, rather than deterministic. The implied mode lifetimes are ∼59 (time-domain estimate) and ∼101 cycles (frequency-domain estimate). Given the period $P_0 = 179$ s, the respective lifetime estimates are ∼2.9 and ∼5.0 h, rather longer than the duration of the observing run. The implication is that the values of Q are very uncertain – a point also made by the substantial difference between the time- and frequency-domain estimates.

It should be noted that the 'signal-to-noise' ratio $\sqrt{C(0)}/\sigma_e = 0.26$ and 0.27 for the frequency- and time-domain estimates in Table 5, respectively, that is, about a factor of 8 smaller than for the simulations reported in Table 4.

Reed et al. (2007) examined 167 hours of photometry of the star, accumulated during a 16-day multisite observing campaign, and found 'a plethora of frequencies with short pulsation lifetimes', which supports the conclusion reached here.

We digress a little to discuss testing whether oscillatory data buried in white measurement errors are due to a single deterministic sinusoid, as opposed to a stochastic variation such as that described by equation (5). [The stipulation of a single sinusoid is necessary, in order to rule out the possibility of beating between several sinusoids with closely similar frequencies, which could have the appearance of stochastic cycles in short data trains.] The deterministic case, corresponding to $Q \to \infty$ in equation (5), is described by

$$y(t) = A\cos(\omega_0 t + \psi) + e(t)\ . \tag{71}$$

The corresponding frequency-domain quasi-likelihood is

$$\mathcal{L}_0 = -\sum_{j\in O}\left[\log \mathrm{E}I_0(\omega_j) + \frac{I(\omega_j)}{\mathrm{E}I_0(\omega_j)}\right],$$

$$\mathrm{E}I_0(\omega) = \frac{N_O}{N}\sigma_e^2 + \frac{A^2}{N}\left|\sum_{j\in O}\mathrm{e}^{-ij\omega}\cos(j\omega_0+\psi)\right|^2, \tag{72}$$

where, as in Section 11, O is the set of times at which observations were acquired. For stochastic oscillations,

$$\mathcal{L}_1 = -\sum_{j\in O}\left[\log \mathrm{E}I_1(\omega_j) + \frac{I(\omega_j)}{\mathrm{E}I_1(\omega_j)}\right] \tag{73}$$

with $\mathrm{E}I_1(\omega)$ given by the sum of the expected values in equation (69) and (70).

In the time-domain, equation (71) implies that

$$\mu = A\cos(\omega t + \psi) + B,$$

$$\mathbf{\Sigma}_c = \mathbf{\Sigma}_e = \sigma_e\mathbf{I},$$

in equations (51) and (62). The constant B allows for a non-zero mean level. It follows that the Gaussian log-likelihood is

$$\mathcal{L}_0 = -\frac{1}{2}\left\{N_O\log 2\pi + N_O\log\sigma_e^2 + \frac{1}{\sigma_e^2}\sum_{j\in O}[y_j - A\cos(j\omega_0+\psi)]^2\right\}\ .$$

Maximizing with respect to σ_e^2 gives

$$\widehat{\sigma_e^2} = \frac{1}{N_O}\sum_{j\in O}[y_j - A\cos(j\omega_0+\psi) - B]^2 \equiv \frac{1}{N_O}\sum_{j\in O} r_j^2 \tag{74}$$

and therefore

$$\mathcal{L}_0 = -\frac{N_O}{2}\left(\log 2\pi + \log\widehat{\sigma_e^2} + 1\right). \tag{75}$$

For the stochastic alternative, the log-likelihood $\mathcal{L}_1$ is calculated as detailed in Sections 8, 10.2 and 11.

A natural test statistic for discriminating between the two alternatives is the likelihood ratio statistic

$$\Lambda = 2(\max\mathcal{L}_1 - \max\mathcal{L}_0), \tag{76}$$

where the maximization is in each case over the relevant parameter values (ω_0, A, ψ and σ_e^2 in the case of $\mathcal{L}_0$; ω_0, Q, σ_ϵ^2 and σ_e^2 in the case of $\mathcal{L}_1$). In standard statistical applications, Λ has a χ^2 distribution, with the number of degrees of freedom determined by the specifications of

the null hypothesis. At first sight, the hypotheses in the present case are simple:

$$\mathrm{H}_0 : A \neq 0, \quad \sigma_\epsilon = 0,$$

$$\mathrm{H}_1 : A = 0, \quad \sigma_\epsilon \neq 0 .$$

However, this specification is far from standard: parameters under the null and alternative are on the boundaries of their allowable values (which violates regularity conditions), and the hypotheses are not nested. The implication is that the distribution of Λ needs to be determined by alternative means.

A simple procedure is to use bootstrapping, under the null hypothesis that the oscillations are deterministic. The method is as follows:

(i) Calculate the likelihood ratio statistic Λ_* of the data from equation (76), using the forms of $\mathcal{L}_0$ and $\mathcal{L}_1$ given above
(ii) Calculate the residuals r_j ($j \in O$), by substituting the estimated values of $\widehat{\omega}_0$, $\widehat{A}$, $\widehat{B}$ and $\widehat{\psi}$ into equation (74)
(iii) Draw, with replacement, a sample of size N_O from the collection of r_j: denote this by $\{r_j^{(1)} | j \in O\}$
(iv) Construct a bootstrap sample $y_j^{(1)}$ by setting

$$y_j^{(1)} = \widehat{B} + \widehat{A}\cos(\widehat{\omega}_0 j + \widehat{\psi}) r_j^{(J)}, \quad j \in O$$

(v) Calculate the likelihood ratio statistic $\Lambda^{(1)}$ for the bootstrap sample
(vi) Repeat steps (iii)–(v) many (typically several thousand) times, obtaining $\Lambda^{(2)}$, $\Lambda^{(3)}$, ...
(vii) The significance level of Λ_* follows by finding its percentile value in the distribution of $\Lambda^{(k)}$

The procedure is computationally expensive, and only 700 bootstrap samples were generated. The results are surprising – the time-domain value of Λ_* is highly significant ($p = 2.3$ per cent), while the frequency-domain statistic is not significant ($p = 22$ per cent). The reason for the discrepancy is certainly not obvious.

13 SUMMARY

The work in this paper could be contrasted with that in Anderson et al. (1990). Those authors dealt with very large data sets, for which the frequency resolution $\Delta\nu$ was sufficiently small that spectral peaks are well defined even if only the Fourier frequencies (4) are considered. The terms in equation (3) are close to being independent, and standard MLE could be used. The authors gave some consideration to shorter data sets (although still long in terms considered here), and showed the degradation of the estimates due to decrease in resolution (their fig. 3). Here, MLE was replaced by QMLE, and the efficacy of frequency oversampling was studied.

Main results of this paper are listed below. The caveat 'for the parameter values investigated' should be borne in mind throughout.

(i) Two viable schemes, one approximate and one exact, for the simulation of time-series satisfying equation (5), are presented in Section 3.

(ii) In finite samples, the periodogram is a biased estimator of the spectrum, and estimation should be based on the expected value (equation 19 or 20) of the periodogram, rather than the spectrum. Furthermore, for finite samples, the periodogram ordinates are not exponentially distributed (see equation 29), nor are they independent. Formulae for the efficient calculation of periodogram covariances are derived in Section 4.

(iii) Formulae for covariances of frequency-domain estimates, by both LS and QML, are given in Section 5. These 'sandwich estimators' incorporate the periodogram covariances.

(iv) QMLE is a much more efficient frequency-domain estimation method than LSE. Weighted LS results are even poorer than those from ordinary LS.

(v) For large quality factors Q, the frequency-domain estimates of Q and σ_ϵ^2 are biased. For particularly regular-looking time-series, $\widehat{Q} \to \infty$. The distribution of $\widehat{Q}$ in the time-domain is better behaved: bias is smaller and estimates do not diverge.

(vi) For large Q, the standard information matrix gives hopelessly optimistic estimates of QMLE SEs; sandwich estimators, on the other hand, give reasonable results. In the time-domain, SE estimates are quite reliable.

(vii) Frequency-domain estimation can be improved by oversampling the periodogram, and by working with only a small frequency interval around ω_0.

(viii) Interestingly, QMLE estimates of Q benefit from the presence of measurement errors (less bias, smaller SEs). This may be due to the dilution of the correlation between periodogram values at different frequencies. Time-domain results show little difference if moderate measurement noise is added.

(ix) Ignorance of measurement error variance does not impact significantly on estimation.

(x) Comparing the hypothesis of stochastic oscillation to that of deterministic variation is a fallow area of study. Research in this field promises to be interesting, as evidenced by the discrepancy between the time- and frequency-domain significance levels for the PG 0048+091 data analysed in Section 12.

Although the results above demonstrate that, for modest sample size, time-domain estimation may be superior to the frequency-domain approach, it should be stressed that only a limited part of the parameter space was explored. In particular, it may be anticipated that frequency-domain methodology will be particularly suited to the study of multiperiodicities, due to the possibility of separation of the different frequencies of interest.

ACKNOWLEDGMENTS

The author is grateful for a grant from the National Research foundation of South Africa.

REFERENCES

Anderson T. W., 1971, The Statistical Analysis of Time Series. John Wiley & Sons, Inc., New York
Anderson E. R., Duvall T. L., Jefferies S. M., 1990, ApJ, 364, 699
Bedding T. R., Kjeldsen H., 2003, PASA, 20, 203
Brillinger D. R., 1981, Time Series: Data Analysis and Theory. Holden Day, Inc., San Francisco
Brockwell P. J., 2001, in Shanbhag D. N., Rao C. R., eds., Handbook of Statistics. Elsevier Science B.V., Amsterdam, p. 249
Cameron A. C., Trivedi P. K., 2005, Microeconometrics. Methods and Applications. Cambridge Univ. Press, New York
Cox D. R., Hinkley D. V., 1974, Theoretical Statistics. Chapman & Hall, London
Finger M. H., 1998, Adv. Space Res., 22, 1007
Fukumura K., Shrader C. R., Dong J. W., Kazana D., 2010, A&A, 524, A34
Gilsing H., Shardlow T., 2007, J. Comput. Appl. Math., 205, 1002
Golub G. H., Van Loan C. F., 1983, Matrix Computations. North Oxford Academic, Oxford
Healy M. J. R., 1986, Matrices for Statistics. Clarendon Press, Oxford
Jakimiec J., Tomczak M., 2010, Sol. Phys., 261, 233
Jones R. H., 1981, in Findley D. F., ed., Applied Time Series Analysis II. Academic Press, New York, p. 651
Koen C., 2005, MNRAS, 361, 887
Koen C., 2009, MNRAS, 392, 190
Koen C., Lombard F., 2004, MNRAS, 353, 98
Koen C., O'Donoghue D., Kilkenny D., Pollacco D. L., 2004, New Astron., 9, 565
Krogstad H. E., 1982, J. Time Ser. Anal., 3, 195
Levinson N., 1947, J. Math. Phys., 25, 261
Mauche C. W., 2002, ApJ, 580, 423
Phadke M. S., Wu S. M., 1974, J. Am. Stat. Assoc., 69, 325
Reed M. D. et al., 2004, MNRAS, 348, 1164
Reed M. D. et al., 2007, ApJ, 664, 518
Török C., 1994, Comput. Math. Appl., 27, 1
Watts A. L., Strohmayer T. E., 2007, Adv. Space Res., 40, 1446
Wei W. W. S., 1990, Time Series Analysis. Univariate and Multivariate Methods. Addison-Wesley Publishing Company, Redwood City, California
Zhang Y., Leithead W. E., Leith D. J., 2005, Proc. IEEE Conf., Time-series Gaussian Process Regression Based on Toeplitz Computation of O(N^2) Operations and O(N)-level Storage. IEEE Conf. 44 on Decision and Control, p. 3711

This paper has been typeset from a TEX/LATEX file prepared by the author.