J. R. Statist. Soc. B (1997)
59, *No.* 1, *pp.* 255–268

Semiparametric Bayesian Inference for Time Series with Mixed Spectra

By C. K. CARTER and R. KOHN†

University of New South Wales, Sydney, Australia

[Received January 1995. Final revision May 1996]

SUMMARY

A Bayesian analysis is presented of a time series which is the sum of a stationary component with a smooth spectral density and a deterministic component consisting of a linear combination of a trend and periodic terms. The periodic terms may have known or unknown frequencies. The advantage of our approach is that different features of the data — such as the regression parameters, the spectral density, unknown frequencies and missing observations — are combined in a hierarchical Bayesian framework and estimated simultaneously. A Bayesian test to detect deterministic components in the data is also constructed. By using an asymptotic approximation to the likelihood, the computation is carried out efficiently using the Markov chain Monte Carlo method in $O(Mn)$ operations, where n is the sample size and M is the number of iterations. We show empirically that our approach works well on real and simulated samples.

Keywords: FREQUENCY ESTIMATION; KALMAN FILTER; MARKOV CHAIN MONTE CARLO METHOD; METROPOLIS–HASTINGS ALGORITHM; MISSING DATA; MIXTURE OF NORMALS; REDUCED CONDITIONALS; SINUSOIDS; STATIONARY SERIES; TREND

1. INTRODUCTION

Consider data collected in time order that are modelled as the sum of a signal and noise. The signal is a deterministic regression function and usually consists of a polynomial trend and trigonometric terms with known or unknown frequencies. The noise is a zero-mean stationary Gaussian time series with a smooth spectral density. Such a model is said to have a mixed spectral density. There are extensive applications of such a model and a large statistical and engineering literature dealing with it. See, for example, Priestley (1981) and Marple (1987).

This paper provides a Bayesian analysis of such a model. The main contribution of our paper is that different features of the data — such as the spectral density of the stationary term, the regression parameters, unknown frequencies and missing observations — are combined in a hierarchical Bayesian framework and estimated simultaneously. A Bayesian test to detect deterministic components in the data is also constructed. Applications of our methods to simulated and real data suggest that they perform well.

We place a smoothness prior, similar to that in Wahba (1980), on the logarithm of the spectral density. To make the estimation of the spectral density computationally tractable, Whittle's (1957) approximation to the Gaussian likelihood is used. This results in a nonparametric regression problem with the logarithm of the periodogram as the dependent variable, the logarithm of the spectral density as the unknown regression

†*Address for correspondence*: Australian Graduate School of Management, University of New South Wales, Sydney 2052, Australia.
E-mail: robertk@agsm.unsw.edu.au

 0035–9246/97/59255

curve and observation errors having log-χ^2-distributions. By approximating the logarithm of a χ^2-distribution as a mixture of normals, the approximate log-likelihood together with the prior for the spectral density can be expressed as a state space model with errors that are mixtures of normals. The computation is carried out efficiently by Markov chain Monte Carlo simulation using the sampling approach in Carter and Kohn (1994).

To make the paper easier to read, the full model is introduced in several steps. Section 2 shows how to estimate the spectral density of a stationary process in the absence of deterministic components. Section 3 extends the estimation to the signal plus noise model with missing observations. Section 4 shows by example how the results in Sections 2 and 3 can be combined to analyse data and studies empirically the performance of the methods on simulated and real data sets. Section 5 reviews the literature and compares it with our approach.

2. ESTIMATING SPECTRAL DENSITY

Suppose that y_t, $t \geqslant 1$, is a zero-mean stationary Gaussian time series with spectral density $f(\omega)$, for $\omega \in [-\pi, \pi]$. Missing observations and deterministic components are introduced in later sections. Given the sample $y_1, \ldots, y_n$, the periodogram at the frequency ω is defined as

$$I(\omega) = \frac{1}{n}\left|\sum_{t=1}^{n} y_t \exp(-\mathrm{i}t\omega)\right|^2.$$

Let $\omega_j = 2\pi j/n$ be the Fourier frequencies, $Y = (y_1, \ldots, y_n)'$ and $\Gamma = \mathrm{var}(Y)$. Whittle (1957) showed that for n large

$$\begin{aligned} p(Y|f) &\propto |\Gamma|^{-1/2} \exp(-Y'\Gamma^{-1}Y/2) \\ &\approx \prod_{j=1}^{n} f(\omega_j)^{-1/2} \exp\left\{-\sum_{j=1}^{n} I(\omega_j)/2f(\omega_j)\right\}. \end{aligned} \tag{2.1}$$

We use approximation (2.1) for the likelihood of Y as it is computationally intractable to generate f by using the exact likelihood.

Let $\lfloor n/2 \rfloor$ be the largest integer less than or equal to $n/2$. It follows from approximation (2.1) that for large samples and for $j = 0, \ldots, \lfloor n/2 \rfloor$ the $I(\omega_j)$ are independent with $I(\omega_j) \sim f(\omega_j)\chi^2_2/2$ for $\omega_j \neq 0, \pi$, and $I(\omega_j) \sim f(\omega_j)\chi^2_1$ for $\omega_j = 0, \pi$. This can also be shown directly as in Priestley (1981), p. 425. An equivalent way of expressing this is that for large samples

$$\log I(\omega_j) = \log f(\omega_j) + e_j, \tag{2.2}$$

where the e_js are independent and $e_j \sim \log(\chi^2_2/2)$, $\omega_j \neq 0, \pi$, and $e_j \sim \log \chi^2_1$, $\omega_j = 0, \pi$. Although approximation (2.1) involves n Fourier frequencies, there are only $\lfloor n/2 \rfloor + 1$ distinct observations in equation (2.2) because $f(\omega) = f(2\pi - \omega)$ and $I(\omega) = I(2\pi - \omega)$. Equation (2.2) shows that estimating the spectral density can be viewed as a nonparametric regression problem with the observation being the log-periodogram ordinates, the unknown regression function being the logarithm of the spectral density and the errors having log-χ^2-distributions.

Hannan (1973) showed that equation (2.2) holds under very general conditions. His results suggest that the methods outlined in this section and Section 3 may perform well for non-Gaussian distributions, except, possibly, when there are missing data.

As in Wahba (1980), we place the following smoothness prior on the logarithm of the spectral density:

$$\frac{\mathrm{d}^2\{\log f(\omega)\}}{\mathrm{d}\omega^2} = \tau \frac{\mathrm{d}W(\omega)}{\mathrm{d}\omega}; \tag{2.3}$$

$W(\omega)$ is a Wiener process, $W(0) = 0$, $\mathrm{var}\{W(\omega)\} = \omega$ and τ is a scale parameter. The symmetry and periodicity of the spectral density mean that

$$\frac{\mathrm{d}\{\log f(\omega)\}}{\mathrm{d}\omega} = 0, \qquad \omega = 0, \pi. \tag{2.4}$$

To indicate our lack of knowledge about $\log f(0)$ we make it diffuse *a priori*, i.e.

$$\log f(0) \sim N(0, k) \tag{2.5}$$

with $k \to \infty$.

To complete the Bayesian model for the spectral density, we give the scale factor τ the non-informative prior $p(\tau^2) \propto 1/\tau^2 \exp(-\rho/\tau^2)$, where ρ is a small positive number; typically $\rho = 10^{-10}$.

The computations that are necessary for the Bayesian analysis are carried out by expressing the observation equation (2.2) and the smoothness prior (2.3)–(2.5) in state space form and by approximating the log-χ^2-distributions of the errors by mixtures of normals. For the rest of this section we assume, for simplicity, that n is even.

Let

$$x_j = (\log f(\omega), \mathrm{d}\{\log f(\omega)\}/\mathrm{d}\omega)'_{\omega=\omega_j}$$

be the state vectors for $j = 0, \ldots, n/2$. The observation equation (2.2) becomes

$$\log I(\omega_j) = h' x_j + e_j,$$

where $h = (1 \quad 0)'$. The state transition equation is

$$x_j = T x_{j-1} + \tau u_j \qquad j \geqslant 2.$$

Let $\delta = 2\pi/n$. The matrix T is 2×2 with $T_{11} = 1 = T_{22}$, $T_{12} = \delta$ and $T_{21} = 0$. The disturbances u_j are independent $N(0, U)$, with $U_{11} = \delta^3/3$, $U_{12} = U_{21} = \delta^2/2$ and $U_{22} = \delta$. Section 3 of Kohn and Ansley (1987) gives further details. The constraint that $\mathrm{d}\{\log f(\omega)\}/\mathrm{d}\omega = 0$ for $\omega = \pi$ is imposed by adding the pseudo-observation $(0\ 1)x_{n/2} = 0$ to the state space model, as in Ansley *et al.* (1993).

We approximate the distributions of $\log \chi_1^2$ and $\log(\chi_2^2/2)$ by five-component mixtures of normals with the weights, means and variances given in Table 1 (see also Fig. 1). Shephard (1994) also suggested a mixture of normals approximation to a log-χ^2-distribution when estimating stochastic volatility.

Let K_j determine which component of the mixture e_j belongs to and let $K = (K_0, \ldots, K_{n/2})'$. The prior probability of K_j is given by the first and fourth columns of Table 1. We assume that the K_j are independent *a priori* since the e_j are independent.

Let $g_j = h' x_j = \log f(\omega_j), j = 0, \ldots, n/2$, and let $G = (g_0, \ldots, g_{n/2})'$. We estimate G

TABLE 1
Five-component approximations to $\log \chi_1^2$ *and* $\log(\chi_2^2/2)$

$\log \chi_1^2$			$\log(\chi_2^2/2)$		
Probability	*Mean*	*Variance*	*Probability*	*Mean*	*Variance*
0.13	−4.63	8.75	0.19	−2.20	1.93
0.16	−2.87	1.95	0.11	−0.80	1.01
0.23	−1.44	0.88	0.27	−0.55	0.69
0.22	−0.33	0.45	0.25	−0.035	0.60
0.25	0.76	0.41	0.18	0.48	0.29

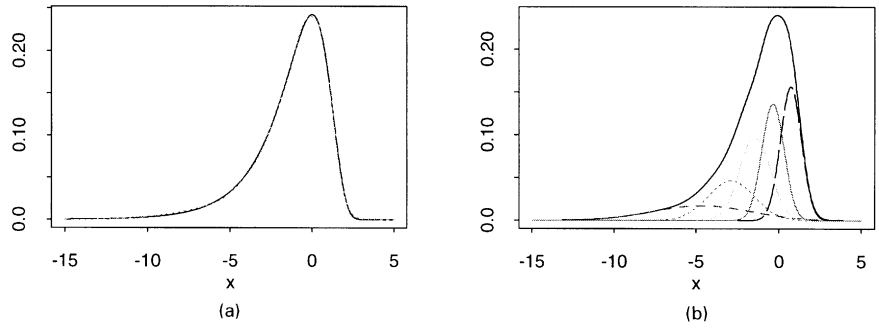

Fig. 1. (a) Density of a $\log(\chi_1^2)$ variable (———) and the normal mixture approximation (·····); (b) approximate density (———) and also the densities of the normal components (·····, - - - -, – – –, — —)

by an estimate of its posterior mean by using Markov chain Monte Carlo simulation. The Markov chain Monte Carlo sampler, outlined in sampling scheme 1 below, starts with some initial value $K = K^{[0]}$ and successively generates τ^2, G and K from conditional distributions. The iterates converge to realizations from the joint posterior distribution $p(\tau^2, G, K|Y)$. The sampler is usually run for a warm-up period at the end of which the sampler is assumed to have converged, followed by a sampling period which is used to estimate posterior moments.

Instead of generating τ^2 directly as in Carter and Kohn (1994), we generate $\gamma = \log(\tau^2)$ by using a Metropolis–Hastings step (Tierney, 1994). The reason for doing so is explained in Section 4.3. The conditional density $p(\gamma|Y, K)$ is approximated by a normal density with a mean chosen to match the mode of $p(\gamma|Y, K)$ and a variance chosen to match the second derivate of $\log p(\gamma|Y, K)$ at the mode. We denote this Metropolis–Hastings proposal by $q(\tau^2|Y, K)$. The vector G is generated from the conditional density $p(G|Y, K, \tau^2)$ as in de Jong and Shephard (1995) and the vector K is generated as in Carter and Kohn (1994).

Sampling scheme 1 is summarized as follows. Given the initial value $K = K^{[0]}$, generate realizations of τ^2, G and K by repeating the following steps.

(a) Generate τ^2 from $q(\tau^2|Y, K)$ and use a Metropolis–Hastings step to determine whether the generated τ^2 is accepted or the old value of τ^2 is retained.

(b) Generate G from $p(G|Y, K, \tau^2)$.
(c) Generate K from $p(K|Y, G, \tau^2)$.

We note that τ^2 and G are generated jointly in sampling scheme 1. A discussion of the convergence of this and other sampling schemes is given in Section 4.

Remark 1. We have found the Metropolis–Hastings proposal described above for generating τ^2 to be adequate for our examples. Its performance in a complex example is described in Section 4.3. However, it can be replaced by any other proposal, such as the proposal in Gilks *et al*. (1995).

3. ESTIMATING OTHER COMPONENTS OF MODEL

3.1. *Introduction*

Suppose that the observations y_t are generated by the model

$$y_t = z_t'\beta + v_t; \tag{3.1}$$

v_t is a zero-mean stationary Gaussian time series with smooth spectral density $f(\omega)$; $z_t'\beta$ is a sum of a polynomial trend or trigonometric terms or both. A Bayesian method for detecting polynomial trends and trigonometric components is given in Section 3.2. Section 3.3 generalizes sampling scheme 1 in Section 2 so that the regression coefficients can be generated as well. Section 3.4 discusses the case of a design matrix representing sinusoids at multiples of an unknown fundamental frequency and a Metropolis–Hastings proposal is given to generate the fundamental frequency. Section 3.5 allows for missing data.

3.2. *Detection of Spikes*

There are many references on using the periodogram or log-periodogram to detect a signal consisting of a polynomial trend and sinusoids. See, for example, Priestley (1981), chapter 8. The basis for these methods is that such a signal usually causes large positive spikes in the log-periodogram. For example, low order polynomials cause a spike around the zero frequency and sinusoidal signals usually cause a spike around their nearest Fourier freqencies, usually with some leakage to nearby frequencies.

We construct a Bayesian test for polynomial or trigonometric terms by using mixture of normals approximations to the log-χ^2-densities with one of the components having a large variance. Because of this component, the approximating density will be heavier tailed than the corresponding χ^2-density. We say that a spike occurs at the frequency ω_j if the indicator variable K_j has a high posterior probability of taking the value corresponding to the high variance component and the residual is positive. Box and Tiao (1968) used a similar approach to model outliers. Specifically, we use the approximation in Table 1 for $\log \chi_1^2$ as it has one component with a variance of 8.75, with the rest of the components having substantially smaller variances. We approximate the density of $\log(\frac{1}{2}\chi_2^2)$ by a mixture of normals having five components, with one of the components having a variance of 25, a mean of $\log(\frac{1}{2})$ and a prior probability of 0.02; the other four components are chosen so that the five-component mixture is a good approximation to the left-hand tail and centre of the $\log(\frac{1}{2}\chi_2^2)$-density. The approximating mixture is given in Table 2, with the quality of the approximation similar to that in Table 1, except that it has a heavier right-hand tail.

TABLE 2
Five-component approximation to $\log(\chi_2^2/2)$ with one component having a variance of 25 and prior probability 0.02

Probability	*Mean*	*Variance*
0.13	−2.26	3.31
0.35	−0.91	0.92
0.02	−0.69	25
0.20	−0.32	0.63
0.30	0.34	0.38

3.3. *Generating Regression Coefficients*

Suppose that the regression coefficient vector β in model (3.1) has a diffuse prior, with the priors on the spectral density f and τ^2 as in Section 2. The conditional density of β given Y and f is

$$p(\beta|Y,f) \propto p(Y|\beta,f)\,p(\beta) \propto p(Y|\beta,f). \tag{3.2}$$

Let $Z = (z_1, \ldots, z_n)'$ and $W(\beta) = Y - Z\beta$. From approximation (2.1),

$$p(Y|\beta,f) \propto \exp\left\{-\sum_{j=1}^{n} I_{W(\beta)}(\omega_j)/2f(\omega_j)\right\} \tag{3.3}$$

where

$$I_{W(\beta)}(\omega) = \frac{1}{n}\left|\sum_{t=1}^{n}(y_t - z_t'\beta)\exp(-\mathrm{i}t\omega)\right|^2 = |a(\omega) - b(\omega)'\beta|^2, \tag{3.4}$$

with

$$\begin{aligned} a(\omega) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} y_t \exp(-\mathrm{i}t\omega) \\ b(\omega) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t \exp(-\mathrm{i}t\omega). \end{aligned} \tag{3.5}$$

Let $\mathcal{R}\{a\}$ be the real part of a. From expressions (3.2)–(3.5), $p(\beta|Y,f) \sim N(\mu, \Sigma)$ where

$$\Sigma^{-1} = \mathcal{R}\left\{\sum_{j=1}^{n} b(\omega_j)\,b(\omega_j)^*/f(\omega_j)\right\}$$

and

$$\Sigma^{-1}\mu = \mathcal{R}\left\{\sum_{j=1}^{n} a(\omega_j)\,b(\omega_j)^*/f(\omega_j)\right\}'.$$

It is therefore straightforward to generate β.

The terms τ^2, G and K are generated as in sampling scheme 1 in Section 2 by also conditioning on β and replacing Y by $Y - Z\beta$.

3.4. *Generating Fundamental Frequency*

Suppose that the signal consists of a linear combination of sinusoids with frequencies 0, λ, . . ., $r\lambda$ and unknown phase. This can be written as model (3.1) with

$$z_t = \{1,\ \sin(\lambda t),\ \cos(\lambda t),\ \ldots,\ \sin(r\lambda t),\ \cos(r\lambda t)\}', \tag{3.6}$$

where the notation is simplified by omitting to indicate the dependence of z_t on λ. The regression parameters β can be generated from the distribution $p(\beta|Y, \lambda, f)$ as in Section 3.3.

To generate the fundamental frequency λ we use a Metropolis–Hastings proposal based on the conditional density $p(\lambda|Y, f)$ and the asymptotic results in Quinn and Thompson (1991). The unknown frequency λ is less than or equal to π/r because $r\lambda \leqslant \pi$. We place a uniform prior for λ in the interval $[0, \pi/r]$. Then,

$$p(\lambda|Y, f) \propto p(Y|\lambda, f) \propto |Z'\Gamma^{-1}Z|^{-1/2}\ \exp\{Y'\Gamma^{-1}Z(Z'\Gamma^{-1}Z)^{-1}Z'\Gamma^{-1}Y/2\}, \tag{3.7}$$

where $\Gamma = \mathrm{var}(v)$. By Quinn and Thompson (1991), for large n expression (3.7) is proportional to

$$\prod_{j=-r}^{r} f(j\lambda)^{1/2}\ \exp\left\{\sum_{j=-r}^{r} I(j\lambda)/2f(j\lambda)\right\} \propto \prod_{j=1}^{r} f(j\lambda)\ \exp\left\{\sum_{j=1}^{r} I(j\lambda)/f(j\lambda)\right\}. \tag{3.8}$$

Let $\hat{\lambda}$ maximize expression (3.8). Our Metropolis–Hastings proposal for generating from $p(\lambda|Y, f)$ is to generate from a constrained normal distribution with mean $\hat{\lambda}$, variance chosen to match the second derivative of expression (3.8) at $\lambda = \hat{\lambda}$ and λ constrained to the interval $[0, \pi/r]$. We denote this Metropolis–Hastings proposal by $q(\lambda|Y, f)$.

Remark 2. A more efficient way of generating β when z_t is of the form (3.6) is to use the asymptotic results discussed in Quinn and Thompson (1991).

3.5. *Generating Missing Data*

Suppose that some of y_1, . . ., y_n are missing. Let Y_{mis} denote the missing data and let $S_{\mathrm{mis}} = \{s_1, \ldots, s_l\}$ denote the indices of the missing data. Let Y_{obs} and R_{obs} denote the corresponding variables for the observed data. We derive the approximate distribution of $p(Y_{\mathrm{mis}}|Y_{\mathrm{obs}}, f)$ for the linear model (3.1). The missing observations can then be generated as part of the Markov chain Monte Carlo algorithm by augmenting to the steps already described in Sections 2 and 3.

Suppose that the observations are of the form (3.1). It is straightforward to show that

$$p(Y_{\mathrm{mis}}|Y_{\mathrm{obs}}, f) \propto p(Y|f) \propto \exp[-\{Y'\Gamma^{-1}Y - Y'\Gamma^{-1}Z(Z'\Gamma^{-1}Z)^{-1}Z'\Gamma^{-1}Y\}/2]. \tag{3.9}$$

Similarly to the derivation of equation (2.1) it is possible to show that for n large

$$p(Y_{\mathrm{mis}}|Y_{\mathrm{obs}}, f) \propto \exp[-\{\tilde{Y}^*F^{-1}\tilde{Y} - \tilde{Y}^*F^{-1}\tilde{Z}(\tilde{Z}^*F^{-1}\tilde{Z})^{-1}\tilde{Z}^*F^{-1}\tilde{Y}\}/2], \tag{3.10}$$

where $\tilde{Y}$ is the Fourier transform of Y, $\tilde{Z}$ is the Fourier transform of Z and F is the diagonal matrix having $f(\omega_1), \ldots, f(\omega_n)$ on the diagonal. The term

$$\tilde{Y}^* F^{-1} \tilde{Y} = \sum_{j=1}^{n} I(\omega_j)/f(\omega_j) \tag{3.11}$$

where

$$I(\omega_j) = |c(\omega_j) + d(\omega_j) Y_{\text{mis}}|^2, \tag{3.12}$$

$$\begin{aligned} c(\omega) &= \frac{1}{\sqrt{n}} \sum_{t \in R_{\text{obs}}} y_t \exp(-\mathrm{i}t\omega), \\ d(\omega) &= \frac{1}{\sqrt{n}} \{\exp(-\mathrm{i}s_1\omega), \ldots, \exp(-\mathrm{i}s_l\omega)\}'. \end{aligned} \tag{3.13}$$

Similarly,

$$\tilde{Z}^* F^{-1} \tilde{Y} = \sum_{j=1}^{n} I_{Z,Y}(\omega_j)/f(\omega_j), \tag{3.14}$$

where

$$\begin{aligned} I_{Z,Y}(\omega) &= \frac{1}{n} \left\{ \sum_{t=1}^{n} z_t \exp(\mathrm{i}t\omega) \right\} \sum_{t=1}^{n} y_t \exp(-\mathrm{i}t\omega) \\ &= \bar{b}(\omega)\{c(\omega) + d(\omega)' Y_{\text{mis}}\} \end{aligned} \tag{3.15}$$

and $b(\omega)$ is defined in equations (3.5). From expressions (3.10)–(3.15), and omitting terms that do not depend on Y_{mis},

$$-2 \log p(Y_{\text{mis}} | Y_{\text{obs}}, f) \approx Y'_{\text{mis}} \Omega_1^{-1} Y_{\text{mis}} - 2\nu_1' \Omega_1^{-1} Y_{\text{mis}}, \tag{3.16}$$

where

$$\Omega_1^{-1} = \mathcal{R}\left[\sum_{j=1}^{n} d(\omega_j)\{1 - b(\omega_j)^*(\tilde{Z}^* F^{-1} \tilde{Z})^{-1} b(\omega_j)/f(\omega_j)\}\, d(\omega_j)^*/f(\omega_j) \right]$$

and

$$\Omega_1^{-1}\nu_1 = \mathcal{R}\left[\sum_{j=1}^{n} c(\omega_j)\{1 - b(\omega_j)^*(\tilde{Z}^* F^{-1} \tilde{Z})^{-1} b(\omega_j)/f(\omega_j)\}\, d(\omega_j)^*/f(\omega_j) \right]'.$$

From expression (3.16), $p(Y_{\text{mis}} | Y_{\text{obs}}, f) \sim N(\nu_1, \Omega_1)$ and it is straightforward to generate Y_{mis} given f and Y_{obs}.

Remark 3. A more efficient way of generating Y_{mis} when z_t is of the form (3.6) is to use asymptotic approximations similar to those in Quinn and Thompson (1991). Details are given in Carter and Kohn (1995), which is available from the authors.

Remark 4. The asymptotic approximations given above are inaccurate for missing data near the beginning or the end of the time series. Thus the approach given above is not suitable for forecasting.

4. EXAMPLES

4.1. *Introduction*

We give two examples to illustrate the results in Sections 2 and 3. The first uses simulated data with a linear mean function and autoregressive errors. The second uses astronomical data on the variable star *S. Carinae*. These data have a periodic mean function, which is modelled as in Section 3.3, and missing data, which are generated as part of the Markov chain Monte Carlo algorithm as in Section 3.5.

4.2. *Simulated Data*

The data are generated from the model $y_t = \beta_0 + \beta_1 t + \upsilon_t$, for $t = 0, \ldots, 255$, $\beta_0 = 1$ and $\beta_1 = 6/255$. The errors υ_t are generated from the third-order autoregression

$$\upsilon_t = 1.4256\upsilon_{t-1} - 0.7344\upsilon_{t-2} + 0.1296\upsilon_{t-3} + \eta_t,$$

where $\eta_t \sim N(0, 1)$. This model was used by Wahba (1980) in her simulations.

From sampling scheme 1 in Section 2 and Section 3.3, the unknown parameters in this model are generated in sampling scheme 2 as follows.

Using some starting values for β and K, the following three steps are repeated: generate from

(a) $q(\tau^2 | Y, \beta, K)$ and use a Metropolis–Hastings step,
(b) $p(G | Y, \beta, K, \tau^2)$ and
(c) $p(\beta, K | Y, G, \tau^2)$.

The proof that this sampling scheme converges to the posterior distribution $p(G, \beta, K, \tau^2 | Y)$ is similar to, but simpler than, the proof given below for sampling scheme 4 in Section 4.3.

Fig. 2 shows the output from one simulation using a warm-up period of 1000 iterations and a sampling period of 1000 iterations. The simulation takes 2 min on a Sun Sparcstation 20. The starting values are $\beta = (0 \quad 0)'$ and $K = (1, \ldots, 1)'$. Fig. 2(a) shows the data y_t, the mean function $z_t'\beta$ and the Markov chain Monte Carlo estimate $z_t'\hat{\beta}$. Fig. 2(b) shows the log-periodogram of the detrended data, the log-spectral-density of the error sequence and the Markov chain Monte Carlo estimate. Fig. 2(c) is a histogram of the generated values of β_1 from the sampling period showing that its posterior distribution is approximately normally distributed about the true value. Results similar to those in Fig. 2 are obtained for other starting values, suggesting that the Markov chain has converged to the posterior distribution and is not stuck in a local mode.

4.3. *Variable Star Data*

The star example uses visual observations on the variable star *S. Carinae* collected over 1189 10-day periods. Each observation is the average over the preceding 10 days and there are 40 missing observations. Further details are in Quinn and Thompson (1991) and the references therein.

To detect deterministic components we use the approach described in Section 3.2 and model the observation as $y_t = \beta_0 + \upsilon_t$ and

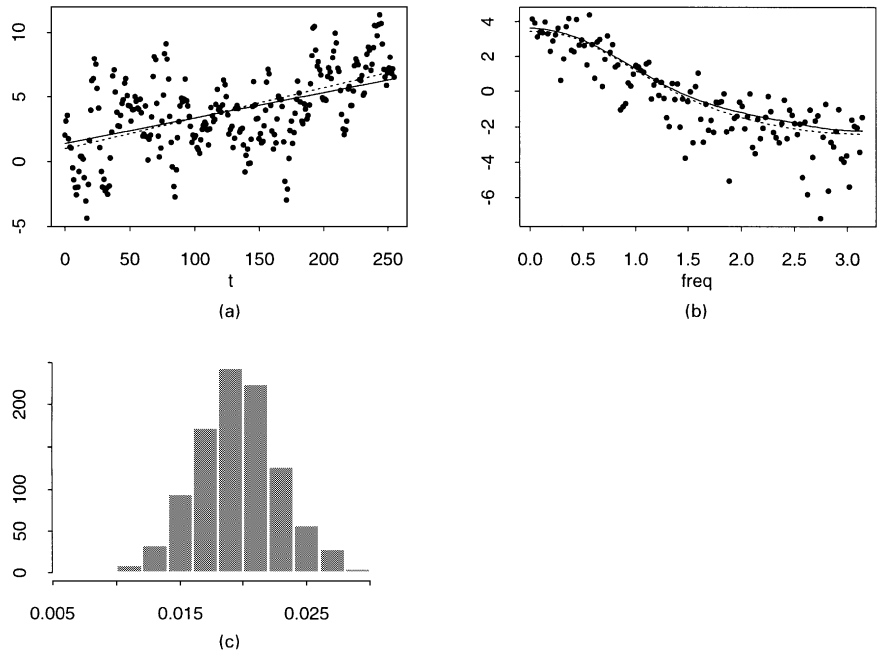

Fig. 2. (a) Data (•), mean function $z_t'\beta$ (- - - - -) and Markov chain Monte Carlo estimate $z_t'\hat{\beta}$ (———); (b) log-periodogram of the detrended data (•), log-spectral density of the error sequence (- - - - -) and Markov chain Monte Carlo estimate (———); (c) histogram of the generated values of β_1 from the sampling period

$$\log I_v(\omega_j) = \log f(\omega_j) + e_j, \tag{4.1}$$

where $I_v(\omega)$ is the periodogram of v_t. The errors e_j are modelled as a mixture of normals as described in Section 3.2. The following sample scheme is run to detect spikes in the spectral density. It is similar to sampling scheme 2 in Section 4.2 except that missing observations are generated as in Section 3.5. The proof that the following sampling scheme 3 converges to the correct posterior distribution follows from the proof for sampling scheme 4 given below.

Generate from

(a) $q(\tau^2|Y, \beta, K)$ and use a Metropolis–Hastings step,
(b) $p(G|Y, \beta, K, \tau^2)$,
(c) $p(Y_{\text{mis}}|Y_{\text{obs}}, G, \tau^2)$ and
(d) $p(\beta, K|Y, G, \tau^2)$.

Fig. 3 shows the output of one simulation using the initial values $\beta_0 = 0$, $Y_{\text{mis}} = (0, 0, \ldots, 0)'$ and $K = (1, 1, \ldots, 1)'$. Fig. 3(a) shows the data; Fig. 3(b) shows the log-periodogram—with the missing data replaced by the Markov chain Monte Carlo estimates; Fig. 3(c) shows the Markov chain Monte Carlo estimate of the log-spectral-density of the error sequence; Fig. 3(d) shows the probabilities of a spike in the log-periodogram, which indicate a deterministic component as discussed in Section 3.2. From Fig. 3(d) there are three main periodic components in the data

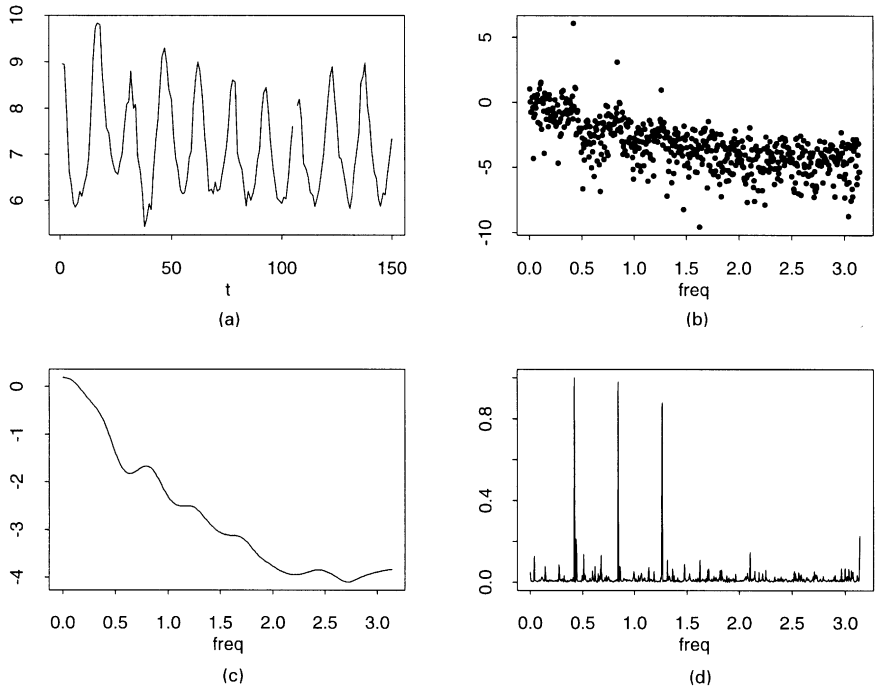

Fig. 3. (a) First 150 observations; (b) log-periodogram; (c) Markov chain Monte Carlo estimate of the log-spectral-density of the error sequence; (d) probabilities of a spike in the log-periodogram

whose frequencies are multiples of a fundamental frequency of about 0.42. Accordingly, we fit model (3.6) with $m = 3$.

The results of Sections 2, 3.3 and 3.5 are used to generate the spectral density and the parameters in model (3.6) in the following sampling scheme 4.

Generate from

(a) $q(\tau^2|Y, \lambda, \beta, K)$ and use a Metropolis–Hastings step,
(b) $p(G|Y, \lambda, \beta, K, \tau^2)$,
(c) $p(Y_{\text{mis}}|Y_{\text{obs}}, G, \lambda, \tau^2)$,
(d) $q(\lambda|Y, G, \tau^2) = q(\lambda|Y, G)$ and use a Metropolis–Hastings step and
(e) $p(\beta, K|Y, G, \lambda, \tau^2)$.

From Tierney (1994) this sampling scheme converges to the posterior distribution $p(G, \lambda, \beta, K, \tau^2, Y_{\text{mis}}|Y_{\text{obs}})$ because it is invariant to $p(G, \lambda, \beta, K, \tau^2, Y_{\text{mis}}|Y_{\text{obs}})$, it is aperiodic and irreducible. Invariance follows from the construction of Gibbs and Metropolis–Hastings sampling schemes and the invariance result in appendix 1 in Carter and Kohn (1996). Aperiodicity and irreducibility follow because all the conditional distributions have strictly positive densities.

Fig. 4 shows the output from one simulation. The starting values are $\lambda = 0.425$, $\beta = (0, \ldots, 0)'$, $Y_{\text{mis}} = (0, \ldots, 0)'$ and $K = (1, \ldots, 1)'$. Other details are similar to those in Section 4.2. The simulation takes 30 min on a Sun Sparcstation 20. Fig. 4(a) shows the log-periodogram of the detrended data—with the missing data replaced

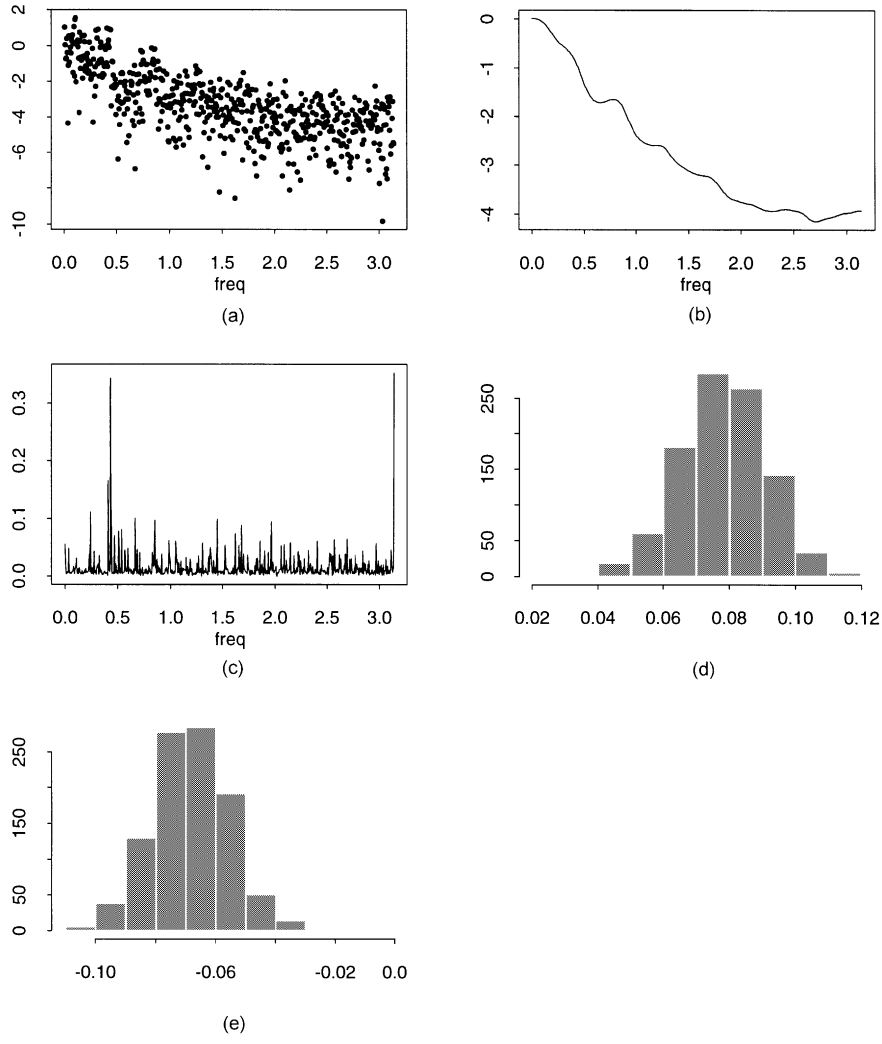

Fig. 4. (a) Log-periodogram of the detrended data; (b) Markov chain Monte Carlo estimates of the log-spectral-density; (c) probabilities of a spike in the log-periodogram; (d), (e) histograms of the generated values of β_6 and β_7 respectively

by the Markov chain Monte Carlo estimates. The spikes in the log-periodogram in Fig. 3(b) do not appear in Fig. 4(a). Fig. 4(b) shows the Markov chain Monte Carlo estimates of the log-spectral-density. Fig. 4(c) shows the probabilities of a spike in the log-periodogram, which are much smaller than the probabilities shown in Fig. 3(d). Figs 4(d) and 4(e) show histograms of the generated values of β_6 and β_7 from the sampling period and suggest that both parameters are significantly different from 0.

Let sampling scheme 5 denote the sampling scheme that generates τ^2 from the distribution $p(\tau^2|Y, G, \lambda, \beta, K) = p(\tau^2|G)$ and is otherwise the same as sampling scheme 4. This is the way that τ^2 is generated by Carter and Kohn (1994). Although sampling scheme 5 is easier to implement than sampling scheme 4 because $p(\tau^2|G)$ is

inverse gamma, we found that, when unknown frequencies need to be estimated as well, sampling scheme 5 can converge unacceptably slowly and it is necessary to use sampling scheme 4. We note that the acceptance rate for the Metropolis–Hastings proposal for generating τ^2 in sampling scheme 4 is 72%.

5. DISCUSSION OF RELATED WORK

Wahba (1980) estimated the logarithm of the spectral density by a smoothing spline, using equation (2.2) as the observation equation with e_j approximated by a normal distribution having the correct mean and variance. She estimated the smoothing parameter τ^2 by an approximation to generalized cross-validation. Pawitan and O'Sullivan (1994) improved Wahba's (1980) approach by taking into account that the errors e_j in equation (2.2) have a log-χ^2-distribution and obtained the minimizer to an approximation to the penalized likelihood. Their estimate of the smoothing parameter is a two-step approximation to an estimate of the integrated squared error. Neither Wahba nor Pawitan and O'Sullivan considered the joint estimation of the spectral density and the detection and estimation of sinusoids. Further, it is unclear how to extend their methods to detect and estimate unknown frequencies.

An alternative approach for estimating the spectral density is to fit an autoregressive process to the data, with the order selected by some order selection criterion. This approach is discussed by Marple (1987), chapter 8, who noted that to obtain high resolution from the autoregressive spectral density estimate the order of the autoregression needs to be high. However, a high order often leads to spurious features in the spectral density estimate.

The detection of polynomial and trigonometric components in a time series is of interest in its own right and also because such components can severely affect the estimate of the spectral density of the non-deterministic component. Recently Kavalieris and Hannan (1994) (see also the references in this paper) determined the number of sinusoidal components r by approximating the noise spectral density by an autoregression of order h, and using penalized likelihood to choose h and r. For a given value of r, the unknown sinusoidal frequencies are estimated by using a procedure that considers the largest values of the periodograms of the data evaluated at the Fourier frequencies and does not require knowledge of the noise spectral density. However, when applying their method to the star data, discussed in Section 4.3, Kavalieris and Hannan pointed out that when the noise spectrum is not flat it may be difficult to determine spikes in the spectrum by an examination of the periodogram alone. They modified their procedure for estimating unknown frequencies by looking for large values of the periodogram of the data divided by a subjective estimate of the noise spectral density. It seems, therefore, that the practical implementation of the Kavalieris and Hannan approach requires a subjective estimate of the noise spectral density.

Given the number of harmonics of a fundamental frequency, Quinn and Thompson (1991) estimated the unknown frequencies by maximizing expression (3.8). Because the spectral density $f(\omega)$ is unknown, they estimated it by using the median of the periodogram values at the m Fourier frequencies nearest ω, with m chosen subjectively.

There seems to be little work on how to deal with missing observations when estimating the two-component model.

We conclude that current work on estimating the signal plus noise model seems incomplete as the bandwidth of the estimate of the spectral density is often chosen subjectively from the data. The approach in this paper estimates all unknowns simultaneously and so seems to solve the problems encountered in previous work.

ACKNOWLEDGEMENTS

This work was partially supported by the Australian Research Council. We would like to thank Barry Quinn for the variable star data and for helpful discussions.

REFERENCES

Ansley, C. F., Kohn, R. and Wong, C. (1993) Nonparametric spline regression with prior information. *Biometrika*, **80**, 75–88.

Box, G. E. P. and Tiao, G. C. (1968) A Bayesian approach to some outlier problems. *Biometrika*, **55**, 119–129.

Carter, C. K. and Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika*, **81**, 541–553.

———(1995) Semiparametric Bayesian inference for time series with mixed spectra. *Working Paper 95-12*. Australian Graduate School of Management, University of New South Wales, Sydney.

———(1996) Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika*, **83**, 589–601.

Gilks, W. R., Best, N. G. and Tan, K. K. C. (1995) Adaptive rejection Metropolis sampling within Gibbs sampling. *Appl. Statist.*, **44**, 455–472.

Hannan, E. J. (1973) The asymptotic theory of linear time series models. *J. Appl. Probab.*, **10**, 130–145.

de Jong, P. and Shephard, N. (1995) Efficient sampling from the smoothing density in time series models. *Biometrika*, **82**, 339–350.

Kavalieris, L. and Hannan, E. J. (1994) Determining the number of terms in a trigonometric regression. *J. Time Ser. Anal.*, **15**, 613–626.

Kohn, R. and Ansley, C. F. (1987) A new algorithm for spline smoothing based on smoothing a stochastic process. *SIAM J. Scient. Statist. Comput.*, **8**, 33–48.

Marple, S. L. (1987) *Digital Spectral Analysis with Applications*. Englewood Cliffs: Prentice Hall.

Pawitan, Y. and O'Sullivan, F. (1994) Nonparametric spectral density estimation using penalized Whittle likelihood. *J. Am. Statist. Ass.*, **89**, 600–610.

Priestley, M. B. (1981) *Spectral Analysis and Time Series*. London: Academic Press.

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

Shephard, N. (1994) Partial non-Gaussian state space models. *Biometrika*, **81**, 115–132.

Tierney, L. (1994) Markov chains for exploring posterior distributions. *Ann. Statist.*, **22**, 1701–1762.

Wahba, G. (1980) Automatic smoothing of the log periodogram. *J. Am. Statist. Ass.*, **75**, 122–132.

Whittle, P. (1957) Curve and periodogram smoothing. *J. R. Statist. Soc.* B, **19**, 38–47.