Bayesian Estimation of the Spectral Density of a Time Series

Nidhan CHOUDHURI, Subhashis GHOSAL, and Anindya ROY

This article describes a Bayesian approach to estimating the spectral density of a stationary time series. A nonparametric prior on the spectral density is described through Bernstein polynomials. Because the actual likelihood is very complicated, a pseudoposterior distribution is obtained by updating the prior using the Whittle likelihood. A Markov chain Monte Carlo algorithm for sampling from this posterior distribution is described that is used for computing the posterior mean, variance, and other statistics. A consistency result is established for this pseudoposterior distribution that holds for a short-memory Gaussian time series and under some conditions on the prior. To prove this asymptotic result, a general consistency theorem of Schwartz is extended for a triangular array of independent, nonidentically distributed observations. This extension is also of independent interest. A simulation study is conducted to compare the proposed method with some existing methods. The method is illustrated with the well-studied sunspot dataset.

KEY WORDS: Bernstein polynomial; Dirichlet process; Metropolis algorithm; Periodogram; Posterior consistency; Posterior distribution; Spectral density; Time series.

1. INTRODUCTION

Spectral analysis is a useful tool in the study of a stationary time series. Let $\{X_t; t = 0, \pm1, \ldots\}$ be a covariance stationary time series with autocovariance function $\gamma(r) = E(X_t X_{t+r})$. The second-order properties of the time series, under mild conditions that primarily exclude purely periodic components, are completely described by the spectral density function

$$f^*(\omega^*) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \gamma(r)e^{-ir\omega^*}, \qquad -\pi < \omega^* \leq \pi. \quad (1)$$

Parametric estimation of spectral density is mostly based on autoregressive moving average (ARMA) models with data-dependent order selection. However, these methods tend to produce biased results when the ARMA approximation to the underlying time series is poor.

The nonparametric estimation procedures mostly use the idea of smoothing the periodogram

$$I_n(\omega^*) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} X_t e^{-it\omega^*}\right|^2. \quad (2)$$

For a fixed $\omega^* \in (0, \pi)$, $I_n(\omega^*)$ has an asymptotic exponential distribution with mean $f^*(\omega^*)$. Moreover, the covariance between the periodogram ordinates evaluated at two different frequencies that are at least $2\pi/n$ apart is of the order n^{-1}. Thus the periodogram, as a process, is randomly fluctuating around the true spectral density. This is the motivation for smoothing the periodogram. Because exponential distributions form a scale family, smoothing is typically applied to the log-periodogram to obtain an estimate of the log-spectral density. Early smoothing techniques were based on kernel smoothers. Smoothing splines for this purpose were suggested by Cogburn and Davis (1974). Wahba (1980) proposed a regression based

method in which the estimator was obtained by spline smoothing the log-periodogram with a data-driven smoothing parameter.

The least squares-type smoothing techniques do not use the additional information of the limiting independent exponential distribution. Let $\omega_l^* = 2\pi l/n$ and $U_l = I_n(\omega_l^*)$, $l = 1, \ldots, \nu$, where $\nu = \nu_n = \lfloor(n-1)/2\rfloor$ is the greatest integer less than or equal to $(n-1)/2$. Then the joint distribution of $(U_1, \ldots, U_\nu)$ may be approximated by the joint distribution of ν independent exponential random variables with mean $f^*(\omega_l^*)$ for the lth component. Whittle (1957, 1962) proposed a "quasi-likelihood,"

$$L_n(f^*|X_1, \ldots, X_n) = \prod_{l=1}^{\nu} \frac{1}{f^*(\omega_l^*)}e^{-U_l/f^*(\omega_l^*)}, \quad (3)$$

known as the Whittle likelihood in the literature. One advantage of the Whittle likelihood is that it directly involves f, whereas the true likelihood involves f indirectly through the autocovariances. Chow and Grenander (1985) used a sieve method to obtain a nonparametric maximum likelihood estimator (MLE) based on the Whittle likelihood. Pawitan and O'Sullivan (1994) defined a penalized MLE as the maximizer of the Whittle likelihood with a roughness penalty. A polynomial spline-fitting approach was described by Kooperberg, Stone, and Truong (1995), and a local likelihood technique was used by Fan and Kreutzberger (1998). Ombao, Raz, Strawderman, and von Sachs (2001) described a Whittle likelihood-based generalized cross-validation method for selecting the bandwidth in periodogram smoothing.

Nonparametric Bayesian approaches to the estimation of a spectral density were studied by Carter and Kohn (1997), Gangopadhyay, Mallick, and Denison (1998), and Liseo, Marinucci, and Petrella (2001), who used the Whittle likelihood to obtain a pseudoposterior distribution of f^*. Carter and Kohn (1997) induced a prior on the logarithm of f^* through an integrated Wiener process and provided elegant computational methods for computing the posterior. Liseo et al. (2001) considered a Brownian motion for the spectral density and modified

it around the zero frequency to incorporate the long-range dependence. Gangopadhyay et al. (1998) fitted piecewise polynomials of a fixed low order to the logarithm of f^* while putting priors on the number of knots, the location of the knots, and the coefficients of the polynomials. Although these methods were applied to some real and simulated data, their theoretical properties are unknown. Only an asymptotic result was provided by Liseo et al. (2001) under a parametric assumption that the time series is a fractional Gaussian noise process.

The goals of the present article are to develop a nonparametric Bayesian method for the estimation of the spectral density and to prove consistency of the posterior distribution. The essential idea behind the construction of the prior probability is that the normalized spectral density is a probability density on a compact interval. A nonparametric prior is thus constructed using Bernstein polynomials in the spirit of Petrone (1999a,b). A pseudoposterior is obtained by updating the prior using the Whittle likelihood. A Markov chain Monte Carlo (MCMC) algorithm for sampling from this pseudoposterior is then described using the Sethuraman (1994) representation of a Dirichlet process. First, a general result is obtained that gives sufficient conditions for posterior consistency for a triangular array of independent, nonidentically distributed observations. Consistency of the pseudoposterior distribution of the spectral density is then established by verifying the conditions of this result and invoking a contiguity argument. Such use of a contiguity argument in proving posterior consistency result is new and may also be applied for different problems.

The article is organized as follows. The prior, together with the preliminaries on the Bernstein polynomial priors, are introduced in Section 2. An MCMC method for sampling from the posterior distribution is described in Section 3. Results of a simulation study and the analysis of the sunspot data is presented in Section 4. In Section 5 our method is justified through consistency of this pseudoposterior distribution. The general consistency result is presented and proved in the Appendix A, and the proof of consistency of the spectral estimates is presented in Appendix B.

2. BERNSTEIN POLYNOMIAL PRIOR

2.1 Preliminaries: Bernstein Polynomial Prior in Density Estimation

The Bernstein polynomial prior was developed by Petrone (1999a,b) to describe a nonparametric prior for probability densities on the unit interval. It is a mixture prior of beta kernels where the choice of the kernel is motivated by the fact that the mixtures of beta densities with only integer parameters can approximate any continuous density on [0, 1]. For any distribution function G with a continuous density q on [0, 1],

$$b(\omega; k, G) = \sum_{j=1}^{k} G\left(\frac{j-1}{k}, \frac{j}{k}\right] \beta(\omega; j, k-j+1) \quad (4)$$

converges uniformly to $q(\omega)$, where $G(u, v] = G(v) - G(u)$ and $\beta(\omega; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\omega^{a-1}(1-\omega)^{b-1}$ is the beta density with parameters a and b.

Motivated by the uniform approximation property, Vitale (1975) suggested a density estimator based on the Bernstein

kernel. In the Bayesian context, Petrone (1999a,b) defined the Bernstein polynomial prior as the distribution of the random density

$$q(\omega) = \sum_{j=1}^{k} w_{j,k}\beta(\omega; j, k-j+1), \quad (5)$$

where k has a probability mass function $\rho(\cdot)$ and, given k, $\mathbf{w}_k = (w_{1,k}, \ldots, w_{k,k})$ has a probability distribution on the k-dimensional simplex

$$\Delta_k = \left\{(w_1, \ldots, w_k) : w_j > 0, \sum_{j=1}^{k} w_j = 1\right\}. \quad (6)$$

In practice, the $w_{j,k}$'s are induced by a probability distribution function G as in (4), and a prior is put on G. Such a construction of the $w_{j,k}$'s provides a simple MCMC algorithm for sampling from the joint posterior distribution of (k, G) via data augmentation, whereas sampling directly from the joint distribution of $(k, \mathbf{w})$ is complicated as the dimension of $\mathbf{w}$ changes with k.

The right side of (5) is referred to as the Bernstein density of order k with weights $\mathbf{w}_k$. Let $\mathcal{B}_k$ denote the class of all Bernstein densities of order k. Then the $\mathcal{B}_k$'s may be viewed as sieves, and the prior may be thought of as a sieve prior in the sense of Ghosal, Ghosh, and Ramamoorthi (1997). Petrone and Veronese (2002) showed that the method of Bernstein kernel approximation may be motivated as a special case of a more general approximation technique due to Feller (1971, chap. VII). For the weak topology, the prior has full topological support if for all k, $\rho(k) > 0$ and $\mathbf{w}_k$ has full support on Δ_k with respect to the Euclidean distance (Petrone 1999a). In fact, any continuous distribution function is in the Kolmogorov–Smirnov support of the prior. If the distribution has a continuous and bounded density, then the conclusion can be strengthened to the variation metric, and the Bernstein polynomial prior gives consistent posterior in probability density estimation (Petrone and Wasserman 2002). The proof uses the techniques developed by Schwartz (1965), Barron, Schervish, and Wasserman (1999), and Ghosal et al. (1999).

2.2 Description of the Prior for the Spectral Density

The principal difference in the case of a probability density is that f^* does not integrate to 1. In addition, the domain of f^* is different form that of the beta densities. Because f^* is symmetric, we may reparameterize f^* by a function f on [0, 1] such that $f(\omega) = f^*(\pi\omega)$, $0 \le \omega \le 1$. Normalizing f to $q = f/\tau$ with the normalizing constant $\tau = \int f$, a prior on f may be induced by first putting a Bernstein polynomial prior on q and then putting an independent prior on τ. Thus the prior on f is described by the following hierarchical scheme:

- $f(\omega) = \tau \sum_{j=1}^{k} G((j-1)/k, j/k]\beta(\omega; j, k-j+1)$.
- G has a Dirichlet process distribution with base measure, $\alpha = MG_0$, where $M > 0$ is a constant and G_0 is a probability distribution function with Lebesgue density g_0.
- k has probability mass function, $\rho(k) > 0$, for $k = 1, 2, \ldots$.
- The distribution of τ has Lebesgue density π on $(0, \infty)$.
- G, k, and τ are a priori independent.

The variance and the frequency components are two different aspects of a stationary time series that are often a priori unrelated. Thus the decomposition of f into τ and q allows one to individually model the prior beliefs for these two components. In the absence of any concrete prior knowledge about one component, a default prior may be used for that part while still using an informative prior for the other component. A complete default prior may be used if no prior information is available.

Any knowledge about τ may be incorporated by centering the density π about the prior guess while adjusting the variance to the precision of this guess. A large variance indicates a lack of confidence in the prior guess. Inverse-gamma distributions form a two-parameter family that is conjugate, with the parameters flexible enough to adjust the prior mean and variance. A flat default prior is obtained by choosing the scale parameter very small.

Modeling prior information about q becomes tricky due to the involvement of the smoothing parameter k. From (4), we obtain

$$\mathrm{E}\big(q(\omega)|k\big) = \sum_{j=1}^{k} G_0\left(\frac{j-1}{k}, \frac{j}{k}\right]\beta(\omega; j, k-j+1), \quad (7)$$

which is the kth-order Bernstein kernel smoothing of the density function g_0. The larger the k, the less the extent of the smoothing. The unconditional prior expectation of q is thus the average of (7) with respect to the prior distribution of k. The prior variance of q is inversely proportional to $(M+1)$. Thus g_0 may be set to the prior guess about q while adjusting M to indicate the faith in this guess. To prevent oversmoothing, the parameter k should be kept as large as possible, with high prior probability satisfying the tail conditions of the consistency theorem. A diffuse prior may be obtained by setting g_0 to the uniform density and M to a small number.

3. A MARKOV CHAIN MONTE CARLO ALGORITHM FOR SIMULATION FROM THE POSTERIOR

The Whittle likelihood, although an approximate likelihood, is the Lebesgue density of a probability measure on $\mathbb{R}^\nu$ given the parameter f. Thus the pseudoposterior will be proper if the prior on f is proper. Note that the pseudoposterior is not the conditional distribution of f given the data $(X_1, \ldots, X_n)$ or given $(U_1, \ldots, U_\nu)$. The expectation of f with respect to this pseudoposterior distribution, say $\hat{f}$, may be used as a Bayes estimator of f. Finally, f^* is estimated by $\hat{f}^*(\omega^*) = \hat{f}(\omega^*/\pi)$ and $\hat{f}^*(-\omega^*) = \hat{f}^*(\omega^*)$. The variance and other features of this posterior distribution may be used for various statistical purposes.

The posterior distribution is analytically intractable, and thus computations are done via Monte Carlo approximations with the help of the MCMC method. Using the Sethuraman (1994) representation for a Dirichlet process, the infinite-dimensional parameter G is represented as

$$G = \sum_{l=1}^{\infty} p_l \delta_{Z_l}, \quad (8)$$

where $p_1 = V_1$, $p_l = (1 - V_1) \cdots (1 - V_{l-1}) V_l$ for $l \geq 2$, $V_l \sim \text{beta}(1, M)$, $Z_l \sim G_0$, and $(Z_1, Z_2, \ldots, V_1, V_2, \ldots)$ are all independent. Thus (G, k, τ) may be reparameterized as $(Z_1, Z_2, \ldots, V_1, V_2, \ldots, k, \tau)$.

Because the infinite series in (8) is almost surely convergent, the random vector (V_l, Z_l), as l increases to infinity, will have a diminishing effect on the prior distribution and thus on the posterior distribution of f. Hence we truncate the series at some large L. Although the truncation has negligible effect, the extent of this effect differs from the prior to the posterior. Thus the truncation point should be chosen based on the posterior distribution according to the tolerance of error in computation, and L may be determined numerically. The effect of truncation on the distribution of functionals of a Dirichlet process has been studied by Muliere and Tardella (1998); however, their results do not apply to our case, because the posterior distribution of G is not Dirichlet.

After the truncation, G may be represented as

$$G = \sum_{l=1}^{L} p_l \delta_{Z_l} + (1 - p_1 - \cdots - p_L)\delta_{Z_0},$$

where $Z_0 \sim G_0$ and is independent of the other parameters. The last term is added to make G a distribution function even after the truncation. Now the problem reduces to a parametric one with finitely many parameters $(Z_0, Z_1, \ldots, Z_L, V_1, \ldots, V_L, k, \tau)$. The functional parameter f may be written as a function of these univariate parameters as

$$f(\omega) = \tau \sum_{j=1}^{k} w_{j,k}\beta(\omega; j, k-j+1),$$

where $w_{j,k} = \sum_{l=0}^{L} p_l I\{\frac{j-1}{k} < Z_l \leq \frac{j}{k}\}$ and $p_0 = 1 - p_1 - \cdots - p_L$. The joint posterior density of $(Z_0, Z_1, \ldots, Z_L, V_1, \ldots, V_L, k, \tau)$ is proportional to

$$\left[\prod_{m=1}^{\nu} \frac{1}{f(2m/n)} e^{-U_m/f(2m/n)}\right] \times \left[\prod_{l=1}^{L} M(1 - v_l)^{M-1}\right]$$
$$\times \left[\prod_{l=0}^{L} g_0(z_l)\right]\rho(k)\pi(\tau).$$

Gibbs sampling techniques are applied for componentwise updating. One may easily sample from the posterior distribution of the discrete parameter k conditioned on the other parameters. The posterior densities of τ, Z_i's, or V_i's, conditional on the other parameters, are known up to a constant. Thus the Metropolis algorithm is applied within the Gibbs sampler steps to update τ, Z_i's, and V_i's. If the prior on τ is an inverse-gamma distribution, then the posterior distribution of τ conditional on the other parameters is also inverse-gamma, from which exact samples can be drawn.

The support of Z_i's and V_i's are [0, 1]. Given a V_l at the tth iteration step, sample a "candidate point" V_l^* from the uniform distribution on $[V_l - \varepsilon_l, V_l + \varepsilon_l]$ "modulo the circular unit interval"; that is, if the sampled point from this uniform distribution is more than 1, take the decimal part only, and if the sampled point is negative, add 1 to make it between 0 and 1. This gives rise to a symmetric "proposal distribution." The value of ε_l should be chosen to match the posterior variance

of the corresponding V_l, which is usually different for different l and may be found numerically for a given dataset. The variables Z_l's are updated similarly.

The starting values of τ may be set to the sample variance divided by 2π, whereas the starting value of k may be set to some large integer K_0. The approximate posterior mode of Z_i's and V_i's given the starting values of τ and k may be considered the starting values for the respective variables. These may be obtained by first setting their values to $1/2$ and then running the foregoing updating algorithm for a while with a larger ε and with fixed (τ, k).

Although we have reduced the problem to a parametric problem, we do not lose any aspect of the nonparametric modeling, because the cutoff points for the number of parameters are chosen depending on the data. Unlike the frequentist methods, a large choice of L will not overfit the data. On the contrary, larger L will always be better in terms of accuracy, because the true posterior distribution involves all of the Z_i's and V_i's. However, a very large L increases computing time while adding little to the precision of the estimator. Liseo et al. (2001) used a similar idea of reparameterization and truncation to reduce an infinite-dimensional problem into a finite-dimensional problem. They considered the Karhunen–Loéve series expansion for the Brownian motion and the prior for the log spectral density, then truncated the infinite series to a finite sum.

Although a computationally efficient implementation of the described MCMC procedure requires some manual input case by case, a default mechanism may be set only at the expense of computational time. The ε_l's in the proposal distribution of the V_l's and Z_l's are chosen to match the posterior variance of the corresponding components. From theoretical results in the parametric literature and some simulation study, we found that $\varepsilon_l = l/(l + 2\sqrt{n})$ works well for almost every dataset. From our simulation study, we found that $L = \max\{20, n^{1/3}\}$ works very well as a conservative truncation point. The number 20 is chosen to keep the truncation point sufficiently high in small samples. The burn-in time and the Monte Carlo sample sizes may be set large, say 3,000 and 10,000. This will result in a default mechanism that will work for almost all spectral densities and all sample sizes without any manual intervention.

4. NUMERICAL ILLUSTRATION

4.1 Simulation Study

This section presents results of a simulation study to compare the performance of our nonparametric Bayes estimator (BE) with some existing estimators of the spectral density. The BE is computed using a diffuse prior. For the Dirichlet distribution on G, G_0 is chosen to be the uniform distribution on $[0, 1]$, whereas M is set to 1, a small number compared with the sample size. The prior on k is $\rho(k) = c \exp(-.05k^2)$, and the prior on τ^{-1} is exponential with mean 10^{-20}. The MCMC algorithm is implemented using the default mechanism. Three estimators are considered for comparison.

Smoothed Periodogram Estimate (SPE). The periodogram at the discrete Fourier frequencies are smoothed using the Bartlett–Priestley quadratic kernel $w(l) = [1 - (l/L)^2] \times I\{|l| < L\}$. The smoothing parameter L is found by minimizing the cross-validation score proposed by Hurvich (1985).

Autoregression Spectral Estimate (ARE). An autoregression (AR) model is fitted to the data, with the order selected by the Akaike information criterion, and the spectral density of the fitted model is used as the estimate. The parameters of the models are estimated via the sample version of Yule–Walker equations that are solved by the Levinson–Durbin algorithm.

Penalized Whittle Likelihood Estimator (PLE). This is the Whittle likelihood-based estimator proposed by Pawitan and O'Sullivan (1994). The estimate is obtained with the automatic selection of the smoothing parameter by using the S–PLUS program provided by the authors.

Three different time series models are considered for the simulation study:

$$X_t = 1.435X_{t-1} - .734X_{t-2} + .129X_{t-3} + \epsilon_t, \quad (9)$$

$$X_t = .9X_{t-4} + .7X_{t-8} - .63X_{t-12} + \epsilon_t, \quad (10)$$

and

$$X_t = \epsilon_t + \sum_{l=1}^{\infty} \psi_l \epsilon_{t-l}, \qquad \psi_l = (-1)^l/l^3; \quad (11)$$

where the ϵ_t's are iid with mean 0 and variance 1. The two AR processes in (9) and (10) were first given by Wahba (1980) and have been used for simulation studies by various authors. The AR(3) process in (9) and the linear process in (11) have relatively smooth spectral density, whereas the AR(12) process in (10) has few sharp spikes. Two different distributions are considered for ϵ_t: (a) $N(0, 1)$, which is symmetric, and (b) exponential(1) minus 1, which is positively skewed.

Data from the AR(p) processes are generated with the first p values of X_t set to 0, and the successive values are generated from (9) or (10). The first 20,000 values are discarded to reach stationarity. For the linear process in (11), the infinite sum is truncated at 1,000 terms for both generating the data and computing the true spectral density.

Datasets are generated for $n = 64, 128, 256$, and 512 with 1,000 replicates in each case. All of the estimates are computed for each sample and compared in terms of the L_1-error or integrated absolute error (IAE),

$$\text{IAE} = \|\hat{f}^* - f^*\|_1 = \int_0^{\pi} |\hat{f}^*(\omega) - f^*(\omega)| \, d\omega. \quad (12)$$

The mean L_1-errors from these 1,000 replicates are given in Table 1, and their boxplots are presented in Figure 1.

As expected, the nonparametric estimators (BE, PLE, and SPE) outperform the ARE in the linear process but underperform in the AR(12) process. Surprisingly, they also outperform the ARE in the AR(3) process for small sample sizes. Perhaps one reason is that as the MLE, the ARE tries to minimize the Kullback–Leibler divergence measure for the model, which may not necessarily minimize the L_1-error of the corresponding spectral density estimates. Moreover, in small sample sizes, the ARE often underestimates the order of the model, leading to large errors. The nonparametric estimates perform worse than the ARE in the AR(12) process, due to a smoothing out of the peaks. The extent of smoothing is greatest for the BE and least for the SPE. Although the BE mostly detects the peak correctly, it underestimates the magnitude of a sharp peak, thus leading to

Table 1. Mean L_1-Error From the 1,000 Monte Carlo Replicates

| Methods	$\epsilon_l \sim N(0, 1)$				$\epsilon_l \sim exp(1) - 1$			
	$n = 64$	$n = 128$	$n = 256$	$n = 512$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
AR(3) process								
BE	.7972	.6303	.5376	.4480	.8689	.6785	.5529	.4801
PLE	.8803	.6917	.5549	.4228	.9259	.7803	.6344	.4786
SPE	1.1665	.8674	.6266	.4832	1.3454	.9654	.7483	.5399
ARE	1.1608	.8802	.6341	.4621	1.3389	.9741	.7459	.5341
AR(12) process								
BE	28.200	19.005	12.981	10.537	28.387	19.009	13.027	10.552
PLE	36.429	20.126	13.127	9.862	36.502	19.736	13.226	9.905
SPE	32.233	19.314	13.809	11.139	32.019	19.466	13.972	11.221
ARE	16.708	13.984	11.231	8.576	16.843	14.283	11.438	8.685
Linear process								
BE	.2698	.2298	.1914	.1594	.3406	.2619	.2068	.1560
PLE	.2812	.2274	.1839	.1486	.3493	.2731	.2044	.1511
SPE	.3252	.2597	.2271	.1815	.4064	.3167	.2364	.1736
ARE	.4039	.3179	.2475	.2235	.4634	.3663	.2678	.1996
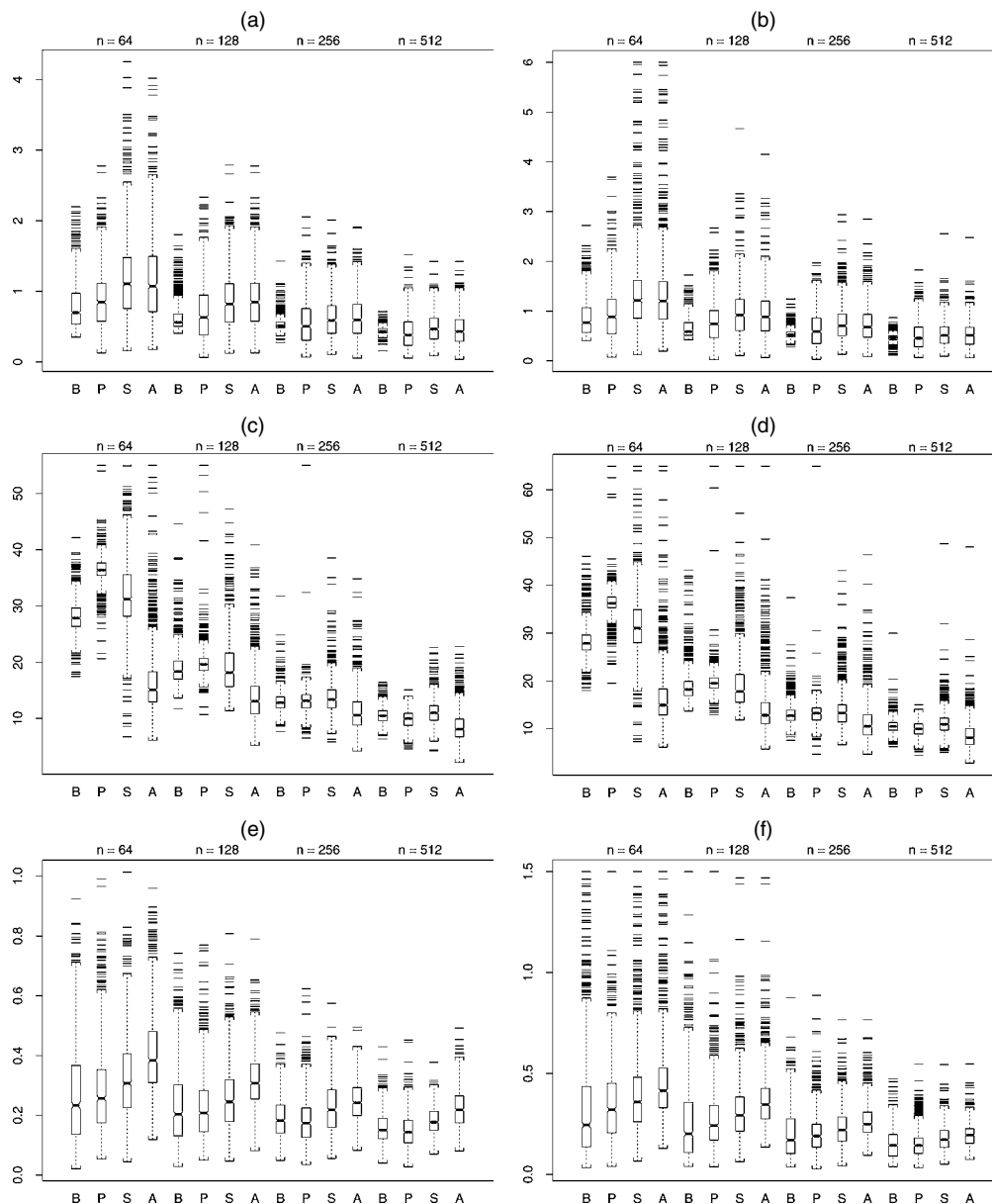
Figure 1. Boxplots of the L_1-Errors From the 1,000 Monte-Carlo Replicates. (a) AR(3) process with normal ϵ_i; (b) AR(3) process with exponential ϵ_i; (c) AR(12) process with normal ϵ_i; (d) AR(12) process with exponential ϵ_i; (e) Linear process with normal ϵ_i; (f) Linear process with exponential ϵ_i. (B = BE; P = PLE; S = SPE; A = ARE.)

a higher L_1-error. In contrast, the PLE and the SPE detect false spikes in small samples. Boxplots in Figure 1 reveal that the L_1-error of the BE has less variability than the other methods. In all three processes, the BE outperform the SPE for all sample sizes and outperforms the PLE for all but very large samples.

4.2 Analysis of the Sunspot Dataset

Here we analyze the sunspot dataset, which consists of the annual average value of the daily index of the number of sunspots for the year 1700–1987. A visual study of the time series plot of these 288 observations reveals a positive skewness. After a square root transformation, the data look more symmetric and stationary, with a constant mean but no trend. We subtract the mean and apply the four procedures to produce the corresponding spectral estimates. The estimates are plotted in the logarithmic scale in Figure 2(a). All of these estimates reveal a large peak of the spectral density at about $\omega_0 = .58$, indicating a strong periodic cycle of period length $2\pi/\omega_0 \approx 11$. The ARE has the sharpest peak. The three nonparametric estimates have comparable peaks. The BE is found to be much smoother than the other two nonparametric estimates. A pointwise 95% credible band is computed from the MCMC samples, and is plotted on a log scale in Figure 2(b).

5. POSTERIOR CONSISTENCY

Consistency of the posterior distribution is an important large-sample property for the validation of a Bayesian method. Because posterior consistency may fail in an infinite-dimensional model, it is important to check the consistency of any nonparametric Bayesian procedure. Here we provide assumptions on the time series and on the prior, which are required for posterior consistency. Let f_0^* be the true spectral density and let $f_0(\omega) = f_0^*(\pi\omega)$, $0 \le \omega \le 1$.

- Assumptions on the time series:

    (M1) The time series is Gaussian with $\sum_{r=0}^{\infty} r^\alpha \times \gamma(r) < \infty$ for some $\alpha > 0$.
    (M2) For all ω^*, $f_0^*(\omega^*) > 0$.

- Assumptions on the prior:

    (P1) For all k, $0 < \rho(k) \le Ce^{-ck\log k}$ for some constants $C, c > 0$,

(P2) g_0 is bounded, continuous, and bounded away from 0.

(P3) The parameter τ is assumed fixed and known.

The following theorem provides the posterior consistency for the proposed method.

Theorem 1. Under the foregoing assumptions, for any $\varepsilon > 0$,

$$\hat{\Pi}_n\{f^*: \|f^* - f_0^*\|_1 > \varepsilon\} \to 0$$

in $P_{f_0^*}^n$-probability, where $\hat{\Pi}_n$ is the pseudoposterior distribution computed using the Whittle likelihood of the periodogram ordinates $(U_1, \dots, U_\nu)$ and $P_{f_0^*}^n$ is the actual distribution of the data $(X_1, \dots, X_n)$.

Remark 1. The conclusion of Theorem 1 still holds if the degenerated prior on τ is replaced by a sequence of prior distributions that asymptotically bracket the true value τ_0; that is, the prior support of τ is in $[\tau_0 - \delta_n, \tau_0 + \delta_n]$ for some $\delta_n \to 0$.

The assumptions about the time series are commonly used in the literature. Assumption (M1) in particular implies that f_0^* is bounded continuous and thus is a short-memory process. Assumption (M2) holds for an invertible linear process. Condition (P1) is satisfied if the tail of the prior distribution of k is no heavier than that of a Poisson distribution. The consistency result requires only that the prior on G has a full weak support. Thus the prior could be different from the Dirichlet process. The hypothesis of Gaussianity is required only to connect the Whittle likelihood to the true distribution of the data through a contiguity argument.

The prior for τ is required to sit in a neighborhood of the true value τ_0 because of the lack of a uniformly exponentially consistent test for testing $\tau = \tau_0$ against $\tau > \tau_0 + \varepsilon$. A two-stage empirical Bayes modification, in which a fraction of the sample is used to determine the neighborhood of τ_0 and the remaining observations are used to update the prior, may be implemented. However, in our simulations we find that the usual BE performs well with a flat prior on τ.

To prove the consistency theorem, first a general result (Thm. A.1) is presented in Appendix A that gives sufficient conditions for posterior consistency for a triangular array of independent, nonidentically distributed observations. This result is then applied to prove posterior consistency assuming that
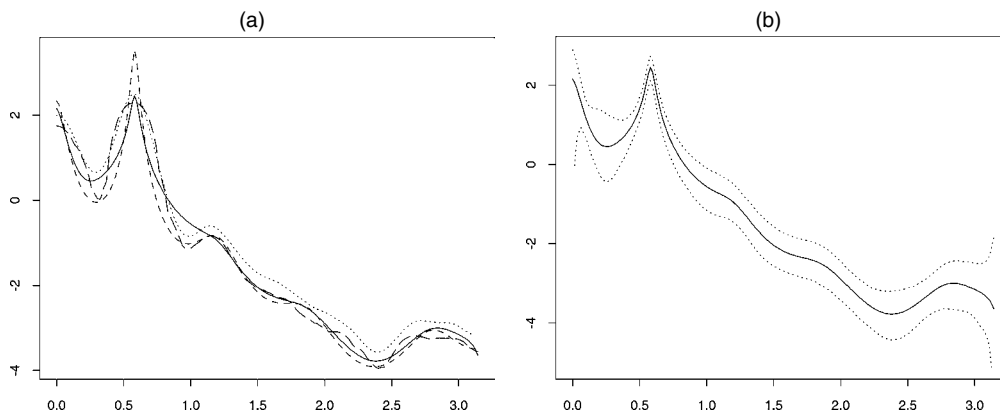

Figure 2. Spectral Estimates of the Square-Root–Transformed Sunspot Data in the Logarithmic Scale. (a) All four spectral estimates; (b) pointwise 95% credible band. (—— BE; ········ PLE; —— SP; - - - AR.)

the distribution of $(U_1, \ldots, U_\nu)$ is the product measure of independent exponential distributions. However, $(U_1, \ldots, U_\nu)$ are independent exponentials only in an asymptotic sense. The consistency under the true distribution of $(U_1, \ldots, U_\nu)$, and thus under the true distribution of $(X_1, \ldots, X_n)$, is proved using the contiguity result of Theorem B.1 in Appendix B.

6. CONCLUSIONS

We have proposed a nonparametric Bayesian method for the estimation of a spectral density of a stationary time series. Although the consistency result is obtained for Gaussian time series with variance approximately known, the method is applicable in a wider context. The simulation results also support that fact. The prior may be generalized to a random linear combination of kernels. The coefficients may come from an independent increment process such as a Lévy process (Hjort 1990). Similar posterior consistency results will still hold provided that the random process has full weak support. However, the assumption of known process variance may not be eliminated with such a prior, because that assumption is used to find an uniformly exponentially consistent test. This is a model restriction and not an artifact of the prior elicitation. The computation will become more prior-specific.

APPENDIX A: A GENERAL CONSISTENCY THEOREM

A celebrated theorem of Schwartz (1965), and its extensions obtained by Barron et al. (1999) and Ghosal et al. (1999) gave sufficient conditions for posterior consistency in terms of existence of appropriate tests and prior positivity of a neighborhood defined by the Kullback–Leibler divergence. However, those theorems are limited to independent and identically distributed (iid) observations. Because the periodogram ordinates are nonidentically distributed and the discrete Fourier frequencies change with the sample size, an extension of Schwartz's result for a triangular array of independent but nonidentically distributed observations is required.

Here we present such a theorem, which is close in spirit to a similar theorem of Amewou-Atisso, Ghosal, Ghosh, and Ramamoorthi (2003, thm. 2.1). There are, however, some important differences. Amewou-Atisso et al. (2003) assumed stronger conditions and conclude almost sure convergence, whereas we consider only convergence in probability, and the corresponding relaxed conditions are easier to verify for our problem. Further, the result of Amewou-Atisso et al. (2003) is tailor-made for regression and semiparametric problems, in which only weak consistency for the nonparametric part is usually sufficient. In contrast, we consider a curve estimation problem, and hence we need to address convergence issues in stronger topologies. This compels us to use sieves like those of Barron et al. (1999) and Ghosal et al. (1999) for the existence of tests.

Theorem A.1. Let $Z_{i,n}$ be independently distributed with density $p_{i,n}(\cdot; \theta)$ $i = 1, \ldots, r_n$, with respect to a common σ-finite measure, where the parameter θ belongs to an abstract measurable space Θ. The densities $p_{i,n}(\cdot, \theta)$ are assumed to be jointly measurable. Let $\theta_0 \in \Theta$ and let $\bar{\Theta}_n$ and $\mathcal{U}_n$ be two subsets of Θ. Let θ have prior Π on Θ. Put $K_{i,n}(\theta_0, \theta) = E_{\theta_0}(\Lambda_i(\theta_0, \theta))$ and $V_{i,n}(\theta_0, \theta) = \text{var}_{\theta_0}(\Lambda_i(\theta_0, \theta))$, where $\Lambda_i(\theta_0, \theta) = \log(p_{i,n}(Z_{i,n}; \theta_0)/p_{i,n}(Z_{i,n}; \theta))$. The following conditions hold:

(C1) Prior positivity of neighborhoods. Suppose that there exists a set B with $\Pi(B) > 0$ such that

(a) $\frac{1}{r_n^2} \sum_{i=1}^{r_n} V_{i,n}(\theta_0, \theta) \to 0$ for all $\theta \in B$, and

(b) $\liminf_{n \to \infty} \Pi(\{\theta \in B : \frac{1}{r_n} \sum_{i=1}^{r_n} K_{i,n}(\theta_0, \theta) < \varepsilon\}) > 0$ for all $\varepsilon > 0$,

(C2) Existence of tests. Suppose that there exists test functions $\{\Phi_n\}$, subsets $\Theta_n \subset \bar{\Theta}_n$, and constants $C_1, C_2, c_1, c_2 > 0$ such that

(a) $E_{\theta_0} \Phi_n \to 0$,

(b) $\sup_{\theta \in \mathcal{U}_n^c \cap \Theta_n} E_\theta(1 - \Phi_n) \leq C_1 e^{-c_1 r_n}$, and

(c) $\Pi(\bar{\Theta}_n \cap \Theta_n^c) \leq C_2 e^{-c_2 r_n}$.

Then

$$\Pi(\theta \in \mathcal{U}_n^c \cap \bar{\Theta}_n | Z_{1,n}, \ldots, Z_{r_n,n}) \to 0 \tag{A.1}$$

in $P_{\theta_0}^n$-probability.

Usually this theorem is applied for $\bar{\Theta}_n = \Theta$ for all n. If, however, condition (C2) can be verified only on a part of Θ, which may possibly depend on n, then the foregoing formulation may be useful. In this case, however, the final conclusion should then be complemented by showing that $\Pi(\bar{\Theta}_n^c | Z_1, \ldots, Z_{r_n}) \to 0$ in $P_{\theta_0}^n$-probability by some alternative method. Posterior consistency is proved by choosing a fixed neighborhood $\mathcal{U}_n^c$, whereas the rate of convergence is obtained by choosing a shrinking neighborhood.

Condition (C1) asserts that every Kullback–Leibler neighborhood of θ_0 has positive prior probability and thus ensures that the true parameter value is not excluded from the prior support. Condition (C2) asserts that, given a neighborhood $\mathcal{U}_n$ of the true parameter value θ_0 in some topology of interest, except in a set Θ_n^c with exponentially small prior probability, the model can separate θ_0 from a $\theta \in \mathcal{U}_n^c$ through an exponentially powerful test.

Proof of Theorem A.1. For ease of notation, we drop the second subscript n from the observations. Also, we write r for r_n. Note that, because $0 \leq \Phi_n \leq 1$ and $0 \leq \Pi(\theta \in \mathcal{U}_n^c | Z_1, \ldots, Z_r) \leq 1$, we have

$\Pi(\theta \in \bar{\Theta}_n \cap \mathcal{U}_n^c | Z_1, \ldots, Z_r)$

$$\leq \Phi_n + (1 - \Phi_n)\Pi(\theta \in \bar{\Theta}_n \cap \mathcal{U}_n^c | Z_1, \ldots, Z_r).$$

The first term on the right side goes to 0 in $P_{\theta_0}^n$-probability by (C2)(a). The second term is bounded above by

$$\left((1 - \Phi_n) \int_{\mathcal{U}_n^c \cap \Theta_n} \prod_{i=1}^{r} \frac{p_i(Z_i, \theta)}{p_i(Z_i, \theta_0)} \, d\Pi(\theta) \right.$$

$$\left. + \int_{\bar{\Theta}_n \cap \Theta_n^c} \prod_{i=1}^{r} \frac{p_i(Z_i, \theta)}{p_i(Z_i, \theta_0)} \, d\Pi(\theta) \right)$$

$$\times \left(\int_\Theta \prod_{i=1}^{r} \frac{p_i(Z_i, \theta)}{p_i(Z_i, \theta_0)} \, d\Pi(\theta) \right)^{-1}.$$

We show that the two terms in the numerator are exponentially small and that

$$e^{\beta r} \int_\Theta \prod_{i=1}^{r} \frac{p_i(Z_i, \theta)}{p_i(Z_i, \theta_0)} \, d\Pi(\theta) \to \infty \tag{A.2}$$

in $P_{\theta_0}^n$-probability for all $\beta > 0$. Then the result follows.

By Fubini's theorem, for any nonnegative function ψ_n of $(Z_1, \ldots, Z_r)$ and subset H of Θ,

$$E_{\theta_0}\left[\int_H \prod_{i=1}^{r} \frac{p_i(Z_i, \theta)}{p_i(Z_i, \theta_0)} \psi_n(Z_1, \ldots, Z_r) \, d\Pi(\theta) \right]$$

$$\leq \int_H E_\theta(\psi_n) \, d\Pi(\theta).$$

Hence, by Conditions (C2)(b) and (c),

$$\mathrm{E}_{\theta_0}\left[(1-\Phi_n)\int_{\Theta_n\cap\mathcal{U}_n^c}\prod_{i=1}^r\frac{p_i(Z_i,\theta)}{p_i(Z_i,\theta_0)}\,d\Pi(\theta)\right]$$

$$\le \sup_{\theta\in\Theta_n\cap\mathcal{U}_n^c}\mathrm{E}_\theta(1-\Phi_n)\le C_1 e^{-c_1 r}$$

and

$$\mathrm{E}_{\theta_0}\left[\int_{\bar{\Theta}_n\cap\Theta_n^c}\prod_{i=1}^r\frac{p_i(Z_i,\theta)}{p_i(Z_i,\theta_0)}\,d\Pi(\theta)\right]\le\Pi(\bar{\Theta}_n\cap\Theta_n^c)$$

$$\le C_2 e^{-c_2 r}.$$

For (A.2), observe that by Chebyshev's inequality, for all $\theta\in B$,

$$P_{\theta_0}\left\{\left|\frac{1}{r}\sum_{i=1}^r\Lambda_i(\theta_0,\theta)-\frac{1}{r}\sum_{i=1}^r K_i(\theta_0,\theta)\right|\ge\varepsilon\right\}\le\frac{1}{r^2\varepsilon^2}\sum_{i=1}^r V_i(\theta_0,\theta).$$

Fix $\varepsilon>0$. Let W_n be the set

$$\left\{(\theta,z_1,\ldots,z_r):\theta\in B,\left|\frac{1}{r}\sum_{i=1}^r\Lambda_i(\theta_0,\theta)-\frac{1}{r}\sum_{i=1}^r K_i(\theta_0,\theta)\right|<\varepsilon\right\}.$$

Let $\delta>0$ to be chosen later and put

$$V_n=\left\{(z_1,\ldots,z_r):\Pi\left(W_n^{(z_1,\ldots,z_r)}\right)>(1-\delta)\Pi(B)\right\},$$

where $W_n^{(z_1,\ldots,z_r)}$ is the $(z_1,\ldots,z_r)$-section of W_n. Let $W_{n,\theta}$ denote the θ-section of W_n. Thus by (C1)(a), for all $\theta\in B$, $P_{\theta_0}^n(W_{n,\theta})\to 1$. Now by Markov's inequality and Fubini's theorem,

$$P_{\theta_0}^n(V_n^c)=P_{\theta_0}^n\left\{\Pi\left(W_n^{(Z_1,\ldots,Z_r)}\right)<(1-\delta)\Pi(B)\right\}$$

$$\le\frac{1}{\delta\Pi(B)}\int\Pi\left(B\cap\left(W_n^{(z_1,\ldots,z_r)}\right)^c\right)dP_{\theta_0}^n(z_1,\ldots,z_r)$$

$$=\frac{1}{\delta\Pi(B)}\int_B P_{\theta_0}^n(W_{n,\theta}^c)\,d\Pi(\theta),$$

which converges to 0 by the dominated convergence theorem.

Let $H_n=\{\theta\in B:\frac{1}{r}\sum_{i=1}^r K_i(\theta_0,\theta)<\varepsilon\}$. By (C1)(b), $\Pi(H_n)\ge 2\delta$ for all sufficiently large n, if $\delta>0$ is chosen sufficiently small. By Bonferroni's inequality, for all $(z_1,\ldots,z_r)\in V_n$ and $n\ge N$,

$$\Pi\left(W_n^{(z_1,\ldots,z_r)}\cap H_n\right)\ge\Pi\left(W_n^{(z_1,\ldots,z_r)}\right)+\Pi(H_n)-\Pi(B)$$

$$\ge\Pi(B)(1-\delta)+2\delta-\Pi(B)\ge\delta.$$

Clearly, if $(Z_1,\ldots,Z_r)\in V_n$ and $\theta\in W_n^{(Z_1,\ldots,Z_r)}\cap H_n$, then

$$\prod_{i=1}^r\frac{p_i(Z_i,\theta)}{p_i(Z_i,\theta_0)}=\exp\left[-\sum_{i=1}^r\Lambda_i(\theta_0,\theta)\right]\ge e^{-2\varepsilon r}.$$

Now if $(Z_1,\ldots,Z_r)\in V_n$, then the left side of (A.2) is bounded by

$$e^{\beta r}\int_{W_n^{(Z_1,\ldots,Z_r)}\cap C_n}\prod_{i=1}^r\frac{p_i(Z_i,\theta)}{p_i(Z_i,\theta_0)}\,d\Pi(\theta)\ge\frac{\delta}{2}e^{(\beta-2\varepsilon)r}.$$

We may choose $\varepsilon<\beta/2$. As $P_{\theta_0}^n(V_n)\to 1$, the assertion holds.

Remark A.1. The foregoing result also holds for a sequence of priors Π_n replacing Π, provided that (C1)(a) is strengthened to uniform convergence. The proof is almost identical.

APPENDIX B: PROOF OF THEOREM 1

Theorem 1 is proved by checking the conditions of Theorem A.1. It is sufficient to prove L_1-consistency for the transformed spectral density f. Thus f plays the role of θ in Theorem A.1 and Θ is the space of all nonnegative integrable functions on $[0,1]$. We consider consistency with respect to an L_1-neighborhood,

$$\mathcal{U}=\mathcal{U}_n=\left\{f:\|f-f_0\|_1=\int_0^1|f(\omega)-f_0(\omega)|\,d\omega<\varepsilon\right\},$$

for some arbitrary fixed $\varepsilon>0$.

Let $P_{f_0^*}^n$ denote the actual distribution of the data $(X_1,\ldots,X_n)$ and Γ_n denote the actual distribution of the periodogram ordinates $(U_1,\ldots,U_\nu)$. Then Γ_n is induced by $P_{f_0^*}^n$ through the periodogram, and it is sufficient to study convergence under Γ_n. Thus $(U_1,\ldots,U_\nu)$ play the roles of observations in Theorem A.1. However, Theorem A.1 requires that the U_i's be independent, which is only asymptotically true. A contiguity result from Choudhuri, Ghosal, and Roy (2004) is presented here that can bridge the gap. Let $\tilde{\Gamma}_n$ denote joint distribution of independent exponential random variables with means $f_0(2i/n)$, $i=1,\ldots,\nu$.

Theorem B.1. Under assumptions (M1) and (M2), Γ_n and $\tilde{\Gamma}_n$ are mutually contiguous.

By virtue of this result, "convergence in probability" is the same under Γ_n and $\tilde{\Gamma}_n$. Hence it suffices to prove posterior consistency assuming that the U_i's are independent exponential random variables, in which case Theorem A.1 is applicable.

B.1 Positivity of the Prior Probabilities

Occasionally, it will be convenient to view f as τq and f_0 as $\tau_0 q_0$. Let Q_0 denote the probability measure corresponding to q_0. First, we show that, for every $\varepsilon>0$,

$$\Pi\{f:\|f-f_0\|_\infty<\varepsilon\}>0,\tag{B.1}$$

where $\|f-f_0\|_\infty=\sup_{0\le\omega\le 1}|f(\omega)-f_0(\omega)|$. Because $b(\cdot;k,Q_0)\to q_0$ uniformly, choose k such that $\|b(\cdot;k,Q_0)-q_0\|_\infty<\varepsilon/(2\tau_0)$. Then

$$|b(\omega;k,G)-b(\omega;k,Q_0)|$$

$$=\left|\sum_{j=1}^k\left\{G\left(\frac{j-1}{k},\frac{j}{k}\right]-Q_0\left(\frac{j-1}{k},\frac{j}{k}\right]\right\}\beta(\omega;j,k+1-j)\right|$$

$$\le k\max_{1\le j\le k}\left|G\left(\frac{j-1}{k},\frac{j}{k}\right]-Q_0\left(\frac{j-1}{k},\frac{j}{k}\right]\right|,$$

as $\sum_{j=1}^k\beta(\omega;j,k+1-j)=k$. Now if $f(\omega)=\tau_0 b(\omega;k,G)$, then

$$\|f-f_0\|_\infty\le\tau_0\|b(\cdot;k,G)-b(\cdot;k,Q_0)\|_\infty+\tau_0\|b(\cdot;k,Q_0)-q_0\|_\infty.$$

Because every index k has positive probability under the prior and the Dirichlet process has full support,

$$\left\{(k,G):\left|G\left(\frac{j-1}{k},\frac{j}{k}\right]-Q_0\left(\frac{j-1}{k},\frac{j}{k}\right]\right|<\frac{\varepsilon}{2k\tau_0},j=1,\ldots,k\right\}$$

has positive probability as well. By assumption (P3), $\int f=\tau_0$ with prior probability 1, and thus (B.1) follows.

Let $\omega_i=2i/n$ denote the ith Fourier frequency, rescaled into $[0,1]$. Assuming that the U_i's are exponential random variables, we have

$$K_i(f_0,f)=-\log\left(1+\frac{f_0(\omega_i)-f(\omega_i)}{f(\omega_i)}\right)+\frac{f_0(\omega_i)-f(\omega_i)}{f(\omega_i)}$$

and

$$V_i(f_0,f)=\left(\frac{f_0(\omega_i)-f(\omega_i)}{f(\omega_i)}\right)^2.$$

By assumption (M1), f_0 is continuous, and by assumption (M2), $f_0 > 0$. Hence, f_0 is bounded and bounded away from 0 on entire $[0, 1]$. Let $s = \inf_{0 \le \omega \le 1} f_0(\omega)$ and

$$B = \{f : \|f - f_0\|_\infty \le s/2\}. \tag{B.2}$$

Then $\Pi(B) > 0$ by (B.1). For $f \in B$, we have $f \ge s/2$ and $-1/2 \le (f_0 - f)/f \le 1$. Hence,

$$\frac{1}{\nu^2} \sum_{i=1}^{\nu} V_i(f_0, f) \le \frac{1}{\nu^2} \sum_{i=1}^{\nu} \left(\frac{f_0(\omega_i) - f(\omega_i)}{f(\omega_i)} \right)^2 \le \frac{1}{\nu} \to 0$$

uniformly on B, so that (C1)(a) holds. For (C1)(b), using the inequality $y - \log(1 + y) \le 2y^2$ on $y \ge -\frac{1}{2}$, we have

$$\frac{1}{\nu} \sum_{i=1}^{\nu} K_i(f_0, f) \le \frac{2}{\nu} \sum_{i=1}^{\nu} \left(\frac{f_0(\omega_i) - f(\omega_i)}{f(\omega_i)} \right)^2$$

$$\le \frac{8}{s^2} \|f - f_0\|_\infty^2$$

on B. Thus (C1)(b) holds.

B.2 Existence of Tests

To find the exponentially powerful consistent tests, consider the sieves of the form

$$\Theta_n = \bigcup_{j=1}^{k_n} \{\tau_0 b(\cdot; j, \mathbf{w}) : \mathbf{w} \in \Delta_j\}. \tag{B.3}$$

Thus the elements of Θ_n are Bernstein polynomials of order k_n or less, with nonnegative coefficients, and integrate to τ_0. The tuning parameter k_n controls both the size and the ability of approximation of the sieve and is chosen later. Let λ denote the Lebesgue measure on the real line.

Lemma B.1. Let $\varepsilon > 0$, f and f_0 be nonnegative integrable functions on $[0, 1]$ with $f_0 \le c$, $\|f - f_0\|_1 > 4\varepsilon$, and $\int f = \int f_0$. Let $A = \{\omega : f(\omega) < f_0(\omega) - \varepsilon\}$. Then $\lambda(A) \ge \varepsilon/c$.

Proof. Note that $\|f - f_0\|_1 = 2 \int_{f < f_0} (f_0 - f)$ and, therefore,

$$2\varepsilon < \int_{f < f_0} (f_0 - f)$$

$$= \int_A (f_0 - f) + \int_{\{f_0 - \varepsilon \le f < f_0\}} (f_0 - f)$$

$$\le c\lambda(A) + \varepsilon.$$

The result now follows.

Lemma B.2. Let $f_0 \le c$ be continuous with $\int f_0 = \tau_0$ and $f \in \Theta_n$, with $\|f - f_0\| > 9\varepsilon$. Then

$$\#\{j : f(\omega_j) < f_0(\omega_j) - \varepsilon\} \ge \frac{n\varepsilon}{c'} - 2k_n$$

for some constant c', where $\#$ stands for the cardinality of a set.

Proof. Get k_0 be such that $\|f_0 - \tau_0 b_0\|_\infty < \varepsilon$, where $b_0 = b(\cdot; k_0, Q_0)$. Then $\|f - \tau_0 b_0\|_\infty > 9\varepsilon - \varepsilon = 8\varepsilon$. Note that b_0 is also a nonnegative integrable function with integral equal to 1 and $b_0 \le c'$ for some c'. Apply Lemma B.1 to $\tilde{A} = \{\omega : f(\omega) < \tau_0 b_0(\omega) - 2\varepsilon\}$ to obtain $\lambda(\tilde{A}) \ge 2\varepsilon/c'$. Note that on $\tilde{A}$, $f(\omega) < \tau_0 b_0(\omega) - 2\varepsilon < f_0(\omega) - \varepsilon$, so that $\tilde{A} \subset A$, where A is as in Lemma B.1. Now $f(\omega) - \tau_0 b_0(\omega)$ is a polynomial of degree at most $\max(k_n, k_0) = k_n$ for large n, so $\tilde{A}$ is the union of at most k_n open intervals. Let $h_1, h_2, \ldots$ be the lengths of these intervals. Because two successive ω_i's are $2/n$ apart, any interval of length h will contain at least $(nh/2) - 2$ many ω_i's. Therefore, the number of ω_i's contained in $\tilde{A}$ is at least $\sum_j \{(nh_j/2) - 2\} \ge (n\varepsilon/c') - 2k_n$.

Lemma B.3. Let Y_j be independent exponential random variables with means μ_j, $j = 1, \ldots, m$. Consider testing $H_0 : \mu_j = \mu_{0j}$ against $H_1 : \mu_j = \mu_{1j}$, where $\mu_{1j} < \mu_{0j} - \varepsilon$ for all j and $\mu_{0j} \le a$, where a and $\varepsilon > 0$ do not depend on m. Then there exists a test Φ_m and constants $\beta_1, \beta_2 > 0$, depending only on a and ε such that $E_{P_0}(\Phi_m) \le e^{-\beta_1 m}$ and $E_{P_1}(1 - \Phi_m) \le e^{-\beta_2 m}$, where P_0 and P_1 are the probability measures under the null and the alternative.

Proof. Consider a test Φ_m that rejects H_0 if $\sum_{j=1}^{m} Y_j/\mu_{0j} < m\{1 - (\varepsilon/2a)\}$. Under H_0, Y_j/μ_{0j}, $j = 1, \ldots, m$, are iid exponential with parameter 1 that has finite moment-generating function. Thus the exponential bound for the first kind of error is obtained by standard large-deviation estimates.

To estimate $E_{P_1}(1 - \Phi_m)$, write $Y_j/\mu_{0j} = (Y_j/\mu_{1j}) \times (\mu_{1j}/\mu_{0j})$. The first factor is exponentially distributed with mean 1, and the second is $\mu_{1j}/\mu_{0j} \le (\mu_{0j} - \varepsilon)/\mu_{0j} \le 1 - (\varepsilon/a)$. Under P_1, Y_j/μ_{0j} is stochastically smaller than $\tilde{Y}_j$, where $\tilde{Y}_j$ is exponentially distributed with mean $1 - (\varepsilon/a)$ and the $\tilde{Y}_j$'s are iid. Thus by standard large-deviation estimates, we obtain

$$E_{P_1}(1 - \Phi_m) \le \Pr\left(\sum_{j=1}^{m} \tilde{Y}_j \ge m\left(1 - \frac{\varepsilon}{2a}\right) \right) \le e^{-\beta_2 m}.$$

Lemma B.4. The ε-covering number $N(\varepsilon, \Theta_n, \|\cdot\|_\infty)$ of Θ_n in the supremum norm is given by $\log N(\varepsilon, \Theta_n, \|\cdot\|_\infty) \le k_n \log k_n + k_n \log(5\tau_0/\varepsilon)$.

Proof. Because $\beta(\omega; j, k - j + 1) \le k$ for all k, j, and ω, we have $\|b(\cdot, k_n, \mathbf{w}) - b(\cdot, k_n, \mathbf{w}^*)\|_\infty \le k_n \|\mathbf{w}_{k_n} - \mathbf{w}^*_{k_n}\|_1$, where $\|\cdot\|_1$ is the ℓ_1 norm on the k_n-simplex Δ_{k_n}. It then follows that $N(\varepsilon, \Theta_n, \|\cdot\|_\infty) \le \sum_{j=1}^{k_n} N(\varepsilon/\tau_0 k_n, \Delta_j, \|\cdot\|_1)$. By lemma A.4 of Ghosal and van der Vaart (2001), $N(\varepsilon, \Delta_j, \|\cdot\|_1) \le (5/\varepsilon)^{j-1}$ for $1 \le j \le k_n$. Thus the estimate follows.

Let $a = \|f_0\|_\infty$. Consider the test $\Phi_{n, f}$, which rejects the simple null f_0 against the simple alternative f if

$$\sum_{i \in I_{n, f}} \frac{U_i}{f_0(\omega_i)} > m\left(1 - \frac{\varepsilon}{2a}\right),$$

where $m = m(n, f) = \#I_{n, f}$ and $I_{n, f} = \{i : f(\omega_i) < f_0(\omega_i) - \varepsilon\}$. If $f \in \Theta_n$ and $\|f - f_0\|_1 > 9\varepsilon$, then by Lemma B.3, there are some constants $\beta'_1, \beta'_2 > 0$ such that $E_{f_0} \Phi_{n, f} \le e^{-\beta'_1 m}$ and $E_f(1 - \Phi_{n, f}) \le e^{-\beta'_2 m}$, where E_f stands for the expectation assuming that the U_i's are independent exponential random variables with mean $f(\omega_i)$. By Lemma B.2, for $k_n \le \delta_1 n$ with sufficiently small $\delta_1 > 0$, the foregoing bound can be reduced to $e^{-\beta_1 n}$ and $e^{-\beta_2 n}$, where $\beta_1, \beta_2 > 0$ are constants not depending on f.

To remove the dependence of test functions on f, we use the standard technique of covering a set by small balls and estimating the covering numbers. Get $f_1, \ldots, f_N \in \Theta_n$ with the property that for any $f \in \Theta_n$, there exists an f_j such that $\|f - f_j\|_\infty < \varepsilon/2$; here N is the $(\varepsilon/2)$-covering number of Θ_n in sup norm.

If $\|f - f_j\|_\infty < \varepsilon/2$, then for every $i \in I_{n, f_j}$, $f(\omega_i) < f_0(\omega_i) - \varepsilon/2$. Hence $E_f(1 - \Phi_{n, f_j}) \le e^{-\beta'' n}$ for some constant β'' depending neither on f nor on f_j. Put $\Phi_n = \max\{\Phi_{n, f_j} : j = 1, \ldots, N\}$. Now, by Lemma B.4, for all sufficiently large n,

$$E_{f_0}(\Phi_n) \le N(\varepsilon, \Theta_n, \|\cdot\|_\infty) e^{-\beta_1 n} \le e^{2k_n \log k_n - \beta_1 n}. \tag{B.4}$$

Clearly, $E_f(1 - \Phi_n) \le e^{-\beta'' n}$, because for any $f \in \Theta_n$, $\|f - f_j\|_\infty < \varepsilon/2$ for some j and $E_f(1 - \Phi_n) \le E_f(1 - \Phi_{n, f_j})$.

For the sieve Θ_n in (B.3), let the tuning parameter be chosen as $k_n = \lfloor \delta n / \log n \rfloor$, $\delta > 0$. Then $k_n \log k_n \le \delta n$ for large n, and the right side of (B.4) tends to 0 for sufficiently small δ. Thus both (C2)(a)

and (C2)(b) hold. To verify (C2)(c), note that by assumption (P3), the prior is concentrated on $\{f : \int f = \tau_0\}$. Thus we need to check that $\rho(k > k_n)$ is exponentially small, which holds by assumption (P1).

[Received May 2003. Revised March 2004.]

REFERENCES

Amewou-Atisso, M., Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. V. (2003), "Posterior Consistency for Semiparametric Regression Problems," *Bernoulli*, 9, 291–312.

Barron, A., Schervish, M., and Wasserman, L. (1999), "The Consistency of Posterior Distributions in Nonparametric Problems," *The Annals of Statistics*, 27, 536–561.

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods*, New York: Springer-Verlag.

Carter, C. K., and Kohn, R. (1997), "Semiparametric Bayesian Inference for Time Series With Mixed Spectra," *Journal of the Royal Statistical Society*, Ser. B, 59, 255–268.

Choudhuri, N., Ghosal, S., and Roy, A. (2004), "Contiguity of the Whittle Measure for a Gaussian Time Series," *Biometrika*, 91, 211–218.

Chow, Y. S., and Grenander, U. (1985), "A Sieve Estimate for the Spectral Density," *The Annals of Statistics*, 13, 998–1010.

Cogburn, R., and Davis, H. T. (1974), "Periodic Splines and Spectral Estimation," *The Annals of Statistics*, 2, 1108–1126.

Fan, J., and Kreutzberger, E. (1998), "Automatic Local Smoothing for Spectral Density Estimation," *Scandinavian Journal of Statistics*, 25, 359–369.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, Vol. II (2nd ed.), New York: Wiley.

Gangopadhyay, A. K., Mallick, B. K., and Denison, D. G. T. (1998), "Estimation of Spectral Density of a Stationary Time Series via an Asymptotic Representation of the Periodogram," *Journal of Statistical Planning and Inference*, 75, 281–290.

Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. V. (1997), "Non-Informative Priors via Sieves and Packing Numbers," in *Advances in Statistical Decision Theory and Applications*, eds. S. Panchapakesan and N. Balakrishnan, Boston: Birkhauser, pp. 119–132.

——— (1999), "Posterior Consistency of Dirichlet Mixtures in Density Estimation," *The Annals of Statistics*, 27, 143–158.

Ghosal, S., and van der Vaart, A. W. (2001), "Entropies and Rates of Convergence of Maximum Likelihood and Bayes Estimation for Mixtures of Normal Densities," *The Annals of Statistics*, 29, 1233–1263.

Hjort, N. L. (1990), "Nonparametric Bayes Estimators Based on Beta Processes in Models for Life History Data," *The Annals of Statistics*, 18, 1259–1294.

Hurvich, C. M. (1985), "Data-Driven Choice of a Spectrum Estimate: Extending the Applicability of Cross-Validation Methods," *Journal of the American Statistical Association*, 80, 933–940.

Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995), "Rate of Convergence for Logspline Spectral Density Estimation," *Journal of Time Series Analysis*, 16, 389–401.

Liseo, B., Marinucci, D., and Petrella, L. (2001), "Bayesian Semiparametric Inference on Long-Range Dependence," *Biometrika*, 88, 1089–1104.

Muliere, P., and Tardella, L. (1998), "Approximating Distributions of Functionals of Ferguson–Dirichlet Priors," *The Canadian Journal of Statistics*, 30, 269–283.

Ombao, H., Raz, J., Strawderman, R., and von Sachs, R. (2001), "A Simple Generalised Cross-Validation Method of Span Selection for Periodogram Smoothing," *Biometrika*, 88, 1186–1192.

Pawitan, Y., and O'Sullivan, F. (1994), "Nonparametric Spectral Density Estimation Using Penalized Whittle Likelihood," *Journal of the American Statistical Association*, 89, 600–610.

Petrone, S. (1999a), "Random Bernstein Polynomials," *Scandinavian Journal of Statistics*, 26, 373–393.

——— (1999b), "Bayesian Density Estimation Using Bernstein Polynomials," *The Canadian Journal of Statistics*, 26, 373–393.

Petrone, S., and Veronese, P. (2002), "Nonparametric Mixture Priors Based on an Exponential Random Scheme," *Statistical Methods and Applications*, 11, 1–20.

Petrone, S., and Wasserman, L. (2002), "Consistency of Bernstein Polynomial Posteriors," *Journal of the Royal Statistical Society*, Ser. B, 64, 79–100.

Schwartz, L. (1965), "On Bayes Procedures," *Zeitschrift für Wahrsheinlichkeitstheorie und verwandte Gebiete*, 4, 10–26.

Sethuraman, J. (1994), "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, 4, 639–650.

Vitale, R. A. (1975), "A Bernstein Polynomial Approach to Density Estimation," in *Statistical Inference and Related Topics*, ed. M. L. Puri, New York: Academic Press, pp. 87–100.

Wahba, G. (1980), "Automatic Smoothing of the Log Periodogram," *Journal of the American Statistical Association*, 75, 122–132.

Whittle, P. (1957), "Curve and Periodogram Smoothing," *Journal of the Royal Statistical Society*, Ser. B, 19, 38–63.

——— (1962), "Gaussian Estimation in Stationary Time Series," *Bulletin of the International Statistical Institute*, 39, 105–129.