

A Shrinkage Estimator for Spectral Densities

Carsten H. Botts
Department of Mathematics and Statistics
Williams College
Williamstown, Massachusetts 01267, USA
cbotts@williams.edu

Michael J. Daniels
Department of Statistics
University of Florida
Gainesville, Florida 32611, USA
mdaniels@stat.ufl.edu

Summary

We propose a shrinkage estimator for spectral densities based on a multilevel normal hierarchical model. The first level captures the sampling variability via a likelihood constructed using the asymptotic properties of the periodogram. At the second level, the spectral density is shrunk towards a parametric time series model. To avoid selecting a particular parametric model for the second level, a third level is added which induces an estimator that averages over a class of parsimonious time series models. The estimator derived from this model, the model averaged shrinkage estimator, is consistent, is shown to be highly competitive with other spectral density estimators via simulations, and is computationally inexpensive.

Keywords: Bayesian inference; Hierarchical model; Model averaging; Laplace method; Smoothing.

1 Introduction

The spectral density of a time series completely captures the dependence structure of the time series. In fact, the covariance structure for any stationary time series $\{X_1, \dots, X_t, \dots\}$ with innovation variance σ^2 can be calculated from its spectral density, $f_X(\omega)$, using the formula

$$\text{cov}(X_t, X_{t+h}) = 2\sigma^2 \int_0^\pi f_X(\omega) \cos(\omega h) d\omega, \quad h = 0, 1, 2, 3, \dots \quad (1)$$

Statistical research in regression and nonparametric function estimation can easily be applied to the estimation of spectral densities. However, estimating the spectral density of a time series has attracted research all of its own. The primary motivation of this research has been the statistical properties of the periodogram, which is an inconsistent estimator of the spectral density. The periodogram is defined on a set of F frequencies, the Fourier frequencies $\left\{ \frac{2\pi j}{n} \right\}_{j=1, \dots, F=\lfloor \frac{n}{2} \rfloor}$, and is calculated as

$$I(\omega_j) = \frac{1}{2\pi} \frac{1}{n} \left[\left\{ \sum_{t=1}^n X_t^* \sin(\omega_j t) \right\}^2 + \left\{ \sum_{t=1}^n X_t^* \cos(\omega_j t) \right\}^2 \right], \quad (2)$$

where $X_t^* = X_t - \bar{X}$, n is the sample size, $\omega_j = \frac{2\pi j}{n}$, and $[a]$ is the largest integer less than or equal to a . A typical correction for the periodogram's lack of consistency is to smooth the periodogram across frequencies, and a large percentage of the research devoted to spectral density estimation focuses on how the periodogram should be smoothed (Blackman & Tukey, 1958; Hall et al., 1994; Ombao et al., 2001). Pawitan and Gangopadhyay (1991) related the spectral density and its second derivative to a smoothed version of the periodogram. Parameter estimates within this linear model gave estimates of the true spectral density.

Bayesian methods may also be used to estimate the spectral density. Shaman (1977) constructed a Bayesian estimator which allows one to incorporate prior information regarding the smoothness and shape of the spectral density. Huerta & West (1999) considered spectral densities of autoregressive processes; they placed priors on the roots of autoregressive polynomials and then sampled

from the corresponding posterior distributions to get estimates of the spectral density. Denison et al. (2002) fitted piecewise polynomials to the log periodogram using reversible jump Markov chain Monte Carlo methods.

A shrinkage estimator of the spectral density, based on a two-level hierarchy that requires one to specify a parametric time series model, was proposed by Daniels & Cressie (2001). Their estimator shrinks toward the parametric estimator of the spectral density. Although the ideas behind this shrinkage estimator are appealing, the asymptotic properties of this estimator have not been explored, it has not been studied via simulations, and it requires the selection of a parametric time series model.

The estimator studied in this paper builds on the ideas in Daniels & Cressie (2001). We derive a consistent shrinkage estimator of the spectral density similar to that of the Daniels-Cressie estimator. We also avoid selection of a particular parametric time series model by averaging over a class of parametric models.

2 The Daniels-Cressie Estimator

Daniels and Cressie constructed a two-level hierarchical model from which their shrinkage estimator was derived. The first level of the hierarchical Daniels-Cressie model describes the random behaviour of the periodogram (2). Asymptotically, the periodogram of any time series has a χ_2^2 distribution, although the last periodogram ordinate has a χ_1^2 distribution if n is even. Since the fourth root of a χ_2^2 random variable is approximately normally distributed (Hawkins & Wixley, 1986),

$$M_j I(\omega_j)^{\frac{1}{4}} \sim N \left\{ f(\omega_j)^{\frac{1}{4}}, V_j f(\omega_j)^{\frac{1}{2}} \right\}, \quad (3)$$

where for n odd, $M_j = 2^{\frac{1}{4}}\Gamma(1.25)$ and $V_j = 2^{\frac{1}{2}}\Gamma(1.5) - M_j^2$, for $j = 1, 2, \dots, F$, where F is the number of Fourier frequencies, and, for n even, $M_j = 2^{\frac{1}{4}}\Gamma(1.25)$ and $V_j = 2^{\frac{1}{2}}\Gamma(1.5) - M_j^2$ for $j = 1, 2, \dots, F - 1$, $M_F = 2^{\frac{1}{4}}\Gamma(0.75)/\Gamma(0.5)$ and $V_F = 2^{\frac{1}{2}}\Gamma(0.5) - M_F^2$. Diggle & al Wasel (1997) used the asymptotic distribution of the periodogram to construct an approximate likelihood, but

without the fourth root transformation.

In the second level of the hierarchical model, a prior was placed on the fourth root of the true spectral density at all of the Fourier frequencies. This prior can be written as

$$f(\omega)^{\frac{1}{4}} = \begin{pmatrix} f(\omega_1)^{\frac{1}{4}} \\ f(\omega_2)^{\frac{1}{4}} \\ \vdots \\ f(\omega_F)^{\frac{1}{4}} \end{pmatrix} \sim TN_0 \left\{ \begin{pmatrix} f_p(\omega_1; \theta)^{\frac{1}{4}} \\ f_p(\omega_2; \theta)^{\frac{1}{4}} \\ \vdots \\ f_p(\omega_F; \theta)^{\frac{1}{4}} \end{pmatrix}, \tau^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix}_{F \times F} \right\}. \quad (4)$$

This is a multivariate normal prior with mean $(f_p(\omega_1; \theta)^{\frac{1}{4}}, \dots, f_p(\omega_F; \theta)^{\frac{1}{4}})'$ and variance matrix $\tau^2 I_F$ truncated below at 0. The function $f_p(\omega; \theta)$ is the spectral density of an ARMA(p, q) time series with unknown parameters θ . Note that θ is a vector which contains all of the parameters that specify a spectral density. For an ARMA(p, q) time series, $\theta = (\psi, \sigma^2)$, where σ^2 is the innovation variance, and $\psi = (\phi, \eta)$, where ϕ are the autoregressive parameters and η are the moving average parameters. Throughout this paper, $f_p(\omega; \theta)$ will be denoted by $f_p(\omega; \psi, \sigma^2)$ when appropriate. The parameter τ^2 measures the uncertainty of the true spectral density around the ARMA model. The motivation behind this hierarchy was to obtain improved estimation in small samples by shrinking towards a parametric form (Chen, 1979; Daniels & Kass, 1999).

With this model, the posterior distribution of the true spectral density at the Fourier frequencies, conditional on τ^2 and θ , is easy to calculate. The mean of the conditional posterior, $E \left\{ f(\omega)^{\frac{1}{4}} \mid \theta, \tau^2, I(\omega) \right\}$, is used as the estimator of the spectral density's fourth root at each of the Fourier frequencies. This mean is

$$E \left\{ f(\omega_j)^{\frac{1}{4}} \mid \hat{\theta}, \hat{\tau}^2, I(\omega) \right\} = \frac{V_j \left\{ M_j I(\omega_j)^{\frac{1}{4}} \right\}^2}{V_j \left\{ M_j I(\omega_j)^{\frac{1}{4}} \right\}^2 + \hat{\tau}^2} f_p(\omega_j; \hat{\theta})^{\frac{1}{4}} + \frac{\hat{\tau}^2}{V_j \left\{ M_j I(\omega_j)^{\frac{1}{4}} \right\}^2 + \hat{\tau}^2} M_j I(\omega_j)^{\frac{1}{4}}, \quad (5)$$

where $\hat{\tau}^2$ was estimated using a method-of-moments estimator, and $\hat{\theta}$ was calculated using maximum likelihood estimation. The estimator given in (5) is a weighted sum of the periodogram and the parametric estimator both on the fourth-root scale. Daniels and Cressie used the fact that the posterior distribution of $f(\omega)^{\frac{1}{4}} \mid \theta, \tau^2, I(\omega)$ was normal when deriving their estimator for the true

spectral density on the original scale, $f(\omega)$. Their estimator for the true spectral density is the fourth moment of a normal random variable with approximate mean $m = E \left\{ f(\omega_j)^{\frac{1}{4}} | \hat{\theta}, \hat{\tau}^2, I(\omega) \right\}$ and variance $\nu^2 = \left[\frac{1}{V_j \left\{ M_j I(\omega_j)^{\frac{1}{4}} \right\}^2} + \frac{1}{\hat{\tau}^2} \right]^{-1}$. The fourth moment of this random variable is

$$\hat{f}_{\text{DC}}(\omega_j) = E \left[\left\{ f(\omega_j)^{\frac{1}{4}} \right\}^4 \middle| \hat{\theta}, \hat{\tau}^2, I(\omega) \right] = 3\nu^4 + 6\nu^2 m^2 + m^4. \quad (6)$$

Although the asymptotic properties of this estimator have not been carefully studied, this estimator is clearly not consistent when the true model, $f_p(\omega; \theta)$, is incorrectly specified. To understand the behaviour of the estimator when $f_p(\omega; \theta)$ is correctly specified, we first examine their estimator of τ^2 , namely

$$\hat{\tau}^2 = \max \left[0, \sum w_j \left\{ M_j I(\omega_j)^{\frac{1}{4}} - f_p(\omega_j; \hat{\theta})^{\frac{1}{4}} \right\}^2 \left\{ \sum w_j - \left(\sum w_j^2 \right) \left(\sum w_j \right)^{-1} \right\}^{-1} \right],$$

where w_j , the weight at frequency ω_j , should ideally be $\left[\text{var} \left\{ M_j I(\omega_j)^{\frac{1}{4}} \right\} \right]^{-1}$. Daniels and Cressie estimated the weights at Fourier frequency ω_j using $V_j \left\{ M_j I(\omega_j)^{\frac{1}{4}} \right\}^2$, which is inconsistent for $V_j f(\omega_j)^{\frac{1}{2}}$. It should be noted, however, that if $w_j \rightarrow \left[\text{var} \left\{ M_j I(\omega_j)^{\frac{1}{4}} \right\} \right]^{-1}$ in probability, and, if $f_p(\omega; \theta)$ is correctly specified, the Daniels-Cressie estimator is consistent. The following theorem is proved in the Appendix.

Theorem 1. *Provided that $w_j \rightarrow \left[\text{var} \left\{ M_j I(\omega_j)^{\frac{1}{4}} \right\} \right]^{-1}$ in probability, \hat{f}_{DC} is consistent for f if the true model, $f_p(\omega; \theta)$, is correctly specified.*

In the next section, we propose a modification to this estimator that is consistent regardless of the choice of $f_p(\omega; \theta)$.

3 The New Estimator

3.1 Definition of the estimator

We modify the methods of Daniels and Cressie in several respects. In the first level of the hierarchy, we minimally smooth the fourth root of the periodogram and adjust the likelihood in (3) accordingly.

Secondly, we add an additional level to the hierarchy by placing a prior distribution on τ^2 . Thirdly, we avoid selection of a particular parametric model for the estimator by model averaging.

We begin by replacing the periodogram with a minimally smoothed nonparametric estimate of the spectral density's fourth root at each of the Fourier frequencies,

$$\hat{f}_s(\omega_j)^{\frac{1}{4}} = \sum_{k=1}^F W_h(\omega_k - \omega_j) M_k I(\omega_k)^{\frac{1}{4}} \quad (7)$$

where $W_h(\omega_k - \omega_j) = K_h(\omega_k - \omega_j) \left\{ \sum_{k=1}^F K_h(\omega_k - \omega_j) \right\}^{-1}$, ω_k is the k th Fourier frequency, and $K_h(\cdot)$ is a normal kernel with bandwidth proportional to $n^{-\frac{7}{8}}$. Note that this is nearly the smallest bandwidth we can choose while maintaining the consistency of our estimator; see Theorem 3. Our motivation for doing this should be clear: smooth the periodogram as little as possible so as not to lose sudden and high peaks. With this smoothing, the likelihood in (3) is replaced with

$$\hat{f}_s(\omega)^{\frac{1}{4}} = \begin{pmatrix} \hat{f}_s(\omega_1)^{\frac{1}{4}} \\ \hat{f}_s(\omega_2)^{\frac{1}{4}} \\ \vdots \\ \hat{f}_s(\omega_F)^{\frac{1}{4}} \end{pmatrix} \sim N \left\{ \begin{pmatrix} f(\omega_1)^{\frac{1}{4}} \\ f(\omega_2)^{\frac{1}{4}} \\ \vdots \\ f(\omega_F)^{\frac{1}{4}} \end{pmatrix}, W V W' \right\},$$

where $W = [W_h(\omega_k - \omega_j)]_{k=1, \dots, F, j=1, \dots, F}$, $V = \text{diag} \left[\text{var} \left\{ \hat{f}_s(\omega_1)^{\frac{1}{4}} \right\}, \text{var} \left\{ \hat{f}_s(\omega_2)^{\frac{1}{4}} \right\}, \dots, \text{var} \left\{ \hat{f}_s(\omega_F)^{\frac{1}{4}} \right\} \right]$, $\text{var} \left\{ \hat{f}_s(\omega_j)^{\frac{1}{4}} \right\} = V_j \hat{f}_s(\omega_j)^{\frac{1}{2}}$, and $\hat{f}_s(\omega_j)^{\frac{1}{2}} = \left\{ \sum_{k=1}^F W_h(\omega_k - \omega_j) M_k I(\omega_k)^{\frac{1}{4}} \right\}^2$. We treat the likelihood above as an actual likelihood. This is similar in spirit to the modified likelihoods constructed by Sun et al. (2000) on pre-whitened residuals. We also smooth the periodogram in our estimate of the variance. This is done to ensure that our estimator of the covariance matrix, $W V W'$, is consistent; see Theorem 3. As an additional benefit, we expect this approximate likelihood to be more accurate than (3) because of the smoothing.

The prior distribution of $f(\omega)^{\frac{1}{4}}$ evaluated at all the Fourier frequencies is given in (4). As in Daniels & Cressie (2001), we replace this truncated distribution with a non-truncated multivariate normal distribution to derive the estimator. We also place a prior distribution $p(\tau^2)$ on τ^2 to exert additional control over the amount of shrinkage towards $f_p(\omega; \theta)$.

With these distributions, the conditional posterior mean of $f(\omega)^{\frac{1}{4}} | \theta, \tau^2, \hat{f}_s(\omega)$ becomes

$$E \left\{ f(\omega)^{\frac{1}{4}} | \theta, \tau^2, \hat{f}_s(\omega) \right\} = \Sigma(\Sigma + \tau^2 I)^{-1} f_p(\omega; \theta)^{\frac{1}{4}} + \tau^2 I(\Sigma + \tau^2 I)^{-1} \hat{f}_s(\omega)^{\frac{1}{4}}, \quad (8)$$

where $\Sigma = W V W'$. Our estimator for the fourth root under squared error loss is then $\tilde{f}(\omega)^{\frac{1}{4}} = E \left\{ f(\omega)^{\frac{1}{4}} | \theta = \hat{\theta}, \tau^2 = \hat{\tau}^2, \hat{f}_s(\omega) \right\}$, where $\hat{\tau}^2$ is the mode of $\log \left[\text{Lik} \left\{ \tau^2 | \theta = \hat{\theta}, I(\omega) \right\} p(\tau^2) \right]$ and $(\hat{\theta})$ is the maximum likelihood estimator of the ARMA(p, q) process fit on original data. The variance of this estimator is given by

$$\begin{aligned} \text{var} \left\{ \tilde{f}(\omega)^{\frac{1}{4}} | \sigma^2, \tau^2 \right\} &= \Sigma (\Sigma + \tau^2 I)^{-1} \text{var} \left\{ f_p(\omega; \hat{\psi}, \sigma^2)^{\frac{1}{4}} \right\} \left\{ \Sigma(\Sigma + \tau^2 I)^{-1} \right\}^T \\ &\quad + \tau^2 I (\Sigma + \tau^2 I)^{-1} \text{var} \left\{ \hat{f}_s(\omega)^{\frac{1}{4}} \right\} \left\{ \tau^2 I(\Sigma + \tau^2 I)^{-1} \right\}^T, \end{aligned}$$

where $\hat{\psi} = (\hat{\phi}, \hat{\eta})$, $\text{var} \left\{ f_p(\omega; \hat{\psi}, \sigma^2)^{\frac{1}{4}} \right\} = R \sigma^2 \left(\begin{array}{cc} E \tilde{U}_t \tilde{U}_t' & E \tilde{U}_t \tilde{V}_t' \\ E \tilde{V}_t \tilde{U}_t' & E \tilde{V}_t \tilde{V}_t' \end{array} \right)^{-1} R^T$,

$$R = \left(\frac{\partial f_p(\omega_1; \theta)^{\frac{1}{4}}}{\partial \psi}, \dots, \frac{\partial f_p(\omega_F; \theta)^{\frac{1}{4}}}{\partial \psi} \right)^T,$$

$\tilde{U}_t = (U_t, U_{t-1}, \dots, U_{t+1-p})'$, $\tilde{V}_t = (V_t, V_{t-1}, \dots, V_{t+1-q})'$, and $\{U_t\}$, $\{V_t\}$ are the processes $U_t = \sum_{i=1}^p \phi_i U_{t-i} + Z_t$ and $V_t + \sum_{i=1}^q \eta_i V_{t-i} = Z_t$, with $Z_t \sim N(0, \sigma^2)$ (Brockwell & Davis, 1990, p. 258). Our estimator of $f(\omega)$, $\tilde{f}(\omega)$, takes a similar form to that of Daniels and Cressie. Rather than use the fourth moment of a univariate normal distribution as our estimator for f , we use the fourth moment of the multivariate normal random vector $f(\omega)^{\frac{1}{4}} | \theta, \tau^2, \hat{f}_s(\omega)$, which is

$$\tilde{f}(\omega) = 3\tilde{\nu}^4 + 6\tilde{\nu}^2 \tilde{m}^2 + \tilde{m}^4, \quad (9)$$

where $\tilde{\nu}$ and \tilde{m} are $F \times 1$ vectors such that $\tilde{\nu} = \text{diag} \left(\Sigma^{-1} + \frac{1}{\hat{\tau}^2} I_F \right)^{-1}$ and $\tilde{m} = E \left\{ f(\omega)^{\frac{1}{4}} | \hat{\theta}, \hat{\tau}^2, \hat{f}_s(\omega) \right\}$.

The variance of this fourth moment can then be computed as

$$\text{var} \left\{ \tilde{f}(\omega) \right\} = 144 \times \text{diag} \left(\tilde{\nu}_1^4 \tilde{m}_1^2, \dots, \tilde{\nu}_F^4 \tilde{m}_F^2 \right) \text{var} \left\{ \tilde{f}(\omega)^{\frac{1}{4}} \right\} + 16 \times \text{diag} \left(\tilde{m}_1^6, \dots, \tilde{m}_F^6 \right) \text{var} \left\{ \tilde{f}(\omega)^{\frac{1}{4}} \right\}.$$

3.2 Choice of $f_p(\omega; \theta)$

The shrinkage estimator derived in the previous section shrinks the nonparametric spectral density estimator, the smoothed periodogram, to some parametric ARMA(p, q) model. However, which

values of p and q should be selected? This has not been addressed in other work with this types of shrinkage prior for dependence (Chen, 1979; Daniels & Kass, 1999, 2001; Daniels & Pourahmadi, 2002). To avoid having to answer this question, we propose an estimator which averages over a class of parametric models.

The estimator we propose averages over a class of ARMA(p, q) models, where $p \in \{0, 1, 2, 3\}$ and $q \in \{0, 1, 2, 3\}$. We limit the number of autoregressive and moving average parameters to three because we would like to study our estimator in small samples. However, any class of parsimonious models would do. This model averaged shrinkage estimator can be written as

$$\tilde{f}_{\text{avg}}(\omega) = \sum_{\text{all } p, q} p \left\{ p, q | \hat{f}_s(\omega) \right\} \tilde{f}_{p, q}(\omega), \quad (10)$$

where $\tilde{f}_{p, q}(\omega)$ denotes the shrinkage estimator derived from a model with p autoregressive and q moving average parameters, and $p \left\{ p, q | \hat{f}_s(\omega) \right\}$ denotes the corresponding posterior probability. With the model averaged shrinkage estimator written as in (10), it is clear that averaging in this way adds an additional level of smoothing to the estimator given in (9). For the terms in (10), it should be noted that

$$p \left\{ p, q | \hat{f}_s(\omega) \right\} = p \left\{ \hat{f}_s(\omega) | p, q \right\} p(p, q) \left[\sum_{p, q} p \left\{ \hat{f}_s(\omega) | p, q \right\} p(p, q) \right]^{-1}, \quad (11)$$

where

$$p \left\{ \hat{f}_s(\omega) | p, q \right\} = \int_{\mathbb{R}^{p+q}} \int_0^\infty \int_0^\infty \text{Lik}(\psi, \sigma^2, \tau^2) p(\psi) p(\sigma^2) p(\tau^2) d\sigma^2 d\tau^2 d\psi,$$

$p(p, q)$ is a prior placed on p and q , and

$$\text{Lik}(\psi, \sigma^2, \tau^2) = \left\{ \frac{1}{(2\pi)^F |\tau^2 I + \Sigma|} \right\}^{\frac{1}{2}} \exp \left[-\frac{1}{2} \left\{ \hat{f}_s(\omega)^{\frac{1}{4}} - f_p(\omega; \psi, \sigma^2)^{\frac{1}{4}} \right\}' (\tau^2 I + \Sigma)^{-1} \left\{ \hat{f}_s(\omega)^{\frac{1}{4}} - f_p(\omega; \psi, \sigma^2)^{\frac{1}{4}} \right\} \right].$$

The marginal likelihood of p and q is calculated using a Laplace approximation. To be specific, we

let

$$p \left\{ \hat{f}_s(\omega) | p, q \right\} = \text{Lik}(\hat{\psi}, \hat{\sigma}^2, \hat{\tau}^2) p(\hat{\psi}) p(\hat{\tau}^2) p(\hat{\sigma}^2) (2\pi)^{\frac{p+q}{2}} \left| - \left[\frac{\partial^2 \{ \log \text{Lik}(\psi, \sigma^2, \tau^2) p(\psi, \sigma^2, \tau^2) \}}{\partial(\psi, \sigma^2, \tau^2)^T \partial(\psi, \sigma^2, \tau^2)} \right]^{-1} \right|_{\psi=\hat{\psi}, \sigma^2=\hat{\sigma}^2, \tau^2=\hat{\tau}^2}^{\frac{1}{2}},$$

where $p(\psi, \sigma^2, \tau^2) = p(\psi) p(\sigma^2) p(\tau^2)$, $\hat{\psi}$ is the maximum likelihood estimates of the ARMA param-

eters, $\hat{\sigma}^2$ is the maximum likelihood estimator of the innovation variance, and $\hat{\tau}^2$ is the mode of the posterior density $p\{\tau^2|\psi = \hat{\psi}, \sigma^2 = \hat{\sigma}^2, I(\omega)\}$.

A variety of priors can be placed on the number of autoregressive and moving average parameters, p and q . The prior we consider places a majority of its weight on the lower-order models. In fact, we let $p(p, q) = p(p)p(q)$, where $p(p)$ and $p(q)$ are both truncated Poisson distributions with $\lambda = 1.5$. Other priors that can be placed may be flat, or may place little weight on small values of p and q , favouring more complex models.

3.3 Priors on ψ , σ^2 and τ^2

In the above model, flat priors are implicit on ψ and on the innovation variance σ^2 ; that is,

$$p(\psi) \propto \mathbb{I}\{\psi \in \mathbb{R}^{p+q}\}, \quad p(\sigma^2) \propto \mathbb{I}\{\sigma^2 \in (0, \infty)\}.$$

We consider two priors for τ^2 : $p(\tau^2) \propto (\tau^2)^{-1}$, which, of course, has more mass towards 0, and $p(\tau^2) \propto 1$, which, compared to the prior given above, should shrink less towards the selected parametric form.

Although we use an empirical Bayesian approach to compute our estimator, a fully Bayesian analysis can be performed. We show that, when we shrink towards an ARMA($p, 0$) model, if $p(\psi)$ is proper and $p(\sigma^2)$ and $p(\tau^2)$ are flat on the positive real line, then $p\{\psi, \sigma^2, \tau^2|I(\omega)\}$ is proper, and a fully Bayesian analysis can be done. If $p(\tau^2) \propto (\tau^2)^{-1}$, however, $p\{\psi, \sigma^2, \tau^2|I(\omega)\}$ is improper.

Theorem 2. *Consider an ARMA($p, 0$) model where the autoregressive parameters are denoted by $\phi = \{\phi_1, \dots, \phi_p\}$. If both $p(\sigma^2)$ and $p(\tau^2)$ are flat on the positive real line, and the two conditions below are met, $p\{\phi, \sigma^2, \tau^2|I(\omega)\}$ is proper.*

Condition 1. $\int_{\mathbb{R}^p} p(\phi) d\phi < \infty$

Condition 2. $\int_{\mathbb{R}^p} |\phi_j \phi_k| p(\phi) d\phi < \infty$ for all pairs j, k .

The proof is given in the Appendix. Theorem 2 provides justification for doing a fully Bayesian analysis with these priors. It should be noted that, if independent normal priors are placed on the

autoregressive parameters, the two conditions listed above are satisfied. For an ARMA(p, q) model with $q \geq 1$, propriety of $p \{ \psi, \sigma^2, \tau^2 | I(\omega) \}$ remains unresolved.

3.4 Consistency and Evaluation of the Estimator at any Set of Frequencies

As developed so far, we obtain an estimator for the spectral density at only the Fourier frequencies. However, it is straightforward to compute an estimate of the spectral density at any set of frequencies. Suppose we choose a set of K frequencies on the interval $(0, \pi)$. Calling the new set of frequencies $\omega_I = \{\omega_{I_1}, \dots, \omega_{I_K}\}$, one can write an estimator for $f(\omega)^{\frac{1}{4}}$ at ω_I as

$$E \left\{ f(\omega_I)^{\frac{1}{4}} | \hat{\theta}, \hat{\tau}^2, \hat{f}_s(\omega) \right\} = \hat{\Sigma}_I (\hat{\Sigma}_I + \hat{\tau}^2 I)^{-1} f_p(\omega_I; \hat{\theta})^{\frac{1}{4}} + \hat{\tau}^2 I (\hat{\Sigma}_I + \hat{\tau}^2 I)^{-1} \hat{f}_s(\omega_I)^{\frac{1}{4}},$$

where $\hat{\Sigma}_I$ is a $K \times K$ symmetric matrix with $\sum_{k=1}^F W_h(\omega_k - \omega_{I_t}) W_h(\omega_k - \omega_{I_l}) V_k \hat{f}_s(\omega_k)^{\frac{1}{2}}$ as its (t, l) th entry, $\hat{f}_s(\omega_I)^{\frac{1}{4}}$ is a $K \times 1$ vector with $\sum_{k=1}^F W_h(\omega_k - \omega_{I_l}) M_k I(\omega_k)^{\frac{1}{4}}$ as its l th entry, and $\hat{\tau}^2$ is that value estimated from the Fourier frequencies; recall that $W_h(\omega_k - \omega_{I_l}) = K_h(\omega_k - \omega_{I_l}) \left\{ \sum_j K_h(\omega_j - \omega_{I_l}) \right\}^{-1}$. The estimator for the true spectral density then becomes $\tilde{f}(\omega_I)$, where

$$\tilde{f}(\omega_I) = 3\tilde{\nu}_I^4 + 6\tilde{\nu}_I^2 \tilde{m}_I^2 + \tilde{m}_I^4, \quad (12)$$

$\tilde{\nu}_I = \text{diag} \left(\hat{\Sigma}_I^{-1} + \frac{1}{\hat{\tau}^2} I_{K \times K} \right)^{-1}$, $\tilde{m}_I = E \left\{ f(\omega_I)^{\frac{1}{4}} | \hat{\theta}, \hat{\tau}^2, \hat{f}_s(\omega) \right\}$, and $\hat{\tau}^2$ and $\hat{\theta}$ are, again, the maximum likelihood estimator calculated from the original data, i.e. the ‘‘smoothed’’ periodogram at the Fourier frequencies. The estimator $\tilde{f}(\omega_I)$ is consistent for the true spectral density at any set of frequencies, ω_I . The following theorem is proved in the Appendix.

Theorem 3. *Let ω_I be a set of K frequencies, not necessarily Fourier frequencies, on the interval $(0, \pi)$. This set can be written as $\omega_I = \{\omega_{I_1}, \dots, \omega_{I_K}\}$. We make the following assumptions:*

- (i) $f(\omega)$ is bounded, and $f''(\omega)$ is continuous on the interval $[0, \pi]$;
- (ii) the kernel, K_h , is symmetric about 0 and supported on $[-\pi, \pi]$;
- (iii) the bandwidth, $h = h_n$, is a sequence satisfying $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$;

(iv) $h\pi < \omega_{I_1} < \omega_{I_K} < \pi - \pi h$ for all n after a particular value of n_0 .

Then, regardless of the parametric structure specified, $\tilde{f}(\omega_I)$, given in (12), is consistent for $f(\omega_I)$.

4 Simulations and Computations

4.1 Description of the simulation study

To explore the behaviour of the estimator in small to medium sample sizes, we simulated 500 realisations from eight different time series models at three different sample sizes, $n = 32, 64$ and 128 . For each realisation, the model averaged shrinkage estimator and six other spectral density estimators were calculated. The mean integrated squared error, $MISE$, and the mean maximum squared deviation, $MMSD$, of each of these estimators are calculated, where

$$MISE = \frac{1}{n_s} \sum_{k=1}^{n_s} \left[\frac{1}{n_K} \sum_{i=1}^{n_K} \left\{ \hat{f}(\omega_i) - f(\omega_i) \right\}^2 \right]_k \quad \text{and} \quad MMSD = \frac{1}{n_s} \sum_{k=1}^{n_s} \left(\max_{\omega_i \in \omega_I} \left[\left\{ \hat{f}(\omega_i) - f(\omega_i) \right\}^2 \right] \right),$$

in which n_s is the number of simulations performed, n_K is the number of frequencies at which the estimator is evaluated, and ω_I is the set of frequencies at which the estimator is evaluated. The quantity $MISE$ measures how each estimator performs in estimating an entire spectral density, while $MMSD$ measures the largest deviation and is introduced to see how each estimator captures peaks.

Of the eight simulated time series, five were true stationary ARMA processes and three were not. The time series were all $n \times 1$ zero-mean multivariate normal random vectors with covariances derived from a given spectral density. The data were simulated using the `fracdiff.sim` procedure in R. Figure 1 shows the spectral densities of the eight simulated time series. The first five of these time series are true ARMA models. The general model is

$$X_t - \sum_{j=1}^p \phi_j X_{t-j} = Z_t + \sum_{i=1}^q \eta_i Z_{t-i},$$

where $Z_t \sim N(0, 1)$, independently for each t , and the parameters used are listed in Table 1. The last three time series simulated from were not ARMA processes. Their spectral densities are as follows.

Spectral Density 6: $f(\omega) = |\cos(1.5\omega)|$

Spectral Density 7: $f(\omega) = .5(\omega - 1.2)^2 + .5 |\cos(6\omega) + \sin(.2\omega)|$

Spectral Density 8: $f(\omega) = N(\omega; 0.25, 0.11) + N(\omega; 0.5, 0.1) + N(\omega; 1, 9)$

where $N(\omega; \mu, \sigma^2)$ denotes the $N(\mu, \sigma^2)$ density function.

In calculating the model averaged shrinkage estimator, a total of 16 parametric models were fitted to the data, namely $ARMA(p, q)$ for $p \in \{0, 1, 2, 3\}$ and $q \in \{0, 1, 2, 3\}$. The maximum likelihood estimates for all models were calculated using the `arima0` function in the `tseries` package of R. We compared the new estimators to the following existing spectral density estimators: the periodogram; an adaptive bandwidth smoother which was calculated using the `glkerns` function in R (Gasser et al., 1986, 1991; Hermann, 1997); the maximum likelihood estimator under each of the ARMA models; a wavelet estimator (Percival & Walden, 2000); and the Daniels-Cressie estimator given in equation (6) and the shrinkage estimator given in equation (12), which correspond to the correct parametric model. The adaptive bandwidth smoother, which is calculated through the `glkerns` function, is a kernel regression estimate of the spectral density of the form

$$\hat{f}_{AB}(\omega; b) = \sum_{i=1}^F w \left(\frac{\omega_i - \omega}{b} \right) I(\omega_i),$$

where b is the bandwidth and $w(\cdot)$ is a polynomial kernel. In this case, the bandwidth is selected to minimise the asymptotic mean integrated squared error $E \left[\int_0^\pi \left\{ f(\omega) - \hat{f}_{AB}(\omega; b) \right\}^2 d\omega \right]$, with the heteroskedasticity of the periodogram taken into account. The wavelet estimator is calculated using the methods outlined in §6 of Chapter 10 of Percival and Walden (2000), and is computed using the functions available in the R package `wavethresh`. For $F = 16, 32, 64$, the values of J_o were set to 4, 5 and 6, respectively; alternative values of J_o either could not be calculated using the functions available, or returned estimates with higher values of *MISE* and *MMSD*. The wavelet estimator was then evaluated at the frequencies of interest using the `approx` function in R.

4.2 Results

The values of $MISE$ and $MMSD$ of all the estimators considered at $n = 64$ are shown in Tables 2-3. In Figure 2, true spectral densities 3, 4 and 8 are plotted with the model averaged shrinkage estimator, Wavelet, and Adaptive estimators averaged over five randomly selected samples with $n = 128$. Figure 2 reveals how the model averaged shrinkage estimator compares in capturing the peaks of a spectral density. Throughout this discussion, the maximum likelihood, Shrinkage, and Daniels-Cressie estimators will be referred to as model-dependent estimators, and the others, as model-independent estimators. On average, the model averaged shrinkage estimator outperforms the model independent estimators. Although it is inferior in a few of the simulated cases shown, it is never consistently inferior to the same estimator.

At $n = 64$, the model averaged shrinkage estimator appears to be the best at estimating an entire spectral density and capturing its peaks. With the exception of spectral densities 4 and 8, the values of $MISE$ and $MMSD$ of the model averaged shrinkage estimator are smaller than those of all the other model-independent estimators. For these two spectral densities, either the Adaptive or Wavelet estimator, but never both, has a smaller $MISE$ than the model averaged shrinkage estimator. At spectral density 4, both the Wavelet and Adaptive estimators have a smaller $MMSD$ than the model averaged shrinkage estimator. The Maximum Likelihood and the Shrinkage estimators occasionally outperform the model averaged shrinkage estimator, but this happens because the model being fitted is the correct structure, which, in practice, is unknown. At sample sizes 32 and 128, the results are similar.

Figure 2 shows that the model averaged shrinkage estimator performs just as well as, if not better than, the Wavelet and Adaptive estimators at capturing peaks. Of the eight spectral densities considered, spectral densities 3,4 and 8 have the most sudden and sharp peaks. They are also the only spectral densities for which the Adaptive and Wavelet estimators are competitive with the model average shrinkage estimator; note that the Wavelet estimator performs poorly on the smoother spectral densities. For spectral density 4, all three estimators behave equivalently. For

spectral densities 3 and 8, the Wavelet estimator tends to overestimate the height of each peak, and the Adaptive estimator underestimates the height of each peak, whereas the model averaged shrinkage estimator captures the magnitude of each peak correctly.

5 Example

We used the model averaged shrinkage, adaptive and wavelet estimators to estimate the spectral density of a few randomly selected patients involved in a medical experiment that was designed to relate clinical depression to the frequency at which luteinising hormone is released in the blood stream. The blood concentration of luteinising hormone fluctuates with time, and Grambsch et al. (2002) hypothesised that the frequency at which this concentration fluctuates differs between those women who are clinically depressed and those who are not. The sample taken in this study includes 26 women who were diagnosed as clinically depressed and 24 women who were not. Blood was taken every 10 minutes for 8 hours from each woman, and the concentration of luteinising hormone was recorded at each time, giving 49 observations in time for each subject.

We estimated the spectral density of nine randomly selected subjects using the first 32 observations in time (this is the largest sample size which allows the Wavelet estimator to be calculated). In these nine cases, we used each spectral density estimator to predict the next five observations. The estimators corresponding to one of these subjects is shown in Figure 3. Recall that in this figure, the estimates cannot be evaluated on how well they estimate the true spectral density, since this is not known. The mean predicted squared errors of the model averaged shrinkage, Adaptive, and Wavelet estimators and the periodogram are, respectively, 0.70, 1.97, 0.74, and 1.05.

6 Discussion

The concepts and models behind this shrinkage estimator may be suited to a variety of other problems involving spectral density estimation. In longitudinal studies, for example, time series are observed for multiple subjects, and a common periodic behavior within each series may be of interest, as in the example in §5 (Grambsch et al., 2002). Hierarchical models, similar to the one

proposed in this paper, might be adapted to such a setting by shrinking the subject-specific spectral densities to a population spectral density. The authors are currently working on such models. Similar models have previously been explored in Diggle & al Wasel (1997). The first author hopes to make a function in R publicly available for computing the model averaged shrinkage estimator.

Acknowledgements

Funding for M.J. Daniels was provided by grants from the National Institutes of Health. We thank the reviewers for helpful comments and suggestions and Professor Grambsch for providing the luteinising hormone data. The research was done as part of Carsten Botts's PhD dissertation at Iowa State University.

Appendix

Proof of Theorems

Proof of Theorem 1. Let $w_j = \text{vâr} \left\{ M_j I(\omega_j)^{\frac{1}{4}} \right\} = V_j \left\{ \hat{f}_s(\omega_j)^{\frac{1}{4}} \right\}^2 \longrightarrow V_j f(\omega_j)^{\frac{1}{2}} = \text{var} \left\{ M_j I(\omega_j)^{\frac{1}{4}} \right\}$, in probability. We begin by showing that $\hat{\tau}^2 \longrightarrow 0$ in probability when the model is correctly specified. To prove this, we consider the function

$$h \{I(\omega_1), \dots, I(\omega_F)\} = \frac{\sum_{j=1}^F \left[V_j \left\{ \hat{f}_s(\omega_j)^{\frac{1}{4}} \right\}^2 \right]^{-1} \left\{ M_j I(\omega_j)^{\frac{1}{4}} - f_p(\omega_j; \theta)^{\frac{1}{4}} \right\}^2 - F}{\sum_{j=1}^F \left[V_j \left\{ \hat{f}_s(\omega_j)^{\frac{1}{4}} \right\}^2 \right]^{-1} - \frac{\left(\sum_{j=1}^F \left[V_j \left\{ \hat{f}_s(\omega_j)^{\frac{1}{4}} \right\}^2 \right]^{-2} \right)}{\left(\sum_{j=1}^F \left[V_j \left\{ \hat{f}_s(\omega_j)^{\frac{1}{4}} \right\}^2 \right]^{-1} \right)}} = \frac{\sum_{j=1}^F A_j - F}{D},$$

where $A_j = \left[V_j \left\{ \hat{f}_s(\omega_j)^{\frac{1}{4}} \right\}^2 \right]^{-1} \left\{ M_j I(\omega_j)^{\frac{1}{4}} - f_p(\omega_j; \theta)^{\frac{1}{4}} \right\}^2$, and D is the denominator in $h \{ \cdot \}$. We show that $h \{I(\omega_1), I(\omega_2), \dots, I(\omega_F)\} \longrightarrow 0$ in probability when the model is correctly specified. This is done by studying the limiting behaviour of $\frac{1}{F} \sum_{j=1}^F A_j$. As long as $\frac{1}{F} \sum_j A_j$ goes to 1 in probability, h should converge to 0 in probability. First consider the array $\{A_j\}$ for $1 \leq j \leq F = \lfloor \frac{n}{2} \rfloor$. Since $V_j \left\{ \hat{f}_s(\omega_j)^{\frac{1}{4}} \right\}^2 \longrightarrow V_j f(\omega_j)^{\frac{1}{2}}$ in probability, it is clear that $E(A_j) \longrightarrow$

$\left[\text{var} \left\{ M_j I(\omega_j)^{\frac{1}{4}} \right\} \right] / \left\{ V_j f(\omega_j)^{\frac{1}{2}} \right\} = 1$. Now consider the variance of the terms in the sequence.

$$\begin{aligned} \text{var}(A_j) &\simeq \frac{1}{\left\{ V_j f(\omega_j)^{\frac{1}{2}} \right\}^2} \text{var} \left[\left\{ M_j I(\omega_j)^{\frac{1}{4}} - f_p(\omega_j; \theta) \right\}^2 \right] \\ &< \frac{1}{\min_j (V_j)^2} \left\{ \max_j (2M_j^4) + \max_j (V_j) + 1 \right\}. \end{aligned}$$

This inequality follows after evaluation of $\text{var} \left[\left\{ M_j I(\omega_j)^{\frac{1}{4}} - f_p(\omega_j; \theta) \right\}^2 \right]$ and substitution of $f(\omega_j)$ for $f_p(\omega_j; \theta)$; remember that the correct parametric form has been specified. From this inequality, one obtains

$$\frac{1}{F^2} \sum_{j=1}^F \text{var}(A_j) < \frac{1}{F} \left[\frac{1}{\min_j (V_j)^2} \left\{ \max_j (2M_j^4) + \max_j (V_j) + 1 \right\} \right] \rightarrow 0.$$

From Theorem 5.4 in Durrett (1996), it is clear that $\frac{1}{F} \sum_{j=1}^F A_j \rightarrow 1$ in probability. This implies that the numerator in the function $h(\cdot)$ goes to 0 in probability. With this result, it follows that $\hat{\tau}^2 \rightarrow 0$ in probability when the correct parametric form is specified.

The limiting behaviour of the Daniels-Cressie estimator must now be examined, given that $\hat{\tau}^2 \rightarrow 0$ in probability when the model is correctly specified. The estimate of the spectral density's fourth root is given by m , and it converges to the truth in this case. The estimated fourth moment of the spectral density's fourth root thus also converges to the truth, making the Daniels-Cressie estimator consistent. \square

Proof of Theorem 2. This proof uses Results 1 and 2 given below. Proofs of these results can be found in the web appendix at www.stat.ufl.edu/~mdaniels/research.html.

Result 1. *If $Y \sim TN_0(\mu_Y, \sigma_Y^2)$ with $\mu_Y > 0$, then $E(Y^3) \leq 6\mu_Y\sigma_Y^2 + 2\mu_Y^3 + 8\sigma_Y^3 + 14\mu_Y^2\sigma_Y$.*

Result 2. *Let Y be a $K \times 1$ random vector such that $Y \sim N(\mu, \Lambda)$ and let $\mu \sim TN_0(\beta, \Gamma)$, a multivariate normal distribution with mean β and covariance matrix Γ truncated at 0. Then $Y|\Lambda, \beta, \Gamma \sim p_Y(y)$ where*

$$p_Y(y) = \frac{c(\beta, \Gamma)}{(2\pi)^{\frac{K}{2}} |\Lambda + \Gamma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (y - \beta)' (\Lambda + \Gamma)^{-1} (y - \beta) \right\} \text{pr}(W \geq 0),$$

$c(\beta, \Gamma) = \{\text{pr}(\tilde{\mu} \geq 0)\}^{-1}$, $\tilde{\mu} \sim N(\beta, \Lambda)$ and

$$W \sim N \left\{ \Lambda (\Lambda + \Gamma)^{-1} \beta + \Gamma (\Lambda + \Gamma)^{-1} Y, (\Lambda^{-1} + \Gamma^{-1})^{-1} \right\}.$$

To begin the proof of Theorem 2, first observe that, by Result 2,

$$\begin{aligned} p \left\{ I(\omega)^{\frac{1}{4}} \mid \phi, \sigma^2, \tau^2 \right\} &= \frac{c \left\{ f_p(\omega)^{\frac{1}{4}}, \tau^2 \right\}}{(2\pi)^{\frac{F}{2}} |\tau^2 I + \Sigma|^{\frac{1}{2}}} \\ &\times \exp \left[-\frac{1}{2} \left\{ I(\omega)^{\frac{1}{4}} - (\sigma^2)^{\frac{1}{4}} g(\omega; \phi) \right\}' (\tau^2 I + \Sigma)^{-1} \left\{ I(\omega)^{\frac{1}{4}} - (\sigma^2)^{\frac{1}{4}} g(\omega; \phi) \right\} \right] \text{pr}(T \geq 0) \end{aligned}$$

where $I(\omega)^{\frac{1}{4}} = \left(I(\omega_1)^{\frac{1}{4}}, \dots, I(\omega_F)^{\frac{1}{4}} \right)'$, $g(\omega; \phi) = \sigma^{-\frac{1}{2}} \left(f(\omega_1; \phi)^{\frac{1}{4}}, \dots, f(\omega_F; \phi)^{\frac{1}{4}} \right)'$, ω_j is the j th Fourier frequency, and $T \sim N \left\{ \Sigma (\Sigma + \tau^2 I)^{-1} f_p(\omega)^{\frac{1}{4}} + \tau^2 (\Sigma + \tau^2 I)^{-1} I(\omega)^{\frac{1}{4}}, (\Sigma^{-1} + \frac{1}{\tau^2} I)^{-1} \right\}$.

Since $c(f_p, \tau^2) \leq (0.5)^{-F}$ and $\text{pr}(T \geq 0) \leq 1$, it is clear that

$$\begin{aligned} &\int_{\mathbb{R}^p} \int_0^\infty \int_0^\infty p \left\{ I(\omega)^{\frac{1}{4}} \mid \phi, \sigma^2, \tau^2 \right\} p(\phi) d\sigma^2 d\tau^2 d\phi \\ &\leq \int_{\mathbb{R}^p} \int_0^\infty \int_0^\infty \frac{(0.5)^{-F}}{(2\pi)^{\frac{F}{2}} \left\{ \prod_{i=1}^F (\tau^2 + \lambda_i) \right\}^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \left(b - \sigma^{\frac{1}{2}} c \right)' \Lambda^{-1} \left(b - \sigma^{\frac{1}{2}} c \right) \right\} p(\phi) d\sigma^2 d\tau^2 d\phi \\ &= \mathfrak{S}, \end{aligned}$$

say, where λ_i is the i th eigenvalue of Σ , $\Lambda = \text{diag}(\sigma^2 + \lambda_1, \dots, \sigma^2 + \lambda_F)$, $c = P^T g(\omega; \phi)$ where $P = [e_1, \dots, e_F]$ and e_j is the j th eigenvector of Σ , and $b = P^T I(\omega)^{\frac{1}{4}}$. Since Σ is a nonsingular positive definite matrix, $\lambda_i > 0$ for all i . Let $\lambda_{(F)}$ denote the smallest eigenvalue of Σ . It then follows that

$$\begin{aligned} \mathfrak{S} &\leq \int_{\mathbb{R}^p} \int_0^\infty \int_0^\infty \prod_{i=1}^F \frac{.5^{-1}}{(2\pi)^{\frac{1}{2}} \left\{ \tau^2 + \lambda_{(F)} \right\}^{\frac{1}{2}}} \exp \left\{ -\frac{c_i^2}{2(\tau^2 + \lambda_i)} \left(\frac{b_i}{c_i} - \sigma^{\frac{1}{2}} \right)^2 \right\} p(\phi) d\sigma^2 d\tau^2 d\phi \\ &\leq \int_{\mathbb{R}^p} \int_0^\infty .5^{-F} \frac{4(\tau^2 + \lambda_1)^{\frac{1}{2}} (2\pi)^{\frac{1}{2}}}{(2\pi)^{\frac{F}{2}} (\tau^2 + \lambda_{(F)})^{\frac{F}{2}} c_1} \\ &\quad \times \underbrace{\left[\int_0^\infty \frac{\sigma^{\frac{3}{2}}}{(2\pi)^{\frac{1}{2}} \left\{ \frac{(\tau^2 + \lambda_1)^{\frac{1}{2}}}{c_1} \right\}} \exp \left\{ -\frac{c_1^2}{2(\tau^2 + \lambda_1)} \left(\frac{b_1}{c_1} - \sigma^{\frac{1}{2}} \right)^2 \right\} d\sigma^{\frac{1}{2}} \right]}_{=6\left(\frac{b_1}{c_1}\right)\left(\frac{\tau^2 + \lambda_1}{c_1^2}\right) + 2\left(\frac{b_1^2}{c_1^2}\right) + 8\left\{\frac{(\tau^2 + \lambda_1)^{\frac{3}{2}}}{c_1^2}\right\} + 14\left(\frac{b_1^2}{c_1^2}\right)\left(\frac{\tau^2 + \lambda_1}{c_1}\right)} p(\phi) d\tau^2 d\phi \end{aligned}$$

The equality under the underbrace follows from Result 1. In the above integral, the only quantity depending on ϕ is c_1 . To remind the reader of this dependence, we will denote c_1 as $c_1(\phi)$. The above integral is finite provided $F \geq 5$, $\int_{\mathbb{R}^p} c_1(\phi)^{-4} p(\phi) d\phi < \infty$, $\int_{\mathbb{R}^p} c_1(\phi)^{-3} p(\phi) d\phi < \infty$, and $\int_{\mathbb{R}^p} c_1(\phi)^{-2.5} p(\phi) d\phi < \infty$. The last three inequalities hold when an ARMA($p,0$) model is specified; see the web appendix for details. \square

Proof of Theorem 3. In proving this theorem, we use six different results (Results 2-7). Results 3-7 are given below; it should be clear that ω_j refers to the j th Fourier frequency and ω_{I_l} refers to the l th frequency in the set ω_I , and that we are letting $p \left\{ f^{\frac{1}{4}}(\omega) \middle| f_p(\omega)^{\frac{1}{4}}, \tau^2 \right\}$ be the density in (4). The reader should also note that proofs to Results 6 and 7 are given in the web appendix.

Result 3. $f_p(\omega_I; \hat{\theta})^{\frac{1}{4}} \rightarrow f(\omega_I; \theta_0)^{\frac{1}{4}}$ in probability for some value θ_0 . When the model is correctly specified, $\theta_0 = \theta$. (Dalhaus & Wefelmeyer, 1996)

Result 4. $\hat{f}_s(\omega_{I_l})^{\frac{1}{4}} = \sum_{j=1}^F W_h(\omega_j - \omega_{I_l}) M_j I(\omega_j)^{\frac{1}{4}} \rightarrow f(\omega_{I_l})^{\frac{1}{4}}$ in probability where ω_{I_l} is the l th frequency in the set ω_I , (Wand & Jones, 1995)

Result 5. $\hat{\Sigma}_I = \text{v\hat{a}r} \left\{ \hat{f}_s(\omega_I)^{\frac{1}{4}} \right\} \rightarrow [0]_{K \times K}$ in probability

Proof of Result 5. Just consider the l th diagonal element of $\hat{\Sigma}_I$, $\text{v\hat{a}r} \left\{ \hat{f}_s(\omega_{I_l})^{\frac{1}{4}} \right\}$. Remember that $\text{v\hat{a}r} \left\{ \hat{f}_s(\omega_{I_l})^{\frac{1}{4}} \right\} = \sum_{j=1}^F W_h^2(\omega_j - \omega_{I_l}) V_j \left\{ \hat{f}_s(\omega_j)^{\frac{1}{4}} \right\}^2$. Result 4 shows that $\hat{f}_s(\omega_j)^{\frac{1}{4}} \rightarrow f(\omega_j)^{\frac{1}{4}}$ in probability. This implies that, for any arbitrarily small numbers $\delta_1 > 0$ and $\delta_2 > 0$, there exists a sample size n^{**} such that

$$\text{pr} \left[\sum_{j=1}^F W_h^2(\omega_j - \omega_{I_l}) V_j \left\{ \hat{f}_s(\omega_j)^{\frac{1}{4}} \right\}^2 < V_j \left\{ \sup_{\omega \in [0, \pi]} f(\omega)^{\frac{1}{2}} + \delta_1 \right\} \sum_{j=1}^F W_h^2(\omega_j - \omega_{I_l}) \right] > 1 - \delta_2$$

for all $n > n^{**}$. Since $\sum W_h^2(\omega_j - \omega_{I_l})$ can be made arbitrarily small with a sufficiently large sample size \tilde{n} , it is clear that, for a given δ_1 and δ_2 and for any $n > \max(n^{**}, \tilde{n})$, $\text{pr} \left[\text{v\hat{a}r} \left\{ \hat{f}_s(\omega_{I_l})^{\frac{1}{4}} \right\} < \epsilon \right] < 1 - \delta_2$.

Result 6. $\hat{\tau}^2 \rightarrow 0$ in probability when the correct parametric model is specified.

Result 7. $\hat{\tau}^2$ is bounded away from 0 when the model is incorrectly specified.

To prove that $\tilde{f}(\omega_I) = 3\tilde{\nu}_I^4 + 6\tilde{\nu}_I^2\tilde{m}_I^2 + \tilde{m}_I^4 \rightarrow f(\omega_I)$ in probability, the limiting behaviour of both $\tilde{\nu}_I^2$ and \tilde{m}_I have to be studied. First observe that $\tilde{\nu}_I = \text{diag}\left(\hat{\Sigma}_I^{-1} + \frac{1}{\hat{\tau}^2}I_{K \times K}\right)^{-1} \rightarrow 0_{K \times 1}$ since $\hat{\Sigma}_I \rightarrow [0]_{K \times K}$ in probability. Secondly, note that regardless of whether the parametric model is specified correctly or incorrectly,

$$\tilde{m}_I = \hat{\Sigma}_I \left(\hat{\tau}^2 I_{K \times K} + \hat{\Sigma}_I \right)^{-1} f_p(\omega_I; \hat{\theta})^{\frac{1}{4}} + \hat{\tau}^2 \left(\hat{\tau}^2 I_{K \times K} + \hat{\Sigma}_I \right)^{-1} \hat{f}_s(\omega_I)^{\frac{1}{4}} \rightarrow f(\omega)^{\frac{1}{4}}$$

in probability. The limiting behaviour of $\hat{\Sigma}_I$ and $\hat{\tau}^2$ ensures this. When the model is incorrectly specified, for instance, the first term written above disappears as $\hat{\Sigma}_I \rightarrow [0]_{K \times K}$ in probability, and the second dominates. Since $\hat{f}_s(\omega_I)^{\frac{1}{4}} \rightarrow f(\omega_I)^{\frac{1}{4}}$ in probability, $\tilde{m}_I \rightarrow f(\omega_I)^{\frac{1}{4}}$ in probability when the model is incorrectly specified. When the model is correctly specified, \tilde{m}_I is just a linear combination of two estimators that are consistent for $f(\omega_I)^{\frac{1}{4}}$. Given these observations, it is clear that $\tilde{f}(\omega_I) \rightarrow f(\omega_I)$ in probability \square

References

- Blackman, R. B. & Tukey, J. (1958). *The Measurement of Power Spectra from the Point of View of Communication Engineering*. New York: Dover.
- Brockwell, P.J. & Davis, R.A. (1990). *Time Series: Theory and Methods*. New York: Springer-Verlag.
- Chen, C. (1979). Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis. *J. R. Statist. Soc. B* **41**, 235-8.
- Dahlhaus, R. & Wefelmeyer, W. (1996). Asymptotically optimal estimation in misspecified time series models. *Ann. Statist.* **24**, 952-74.
- Daniels, M.J. & Cressie, N. (2001). A hierarchical approach to covariance function estimation for time series. *J. Time Ser. Anal.* **22**, 253-66.

- Daniels, M. & Kass, R. (1999). Nonconjugate Bayesian estimation of covariance matrices in hierarchical models. *J. Am. Statist. Assoc.* **94**, 1254-63.
- Daniels, M.J. & Kass, R.E. (2001). Shrinkage estimators for covariance matrices. *Biometrics* **57**, 1173-84.
- Daniels, M. & Pourahmadi, M. (2002). Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika* **89**, 553-66.
- Denison, D.G.T., Mallick, B.K. & Gangopadhyay, A.K. (2002). A Bayesian curve fitting approach to power spectrum estimation. *J. Nonparam. Statist.* **14**, 141-53.
- Diggle, P.J. & al Wasel, I. (1997). Spectral analysis of replicated biomedical time series (with Discussion). *Appl. Statist.* **46**, 31-71.
- Durrett, R. (1996). *Probability: Theory and Examples*. Cincinnati: Duxbury Press.
- Gasser, T., Kneip, A. & Koehler, W. (1991). A flexible and fast method for automatic smoothing. *J. Am. Statist. Assoc.* **86**, 643-52.
- Gasser, T., Stroka, L. & Jennen-Steinmetz, C. (1986). Residual and residual pattern in nonlinear regression. *Biometrika* **73**, 625-33.
- Grambsch, P., Meller, W.H. & Grambsch, P. (2002). Periodograms and pulse detection methods for pulsatile hormone data. *Statist. Med.* **21**, 2331-44.
- Hall, P., Fisher, N. & Hoffman, B. (1994). On the nonparametric estimation of covariance functions. *Ann. Statist.* **22**, 2115-34.
- Hawkins, D.M. & Wixley, R.A.J. (1986). A note on the transformation of chi-squared variables to normality. *Am. Statistician* **40**, 296-98.
- Hermann, E. (1997). Local bandwidth choice in kernel regression estimation. *Comp. Graph. Statist.* **6**, 35-54.
- Huerta, G. & West, M. (1999). Bayesian inference on periodicities and component spectral structure in time series. *J. Time Series Anal.* **20**, 401-16.
- Ombao, H.C., Raz, J.A., Strawderman, R.L., & Von Sachs, R. (2001). A simple generalised cross-validation method of span selection for periodogram smoothing. *Biometrika* **88**, 1186-92.

- Pawitan, Y. & Gangopadhyay, A.K. (1991). Efficient bias corrected nonparametric spectral estimation. *Biometrika* **78**, 825-32.
- Percival, D.B. & Walden, A.T. (2000). *Wavelet Methods for Time Series Analysis*. New York: Cambridge University Press.
- Shaman, P. (1977). Some Bayesian considerations in spectral estimation. *Biometrika* **64**, 79-84.
- Sun, L., Zidek, J.V., Le, N.D., & Ozkaynak, H. (2000). Interpolating Vancouver's daily ambient PM_{10} field. *Environmetrics* **11**, 651-63.
- Wand, M.P. & Jones, M.C. (1995). *Kernel Smoothing. Monographs on Statistics and Applied Probability*. London: Chapman and Hall.

Spectral Density	ϕ_1	ϕ_2	ϕ_3	ϕ_4	η_1	η_2	η_3
1	0.423	0	0	0	0	0	0
2	0.449	0	0	0	1.473	-1.849	0
3	0.643	-0.517	0.547	0	0	0	0
4	0.842	-0.845	0.786	0	-1.250	-1.440	0
5	0.1	0.07	-0.05	0.2	0.01	-0.2	0.27

Table 1: Parameters for spectral densities 1 - 5

SD ^b	Spectral Density Estimators ^a						
	Prd	DC	Wvlt	Adpt	ML	Shrnk	MASE
1	0.035 (0.001)	0.015 (0.001)	0.029 (0.002)	0.011 (0.001)	0.004 (0.0002)	0.004 (0.0002)	0.005 (0.0002)
2	0.52 (0.02)	0.29 (0.05)	0.48 (0.02)	0.18 (0.01)	0.09 (0.01)	0.09 (0.01)	0.098 (0.004)
3	0.088 (0.003)	0.040 (0.002)	0.066 (0.003)	0.043 (0.002)	0.029 (0.001)	0.029 (0.001)	0.037 (0.001)
4	53 (3)	29.3 (0.8)	34.3 (0.8)	46 (3)	113 (31)	72 (14)	42 (2)
5	0.042 (0.003)		0.039 (0.003)	0.020 (0.002)			0.013 (0.001)
6	0.318 (0.008)		0.33 (0.02)	0.090 (0.004)			0.086 (0.003)
7	0.84 (0.04)		0.78 (0.05)	0.38 (0.03)			0.22 (0.02)
8	1.46 (0.08)		1.7 (0.1)	0.66 (0.03)			0.94 (0.03)

^aPrd: Periodogram; DC: Daniels-Cressie estimator; Wvlt: Wavelet estimator; Adpt: Adaptive bandwidth estimator; ML: Maximum likelihood estimator; Shrnk: Shrinkage estimator; MASE: Model averaged shrinkage estimator

^bSD: Spectral Density

Table 2: Simulation study. Values of *MISE* of estimators based on sample size $n = 64$. Monte Carlo standard errors are given in parentheses.

SD ^b	Spectral Density Estimators ^a						
	Prd	DC	Wvlt	Adpt	ML	Shrnk	MASE
1	0.69 (0.04)	0.17 (0.01)	0.34 (0.03)	0.11 (0.01)	0.037 (0.002)	0.037 (0.002)	0.041 (0.002)
2	9.8 (0.6)	4 (2)	6.5 (0.5)	2.1 (0.2)	0.45 (0.04)	0.45 (0.04)	0.60 (0.05)
3	1.8 (0.1)	0.63 (0.04)	0.89 (0.06)	0.66 (0.06)	0.54 (0.03)	0.51 (0.02)	0.53 (0.02)
4	1397 (73)	889 (21)	961 (22)	1139 (75)	6198 (2000)	3431 (840)	1291 (90)
5	1.0 (0.1)		0.76 (0.08)	0.36 (0.07)			0.21 (0.02)
6	4.6 (0.2)		3.9 (0.2)	0.64 (0.05)			0.51 (0.03)
7	20 (2)		14 (1)	6.4 (0.7)			3.2 (0.8)
8	36 (2)		37 (4)	9.2 (0.5)			17 (1)

^aPrd: Periodogram; DC: Daniels-Cressie estimator; Wvlt: Wavelet estimator; Adpt: Adaptive bandwidth estimator; ML: Maximum likelihood estimator; Shrnk: Shrinkage estimator; MASE: Model averaged shrinkage estimator

^bSD: Spectral Density

Table 3: Simulation study. Values of $MMSD$ of estimators based on sample size $n = 64$. Monte Carlo standard errors are given in parentheses.

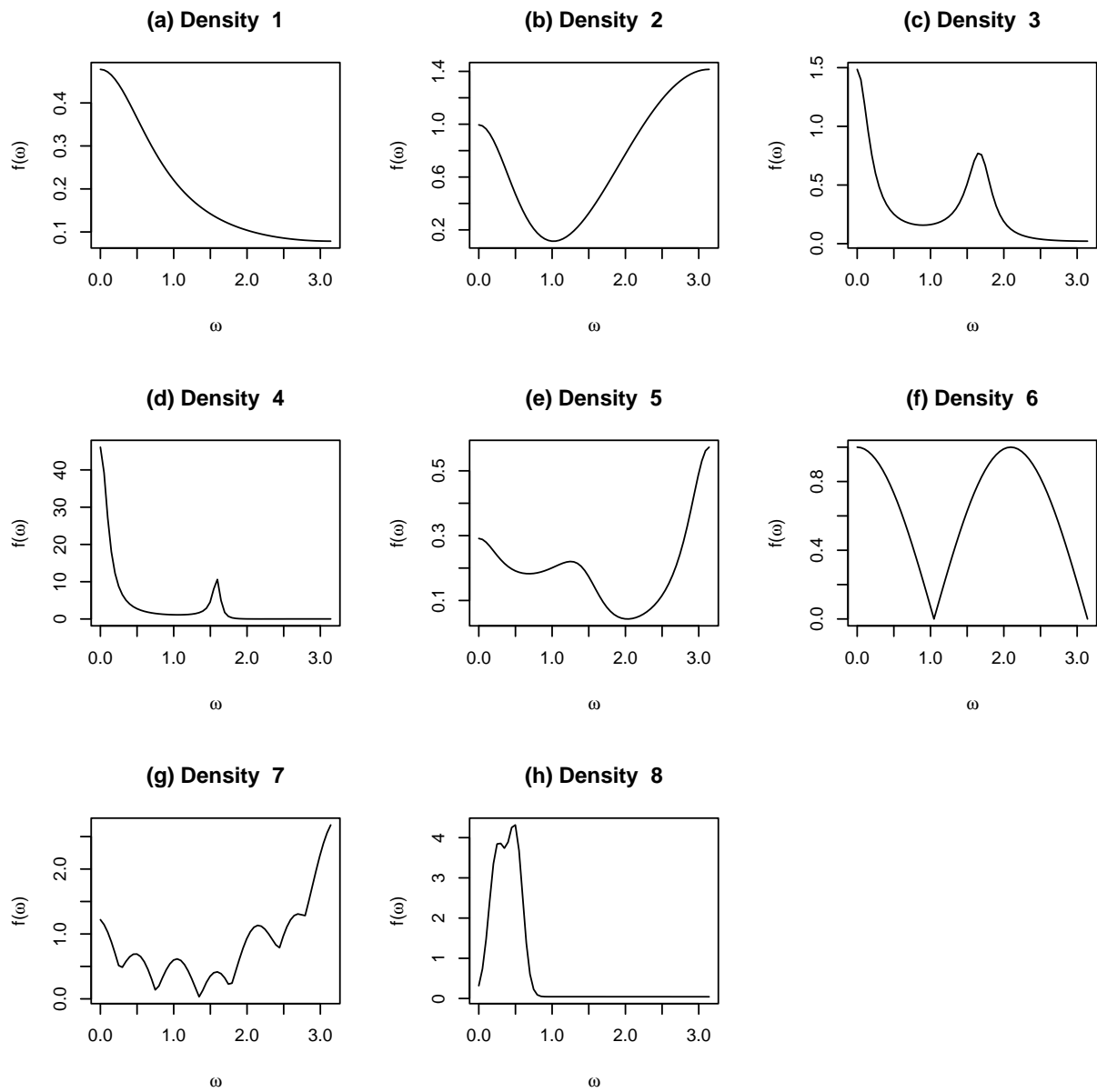


Figure 1: Spectral densities of the eight time series models used in the simulation study. (a) - (h) correspond to spectral densities 1-8.

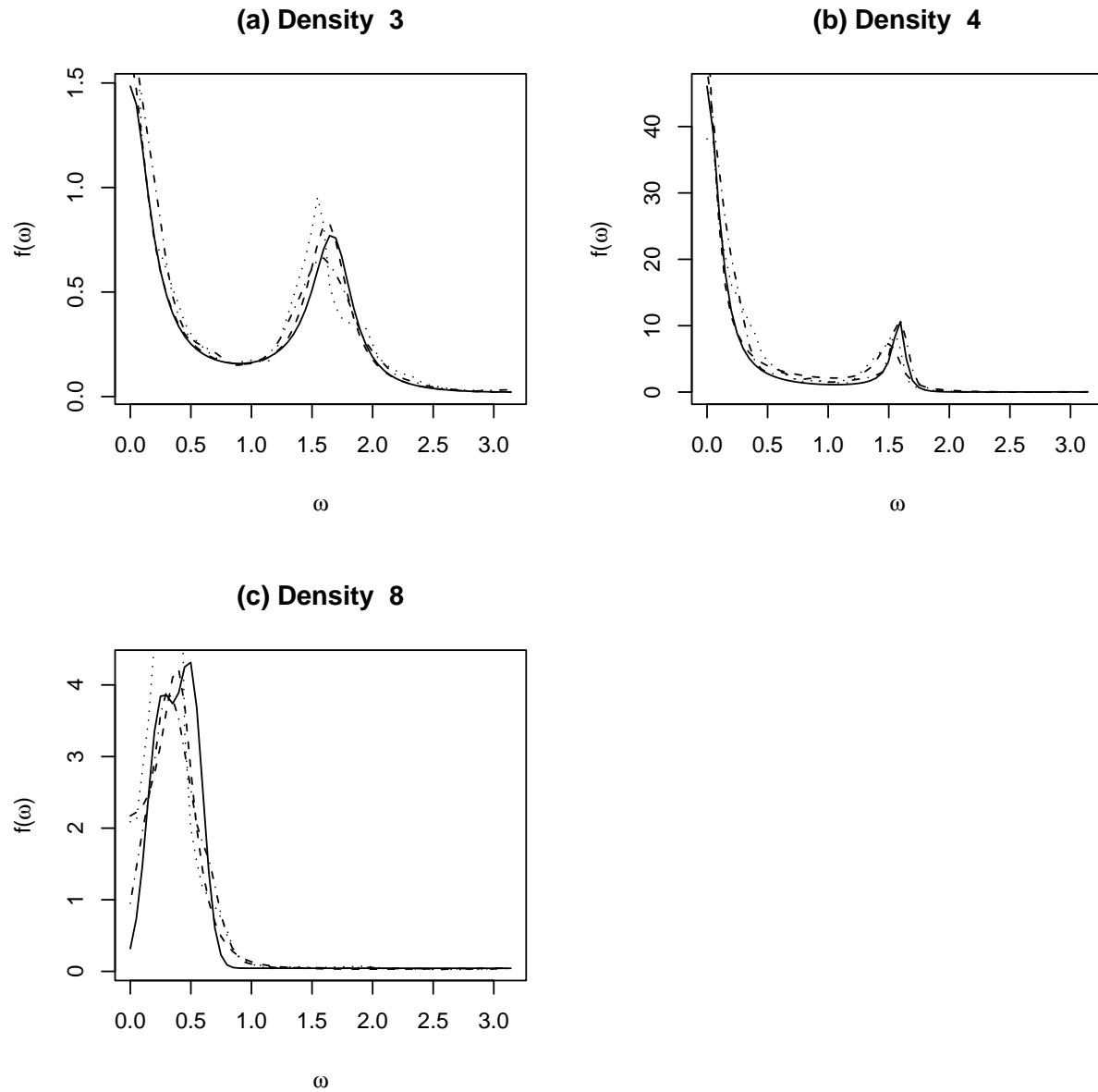


Figure 2: Simulation study. Estimates averaged over 5 randomly selected samples of size $n = 128$ for (a) spectral density 3, (b) spectral density 4, (c) spectral density (8). ($\cdots\cdots$) Wavelet Estimator, ($-\cdot-\cdot-$) Adaptive Estimator, ($-\ -$) Model Averaged Shrinkage Estimator, ($-\ -$) True Spectral Density

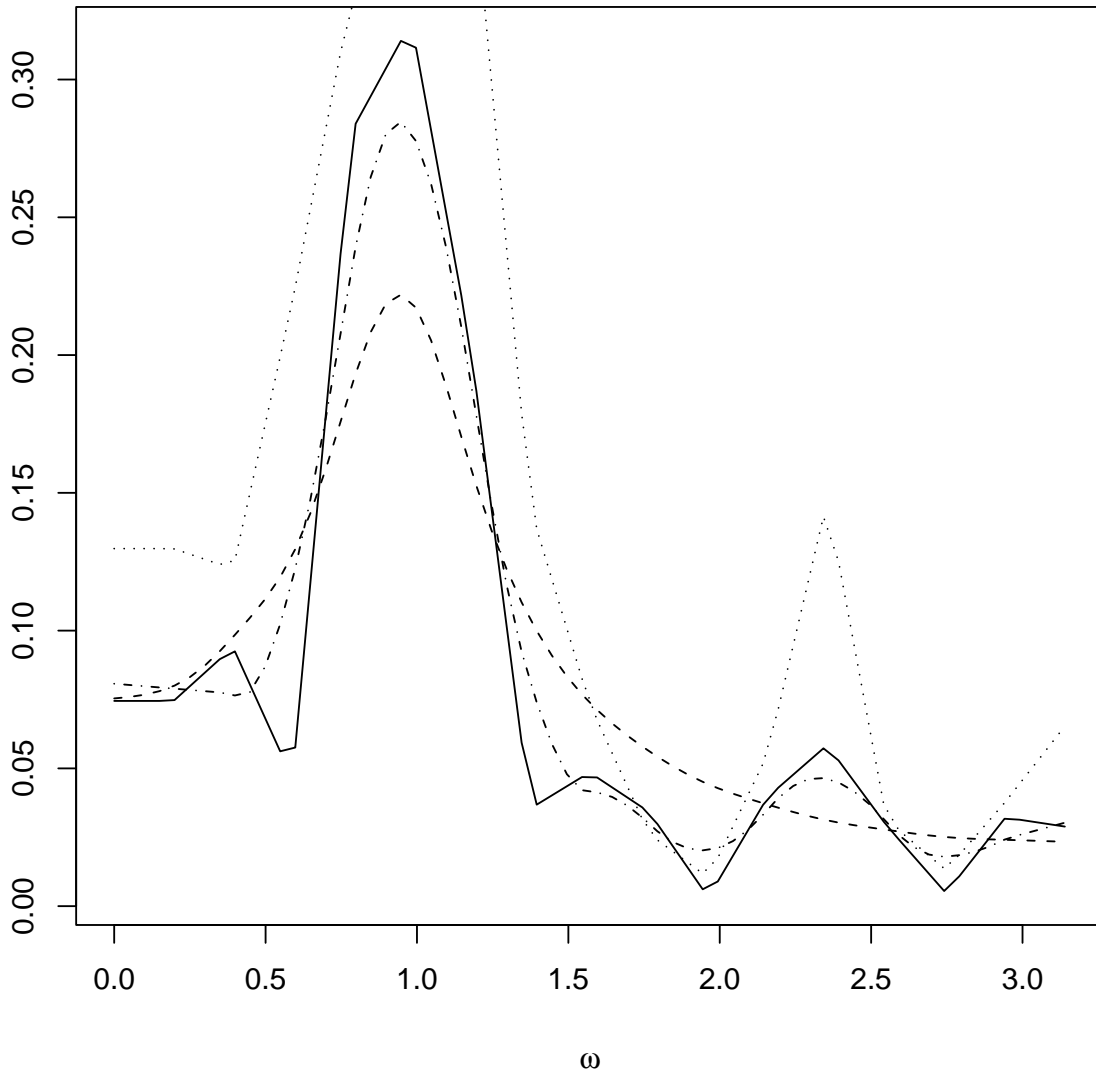


Figure 3: Luteinising hormone experiment. Spectral density estimates for one subject. (·····) Wavelet Estimator, (— · — ·) Adaptive Estimator, (— —) Model Averaged Shrinkage Estimator, (—) Periodogram.