PSYCHOMETRIKA—VOL. 79, NO. 1, 105–129
JANUARY 2014
DOI: 10.1007/S11336-013-9354-0

SPECTRAL DECOMPOSITIONS OF MULTIPLE TIME SERIES: A BAYESIAN NON-PARAMETRIC APPROACH

CHRISTIAN MACARO

SAS INSTITUTE

RAQUEL PRADO

DEPARTMENT OF APPLIED MATHEMATICS AND STATISTICS, UNIVERSITY OF CALIFORNIA, SANTA CRUZ

We consider spectral decompositions of multiple time series that arise in studies where the interest lies in assessing the influence of two or more factors. We write the spectral density of each time series as a sum of the spectral densities associated to the different levels of the factors. We then use Whittle's approximation to the likelihood function and follow a Bayesian non-parametric approach to obtain posterior inference on the spectral densities based on Bernstein–Dirichlet prior distributions. The prior is strategically important as it carries identifiability conditions for the models and allows us to quantify our degree of confidence in such conditions. A Markov chain Monte Carlo (MCMC) algorithm for posterior inference within this class of frequency-domain models is presented.

We illustrate the approach by analyzing simulated and real data via spectral one-way and two-way models. In particular, we present an analysis of functional magnetic resonance imaging (fMRI) brain responses measured in individuals who participated in a designed experiment to study pain perception in humans.

Key words: spectral one-way and two-way models, Bayesian nonparametrics, Whittle's approximation, Bernstein–Dirichlet priors.

1. Introduction

Time series data from several subjects that can be classified into groups according to a set of features are oftentimes recorded in clinical and non-clinical studies. For instance, a set of brain signals recorded from several individuals may be grouped by the type of stimulus each individual received in a given experimental setting. In such cases, one of the main goals of the time series analysis is to determine if there are differences in the spectral characteristics of the signals across the groups.

More specifically, suppose that a collection of time series $\{y_{i_1,i_2,h}(t)\}$ are recorded during a particular experimental setting in which i_1 indexes the level of some factor, say A, i_2 indexes the level of some other factor, say B, and h indexes the time series within given levels of the A and B factors, where $i_1 = 1 : N_1$, $i_2 = 1 : N_2$, $h = 1 : H_{i_1,i_2}$, and $t = 1 : T$. Such data may arise in designed experiments where factor A is a given treatment or experimental condition—e.g., the type of stimulus in a neuroscience experiment or the disease level—, factor B is another treatment or experimental condition, and h indexes the individuals who were assigned to levels i_1 and i_2 of the A and B factors, respectively. We are interested in fitting models that allow us to characterize the spectral features of the time series in terms of effects associated to the levels of the factors and, possibly also, in terms of effects associated with the interaction of such factors.

Requests for reprints should be sent to Christian Macaro, SAS Institute, Cary, NC, USA. E-mail: christianmacaro@gmail.com

© 2013 The Psychometric Society. This article is published with open access at Springerlink.com

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

Assume for instance that each time series can be decomposed as

$$y_{i_1,i_2,h}(t) = \beta_{i_1}^{(1)}(t) + \beta_{i_2}^{(2)}(t) + \varepsilon_{i_1,i_2,h}(t), \tag{1}$$

where the unobserved processes $\beta_{i_1}^{(1)}(t)$ and $\beta_{i_2}^{(2)}(t)$ are assumed to be uncorrelated, i.e., $\mathrm{Cov}(\beta_{i_1}^{(1)}(t), \beta_{i_2}^{(2)}(t)) = 0$ for all i_1, i_2 and all $t = 1:T$. Furthermore, for a given k with $k = 1, 2$ assume that $\mathrm{Cov}(\beta_i^{(k)}(t), \beta_j^{(k)}(t)) = 0$ when $i \neq j$ for all t. The $\varepsilon_{i_1,i_2,h}(t)$s are idiosyncratic—i.e., individual specific—error series assumed to be independent of the other unobserved processes and also mutually independent. This implies that the spectral density of $y_{i_1,i_2,h}(t)$, which is the Fourier transform of the theoretical autocovariance function of the $y_{i_1,i_2,h}(t)$ process and is denoted as $f^*_{y_{i_1,i_2,h}}(\lambda)$, can be written as

$$f^*_{y_{i_1,i_2,h}}(\lambda) = f^{*,(1)}_{i_1}(\lambda) + f^{*,(2)}_{i_2}(\lambda) + f^*_{\varepsilon_{i_1,i_2,h}}(\lambda), \tag{2}$$

where $f^{*,(1)}_{i_1}(\lambda)$ and $f^{*,(2)}_{i_2}(\lambda)$ are the spectral densities of the unobserved processes associated with levels i_1 and i_2 of factors A and B, respectively; and $f^*_{\varepsilon_{i_1,i_2,h}}(\lambda)$ is the spectral density of the idiosyncratic error component associated to individual h within such levels.

There is a relatively rich literature on Bayesian approaches for estimation of functions that can be used in the context of analyzing multiple signals. DiMatteo, Genovese, and Kass (2001) describe a Bayesian method for fitting curves to data drawn from a distribution in the exponential family. This approach assumes that the curves can be well approximated by splines with an unknown number of knots and unknown knot locations that are then inferred via reversible jump Markov chain Monte Carlo. The method is used to analyze individual (not multiple) time series data obtained from functional magnetic resonance imaging (fMRI) experiments. In such context the method of DiMatteo et al. (2001) is a time-domain approach that assumes that a given time series $y(t)$ can be modeled as $y(t) = f(t) + \epsilon(t)$, with $f(t) = \sum_{j=1}^{k+2} \theta_j b_j(t)$, where $b_j(t)$ is the jth function of a cubic B-spline. In contrast, our proposed frequency-domain method aims to infer the spectral characteristics of multiple, not just a single, time series. A possible way of analyzing multiple times series in the spectral domain would consist on extending the method of DiMatteo et al. (2001) for fitting curves to the periodogram ordinates of the multiple time series recorded at various levels of the factors. Such extension would involve developing and implementing models that can handle multiple series given that the approach of DiMatteo et al. (2001) can only be directly applied to a single data set, e.g., the periodogram ordinates of an individual time series. Furthermore, the method of DiMatteo et al. (2001) applies to independent data $(y_1, x_1), \ldots, (y_n, x_n)$ that satisfy a model of the form $y_i | x_1, \ldots, x_n \sim p(y_i | f(x_i), \sigma)$. However, the periodogram ordinates are only asymptotically independent. Therefore, it would not be advisable to directly apply the free-knots splines method in order to estimate spectral densities. On the other hand, the method proposed here is based on the approaches of Choudhuri, Ghosal, and Roy (2004) and Macaro (2010), which provide consistent Bayesian estimates even if the periodogram ordinates are only asymptotically independent (see Choudhuri et al., 2004, pp. 1056–1057, and Macaro, 2010, p. 384).

Regarding factorial temporal data, several authors have considered various time-domain and frequency-domain approaches. Some key references include, among others, Shumway (1970), Brillinger (1973, 1980), and Stoffer (1999). Shumway and Stoffer (2006) also summarize and illustrate several aspects of some of such approaches. As mentioned above, we follow Choudhuri et al. (2004) and Macaro (2010) to obtain Bayesian non-parametric posterior inference of the spectral representation of multiple time series, including the particular case of factorial spectral decompositions, by using Bernstein–Dirichlet prior distributions (Petrone, 1999a, 1999b) on the spectral densities in (2). More specifically, Choudhuri et al. (2004) described a Bayesian

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

approach to estimating the spectral density of a single stationary time series by imposing a non-parametric prior on such density through Bernstein polynomials. Then, Macaro (2010) proposed a mixture generalization of such approach in which each component in the spectral decomposition is identified using informative prior distributions. We extend the methods and algorithms of Macaro (2010) to provide a Bayesian non-parametric spectral analysis of multiple time series recorded during designed experiments that may involve several factors, various levels within each factor and several individuals. Bernstein–Dirichlet prior distributions have been successfully applied to analyze other types of data that arise in psychometric applications. For example, Karabatsos and Walker (2009) use a bivariate Bernstein–Dirichlet prior in the context of modeling test equating. Alternative Bayesian approaches to spectral density estimation of a single time series, including stationary and long-range dependence time series, as well as nonstationary time series with slowly varying dynamics or piecewise stationary time series, can be found in Carter and Kohn (1997), Gangopadhyay, Mallick, and Denison (1998), Liseo, Marinucci, and Petrella (2001), and Rosen, Stoffer, and Wood (2009), among others. A Bayesian approach for estimating the spectral density of multivariate stationary time series is presented in Rosen and Stoffer (2007).

The article is organized as follows. The details of the modeling approach, including some examples, are provided in Section 2. Section 3 summarizes the Markov chain Monte Carlo algorithm[1] for posterior inference. In Section 4 we present the results of several simulation studies and discuss aspects of identification through the prior distributions. In Section 5 we analyze multiple time series of fMRI brain responses measured in individuals who participated in an experiment designed to study pain perception in humans. Finally, Section 6 presents the conclusions.

2. Bayesian Spectral Decompositions of Factorial Time Series Data

2.1. General Model Formulation

Let $\{y_{i_1,\ldots,i_D,h}(t)\}$ be a set of time series for $i_d = 1 : N_d$, $h = 1 : H_{i_1,\ldots,i_D}$ and $t = 1 : T$. Assume that each time series $y_{i_1,\ldots,i_D,h}(t)$ can be decomposed as a sum of D unobservable components plus an idiosyncratic error term. That is,

$$y_{i_1,\ldots,i_D,h}(t) = \sum_{d=1}^{D} \beta_{i_d}^{(d)}(t) + \varepsilon_{i_1,\ldots,i_D,h}(t), \tag{3}$$

where $\beta_{i_j}^{(j)}(t)$ and $\beta_{i_k}^{(k)}(t)$ are assumed to be independent for all $j \neq k$ and all t, and the $\varepsilon_{i_1,\ldots,i_D,h}(t)$ are assumed to be independent of the $\beta_{i_j}^{(d)}(t)$ processes for all d and also mutually independent. Furthermore, it is assumed that for a given d, $\beta_{i_j}^{(d)}(t)$ and $\beta_{i_k}^{(d)}(t)$ are independent for all $i_j \neq i_k$ and all t, and that $\varepsilon_{i_1,\ldots,i_D,h}(t_0)$ and $\varepsilon_{i_1,\ldots,i_D,h}(t_1)$ are independent for all $t_0 \neq t_1$. This implies that the spectral density of $y_{i_1,\ldots,i_D,h}(t)$, denoted as $f^*_{y_{i_1,\ldots,i_D,h}}(\lambda)$, can be written as

$$f^*_{y_{i_1,\ldots,i_D,h}}(\lambda) = \sum_{d=1}^{D} f^{*,(d)}_{i_d}(\lambda) + f^*_{\varepsilon_{i_1,\ldots,i_D,h}}(\lambda), \tag{4}$$

[1]The code is available upon request. Please email C. Macaro at christianmacaro@gmail.com.

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

where $f_{i_d}^{*,(d)}(\lambda)$ and $f_{\varepsilon_{i_1,\ldots,i_D,h}}^{*}(\lambda)$ are, respectively, the spectral densities of the unobserved factors and the idiosyncratic error components. The spectral density is the Fourier transform of the theoretical autocovariance function. Under some regularity conditions it is a real and continuous function over $\lambda \in (-\pi, \pi]$. A further assumption requires the spectral densities to be bounded and bounded away from zero, ruling out long memory and band limited processes.

2.2. Examples

2.2.1. Baseline Plus Single Factor Model In this example we consider a one-way model represented as a baseline process plus one factor with two levels. Specifically, let $d = 1:2$, $i_1 = 1$, $i_2 = 1:2$, $h = 1:2$, and assume the following structure on $y_{1,i_2,h}(t)$:

$$\text{Observations:} \quad y_{1,i_2,h}(t) = \beta_1^{(1)}(t) + \beta_{i_2}^{(2)}(t) + \varepsilon_{1,i_2,h}(t), \quad \varepsilon_{1,i_2,h}(t) \sim N(0,1), \tag{5}$$

$$\text{Baseline:} \quad \beta_1^{(1)}(t) \sim \mathrm{AR}\big(\phi^{(1)} = 0.9, 1\big), \tag{6}$$

$$\text{One factor, two levels:} \quad \beta_{i_2}^{(2)}(t) \sim \begin{cases} \mathrm{AR}(\phi_{1,1}^{(2)} = -0.8, \phi_{1,2}^{(2)} = -0.9, 1) & i_2 = 1, \\ \mathrm{AR}(\phi_{2,1}^{(2)} = -1.5, \phi_{2,2}^{(2)} = -0.8, 1) & i_2 = 2. \end{cases} \tag{7}$$

Therefore, the baseline process $\beta_1^{(1)}(t)$ is an autoregressive process of order one, or AR(1), with coefficient 0.9, while the factor has two levels, the first level corresponds to an autoregressive process of order two, or AR(2), with two reciprocal characteristic roots each with modulus 0.949 and frequency 2.006 (period 3.132), and the second level is an AR(2) with two reciprocal characteristic roots each with modulus 0.894 and frequency 2.566 (period 2.449). The spectral representation is

$$f_{y_{1,i_2,h}}^{*}(\lambda) = f_1^{*,(1)}(\lambda) + f_{i_2}^{*,(2)}(\lambda) + f_{\varepsilon_{i_1,i_2,h}}^{*}(\lambda), \tag{8}$$

with $f_1^{*,(1)}(\lambda) = 1/[2\pi(1 + (\phi^{(1)})^2 - 2\phi^{(1)}\cos(\lambda))]$, and

$$f_{i_2}^{*,(2)}(\lambda) = 1/\big[2\pi\big(1 + \big(\phi_{i_2,1}^{(2)}\big)^2 + 2\phi_{i_2,2}^{(2)} + \big(\phi_{i_2,2}^{(2)}\big)^2 + 2\big(\phi_{i_2,1}^{(2)}\phi_{i_2,2}^{(2)} - \phi_{i_2,1}^{(2)}\big)\cos(\lambda) - 4\phi_{i_2,2}^{(2)}\cos^2(\lambda)\big)\big].$$

Therefore, the decomposition in (8) is such that $f_1^{*,(1)}(\lambda)$ captures constant and persistent effects, while the $f_{i_2}^{*,(2)}(\lambda)$ spectra capture quasiperiodic dynamics that shift across the levels, as shown in Figure 1.

2.2.2. Two-Way Models The probabilistic spectral decomposition in the example below corresponds to that of a two-way temporal model. In particular, assume that a two-way factor model with two levels in each factor is obtained by letting $d = 1:2$, $i_1 = 1:2$, $i_2 = 1:2$, $h = 1:H$ and by assuming the following structure on $y_{i_1,i_2,h}(t)$:

$$\text{Observations:} \quad y_{i_1,i_2,h}(t) = \beta_{i_1}^{(1)}(t) + \beta_{i_2}^{(2)}(t) + \varepsilon_{i_1,i_2,h}(t), \quad \varepsilon_{i_1,i_2,h}(t) \sim N(0,1), \tag{9}$$

$$\text{First factor, two levels:} \quad \beta_{i_1}^{(1)}(t) \sim \begin{cases} N(0,1), & i_1 = 1, \\ \mathrm{AR}(\phi_{2,1}^{(1)} = 1.46, \phi_{2,2}^{(1)} = -0.81, 1), & i_1 = 2, \end{cases} \tag{10}$$

$$\text{Second factor, two levels:} \quad \beta_{i_2}^{(2)}(t) \sim \begin{cases} \mathrm{AR}(\phi_{1,1}^{(2)} = 0.9, 1), \\ \mathrm{AR}(\phi_{2,1}^{(2)} = -0.9, \phi_{2,2}^{(2)} = -0.81, 1). \end{cases} \tag{11}$$

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

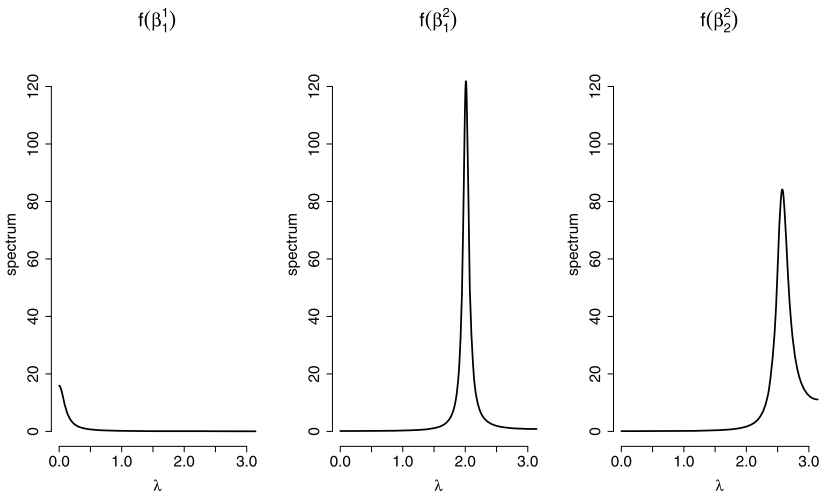


FIGURE 1.
Spectral representation of the one-way factor model given in Equations (5)–(7).

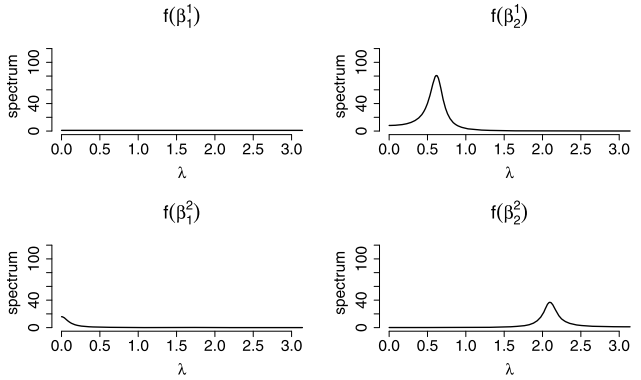


FIGURE 2.
Spectral representation of the two-way factor model given in Equations (9)–(11).

Therefore, the spectral representation of $y_{i_1,i_2,h}(t)$ is given by

$$f^*_{y_{i_1,i_2,h}}(\lambda) = f^{*,(1)}_{i_1}(\lambda) + f^{*,(2)}_{i_2}(\lambda) + f^*_{\varepsilon_{i_1,i_2,h}}(\lambda),$$

with $f^{*,(1)}_1(\lambda)$ the flat spectrum of a white-noise process and $f^{*,(1)}_2(\lambda)$, $f^{*,(2)}_1(\lambda)$ and $f^{*,(2)}_2(\lambda)$ the spectra of quasiperiodic AR(2) processes. The densities in the spectral decomposition of $y_{i_1,i_2,h}(t)$ are shown in Figure 2.

2.2.3. Interaction Effects Oftentimes studies are characterized by the presence of one or more interaction terms. For example, a two-way time-domain model with interactions can be written as

$$y_{i_1,i_2,h}(t) = \beta^{(1)}_{i_1}(t) + \beta^{(2)}_{i_2}(t) + \gamma^{(1,2)}_{i_1,i_2}(t) + \varepsilon_{i_1,i_2,h}(t), \tag{12}$$

where $\gamma^{(1,2)}_{i_1,i_2}(t)$ are time series processes that represent possible interaction effects between factors 1 and 2 at the i_1 and i_2 levels. In general, if we have D factors, we can extend the model

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

in (3) to allow for second order interactions as follows:

$$y_{i_1,\ldots,i_D}(t) = \sum_{d=1}^{D} \beta_{i_d}^{(d)}(t) + \sum_{k=1}^{D}\sum_{l=1}^{D} \gamma_{i_k,i_l}^{(k,l)}(t) + \varepsilon_{i_1,i_2,h}(t). \tag{13}$$

The Bayesian non-parametric approach described in Section 3 is illustrated with one-way and two-way models with no interactions in Sections 4 and 5. However, the methodology can be applied to models that include interaction effects by adding such terms in Whittle's likelihood approximation and by directly modeling the spectra of the interaction processes using the Bernstein–Dirichlet priors discussed in Section 2.4. Note that higher order interaction processes (e.g., third or higher order) can be added to (13) and modeled similarly.

As it will be shown in Section 4 with simulated data, when the number of factors increases, more restrictions need to be added to guarantee the identifiability of the different processes in the spectral decomposition. Therefore, instead of adding interaction effects and directly modeling the spectral densities of such processes, we propose simple post-processing calculations that can be used to explore if such interaction processes are present. We summarize such calculations in the case of a two-way model. Assume that $y_{i_1,i_2,h}(t)$ is represented as the two-way model process in (12). Such model can be written as $y_{i_1,i_2,h}(t) = \beta_{i_1}^{(1)}(t) + \beta_{i_2}^{(2)}(t) + \varepsilon'_{i_1,i_2,h}(t)$, with $\varepsilon'_{i_1,i_2,h}(t) = \gamma_{i_1,i_2}(t) + \varepsilon_{i_1,i_2,h}(t)$. If we were to analyze this scenario without implicitly considering the interaction terms, such terms would appear in the idiosyncratic factors $\varepsilon'_{i_1,i_2,h}(t)$. Therefore, we could obtain estimates of the spectral densities of the interaction terms as follows:

1. Evaluate the common baseline across all the individuals (i_1, i_2, h). Compute

$$f_{\bar{\varepsilon}}^{*}(\lambda) = \min_{i_1,i_2,h} f_{\varepsilon_{i_1,i_2,h}}^{*\prime}(\lambda).$$

2. Evaluate the corresponding individual residual. For each λ and each (i_1, i_2, h) compute

$$f_{[\text{residual}]_{i_1,i_2,h}}^{*\prime}(\lambda) = f_{\varepsilon_{i_1,i_2,h}}^{*\prime}(\lambda) - f_{\bar{\varepsilon}}^{*}(\lambda).$$

3. Evaluate the interaction term across all the individuals within the same group. For each λ and each (i_1, i_2) compute an estimate of the spectrum of the interaction process $\gamma_{i_1,i_2}^{(1,2)}(t)$ as

$$f_{\gamma_{i_1,i_2}}^{*}(\lambda) = \min_{h} f_{[\text{residual}],i_1,i_2,h}^{*\prime}(\lambda).$$

By inspecting the posterior distribution of $f_{\gamma_{i_1,i_2}}^{*}(\lambda)$ for each combination of (i_1, i_2), it can be determined if one or more interaction terms should be included in the analysis.

2.3. Spectral Representation

Following a similar approach to that in Macaro (2010), we use a discrete version of Whittle's approximation of the likelihood function (Whittle, 1957, 1962) to estimate the factors. That is, the likelihood is approximated as

$$L_T\left[I_{y_{i_1,\ldots,i_D,h}}(\lambda_{1:J}) | f_{\beta}^{*,(1:D)}(\lambda_{1:J}), f_{\varepsilon}^{*}(\lambda_{1:J})\right] = \prod_{j=1}^{J} \frac{\exp\{-\frac{I_{y_{i_1,\ldots,i_D,h}}(\lambda_j)}{\sum_{d=1}^{D} f_{\beta_{i_d}}^{*,d}(\lambda_j) + f_{\varepsilon_{i_1,\ldots,i_D,h}}^{*}(\lambda_j)}\}}{\sum_{d=1}^{D} f_{\beta_{i_d}}^{*,d}(\lambda_j) + f_{\varepsilon_{i_1,\ldots,i_D,h}}^{*}(\lambda_j)}, \tag{14}$$

where $I_{y_{i_1,\ldots,i_D,h}}(\lambda_j)$ are the periodogram ordinates evaluated at the Fourier frequencies,

$$\lambda_j = \frac{2\pi j}{T} \quad \text{for } j = 0, 1, \ldots, \lfloor T/2 \rfloor := J.$$

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

These summarize the information obtained from the data, while the $f^{*,d}_{\beta_{i_d}}(\lambda)$s are spectral ordinates to be estimated.

We now combine Whittle's approximation to the likelihood in (14) with prior distributions on each $f^{*,d}_{\beta_{i_d}}(\lambda)$ to obtain posterior inference in the spectral domain. As discussed in Macaro (2010), prior elicitation is difficult since it involves the approximation of continuous spectral densities over the set of Fourier frequencies. Choudhuri et al. (2004) propose the use of Bernstein–Dirichlet prior distributions (Petrone, 1999a, 1999b), while Macaro (2010) presents a mixture generalization of their work. Specifically, the spectrum of an observable time series is modeled as a weighted sum of several unobserved spectral densities in Macaro (2010). The resulting procedure is indeed more flexible, although the real novelty of the approach is the possibility of studying unobserved components from a non-parametric viewpoint. Here we extend Macaro (2010) to consider factorial designs with multiple time series. In other words, the models and methods described in this section lead to a non-parametric spectral representation of the unobserved components underlying multiple time series recorded in factorial experimental settings. Below we describe the prior structure and how posterior inference can be achieved under such prior structure. We also discuss some issues related to prior choice and summarize the steps of a Markov chain Monte Carlo (MCMC) algorithm for posterior inference.

2.4. The Bernstein–Dirichlet Prior Distribution

The Bernstein polynomial prior developed by Petrone (1999a, 1999b) was used as a non-parametric prior to estimate densities on the [0, 1] interval, and was then used by Choudhuri et al. (2004) for Bayesian estimation of spectral densities. This non-parametric prior hinges on the uniform convergence of

$$\sum_{k=1}^{K} Q^{\dagger}\left(\frac{k-1}{K}, \frac{k}{K}\right)\text{Beta}_{(k,K-k+1)}(\lambda)$$

to $q(\lambda)$, a continuous probability density on [0, 1]. Here $Q^{\dagger}(\epsilon_1, \epsilon_2) := Q(\epsilon_2) - Q(\epsilon_1)$, with $Q(\lambda)$ being the probability function associated to $q(\lambda)$. In particular, the Bernstein–Dirichlet prior induces the weights $w_1, \ldots, w_K$ of the mixture through a Dirichlet process, i.e.,

$$q(\lambda) = \sum_{k=1}^{K} w_k \text{Beta}_{(k,K-k+1)}(\lambda), \tag{15}$$

with

$$\Pi(w_1, \ldots, w_K | K) = M\ \text{Dirichlet}_{(\theta_1,\ldots,\theta_K)}(w_1, \ldots, w_K), \tag{16}$$

where $\theta_1, \ldots, \theta_K$ are the shape parameters, $M > 0$ is the concentration parameter, and K determines how flexible is the prior, with larger values of K leading to more flexible distributions.

We now represent each of the spectral densities, $f^{*,(d)}_{\beta_{i_d}}(\lambda)$ in (4), through a Bernstein kernel. In order to implement Bernstein–Dirichlet priors in the spectral domain, such densities must be normalized to the interval [0, 1]. Therefore, we can define a pseudo-spectral density function, $q_{y_{i_1,\ldots,i_D},h}(\lambda)$, and a normalization parameter, $\tau_{y_{i_1,\ldots,i_D},h}$, such that

$$\tau_{y_{i_1,\ldots,i_D},h} q_{y_{i_1,\ldots,i_D},h}(\lambda) := \frac{\sigma^2_{y_{i_1,\ldots,i_D}}}{2\pi} q_{y_{i_1,\ldots,i_D},h}(\lambda) = f_{y_{i_1,\ldots,i_D},h}(\lambda),$$

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

where

$$\sigma^2_{y_{i_1,\ldots,i_D,h}} = 2\int_0^{\pi} f^*_{y_{i_1,\ldots,i_D,h}}(u)\,du = 2\int_0^{1} f^*_{y_{i_1,\ldots,i_D,h}}(\pi u)\,du := 2\pi\int_0^{1} f_{y_{i_1,\ldots,i_D,h}}(u)\,du.$$

Then, the normalization parameters for the spectral representation of the unobserved factors $\beta^{(d)}_{i_d}(t)$ and the idiosyncratic errors $\varepsilon_{i_1,\ldots,i_D,h}(t)$ can be derived as follows:

$$\begin{aligned} q_{y_{i_1,\ldots,i_D,h}}(\lambda) &= \sum_{d=1}^{D} q^{(d)}_{\beta_{i_d}}(\lambda) + q_{\varepsilon_{i_1,\ldots,i_D,h}}(\lambda) \\ &= \sum_{d=1}^{D} \frac{f^{(d)}_{\beta_{i_d}}(\lambda)}{\tau_{\beta^{(d)}_{i_d}}} + \frac{f_{\varepsilon_{i_1,\ldots,i_D,h}}(\lambda)}{\tau_{\varepsilon_{i_1,\ldots,i_D,h}}}, \end{aligned}$$

such that $\tau_{y_{i_1,\ldots,i_D,h}} = \sum_{d=1}^{D} \tau_{\beta^{(d)}_{i_d}} + \tau_{\varepsilon_{i_1,\ldots,i_D,h}}$, with

$$\tau_{\beta^{(d)}_{i_d}} = \frac{\sigma^2_{\beta^{(d)}_{i_d}}}{2\pi} \quad \text{and} \quad \tau_{\varepsilon_{i_1,\ldots,i_D,h}} = \frac{\sigma^2_{\varepsilon_{i_1,\ldots,i_D,h}}}{2\pi}.$$

Since $\tau_{\beta^{(d)}_{i_1}}, \ldots, \tau_{\beta^{(d)}_{i_D}}$ and $\tau_{\varepsilon_{i_1,\ldots,i_D,h}}$ are proportional to the variances of the factors, inverse-gamma distributions can be used as priors. That is,

$$\Pi[\tau_{(\cdot)}|a_{(\cdot)}, b_{(\cdot)}] = \frac{b^{a_{(\cdot)}}_{(\cdot)}}{\Gamma[a_{(\cdot)}]}\left[\frac{1}{\tau_{(\cdot)}}\right]^{a_{(\cdot)}+1} \exp\left(-\frac{b_{(\cdot)}}{\tau_{(\cdot)}}\right). \tag{17}$$

The pseudo-spectral densities $q^{(d)}_{\beta_{i_d}}(\lambda)$ and $q_{\varepsilon_{i_1,\ldots,i_D,h}}(\lambda)$ will be discretized over the set of Fourier frequencies and estimated using Bernstein–Dirichlet priors.

Note that infinitely many choices of $\beta^{(d)}_{i_d}$ and $\varepsilon_{i_1,\ldots,i_D}(t)$ would lead to the same $y_{i_1,\ldots,i_D,h}(t)$. The same is true in the spectral domain. Therefore, identifiability conditions need to be imposed. One way of doing so is by adding ANOVA-type restrictions (see Section 4.2). Alternatively, following Macaro (2010), identifiability can be achieved through proper specification of informative Bernstein–Dirichlet prior as follows. For each $q^{(d)}_{\beta_{i_d}}(\lambda)$ we have the representation in (15) and (16) characterized by weights $w^{i_d,d}_{k_d}$ for $k_d = 1 : K^d_{i_d}$ and concentration parameter $M^d_{i_d}$. Similarly, for each $q_{\varepsilon_{i_1,\ldots,i_D,h}}(\lambda)$ we have a Bernstein–Dirichlet prior characterized by mixture weights $w^{\varepsilon_{i_1,\ldots,i_D,h}}_{k}$ for $k = 1 : K_{i_1,\ldots,i_D,h}$ and concentration parameter $M^{\varepsilon}_{i_1,\ldots,i_d,h}$. The concentration parameters control the variance of the Dirichlet processes. In other words, as $M^d_{i_d} \to 0$ the variance of the corresponding Dirichlet process increases; and as $M^d_{i_d} \to \infty$ such variance goes to zero. Therefore, instead of letting $M^d_{i_d} \to \infty$, we choose informed priors ($M^d_{i_d} < \infty$) on the $f^{(d)}_{i_d}(\lambda)$s. In practice, if we use the Dirichlet process characterization proposed by Sethuraman (1994), a natural choice is to fix $M^d_{i_d} = 1$ (Choudhuri et al., 2004), which is equivalent to saying that the prior for $v^{i_d,d}_{k_d,l}$ for $l = 1 : L$ is flat (see Section 3 for a detailed description of such parameters).

Theoretically, we could also set a prior on $K^d_{i_d}$ as proposed by Choudhuri et al. (2004), however, we fix these parameters in order to improve the speed and the stability of the MCMC algorithm (see Section 3 for further discussion). This will be illustrated in the analyses of simulated and real data sets in Sections 4 and 5. One of the main reasons for choosing this approach

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

is that in many practical scenarios we have information about the unobserved components in the decomposition of the series that we may want to incorporate into the model through the prior distribution. For instance, electroencephalographic signals typically display activity in four main frequency bands (see, e.g., Prado, 2010), which induces a natural way of decomposing each signal into four unobserved components, one per frequency band.

Finally, it is important to notice the difference between $K_{i_d}^d$ and $M_{i_d}^d$. The parameter $K_{i_d}^d$ is similar to the bandwidth parameter used when smoothing the periodogram: a smaller value of $K_{i_d}^d$ leads to a higher degree of smoothing. On the other hand, $M_{i_d}^d$ indicates the degree of trust that we have in our prior distribution.

3. Posterior Inference

The posterior associated with the likelihood function (14) and with the priors described above is not available in closed form. Nevertheless, samples from this distribution can be obtained with a Metropolis–Hastings Markov chain Monte Carlo (MCMC) algorithm. The original construction of the Dirichlet process is not suitable for MCMC schemes. Sethuraman (1994) proposed an alternative construction which has been implemented by Gelfand and Kottas (2002) for a more flexible algorithm. Our sampling procedure is a generalization of the algorithm in Macaro (2010) which, in turn, is an adaptation of the work of Choudhuri et al. (2004) for the analysis of spectral densities. Particularly, the MCMC algorithm outlined in Section 3.2 is not built for the weight parameters $w_1^{i_d,d},\ldots,w_{K_{i_d}^d}^{i_d,d}$ and $w_1^{\varepsilon_{i_1,\ldots,i_D,h}},\ldots,w_{K_{i_1,\ldots,i_D,h}}^{\varepsilon_{i_1,\ldots,i_D,h}}$, but for the set of parameters $r_{k_d,1}^{i_d,d},\ldots,r_{k_d,L}^{i_d,d}, v_{k_d,1}^{i_d,d},\ldots,v_{k_d,L}^{i_d,d}, r_{k_h,1}^{\varepsilon_{i_1,\ldots,i_D,h}},\ldots,r_{k_h,L}^{\varepsilon_{i_1,\ldots,i_D,h}}$, and $v_{k_h,1}^{\varepsilon_{i_1,\ldots,i_D,h}},\ldots,v_{k_h,L}^{\varepsilon_{i_1,\ldots,i_D,h}}$ such that

$$w_{k_d}^{i_d,d}=\sum_{l=0}^{L}p_{k_d,l}^{i_d,d}\mathbf{1}_{\{\frac{k_d-1}{K_{i_d}^d}<r_{k_d,l}^{i_d,d}\leq\frac{k_d}{K_{i_d}^d}\}}, \tag{18}$$

with

$$p_{k_d,1}^{i_d,d}=v_{k_d,1}^{i_d,d},\qquad p_{k_d,l}^{i_d,d}=v_{k_d,l}^{i_d,d}\prod_{i=1}^{l-1}\left(1-v_{k_d,i}^{i_d,d}\right),\qquad p_{k_d,0}^{i_d,d}=1-\sum_{l=1}^{L}p_{k_d,l}^{i_d,d}, \tag{19}$$

and

$$w_{k_h}^{\varepsilon_{i_1,\ldots,i_D,h}}=\sum_{l=0}^{L}p_{k_h,l}^{\varepsilon_{i_1,\ldots,i_D,h}}\mathbf{1}_{\{\frac{k_h-1}{K_{i_1,\ldots,i_D,h}}<r_{k_h,l}^{\varepsilon_{i_1,\ldots,i_D,h}}\leq\frac{k_h}{K_{i_1,\ldots,i_D,h}}\}}, \tag{20}$$

with

$$\begin{aligned}&p_{k_h,1}^{\varepsilon_{i_1,\ldots,i_D,h}}=v_{k_h,1}^{\varepsilon_{i_1,\ldots,i_D,h}},\qquad p_{k_h,l}^{\varepsilon_{i_1,\ldots,i_D,h}}=v_{k_h,l}^{\varepsilon_{i_1,\ldots,i_D,h}}\prod_{i=1}^{l-1}\left(1-v_{k_h,i}^{\varepsilon_{i_1,\ldots,i_D,h}}\right),\\&p_{k_d,0}^{\varepsilon_{i_1,\ldots,i_D,h}}=1-\sum_{l=1}^{L}p_{k_h,l}^{\varepsilon_{i_1,\ldots,i_D,h}},\end{aligned} \tag{21}$$

where the truncation level of the Dirichlet process is $L=\max(20,T^{1/3})$ (Choudhuri et al., 2004).

3.1. Prior Choice

The prior probabilities can be chosen by performing the following steps:

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

1. Choose the number and types of factors and idiosyncratic components. This will depend on the data structure and the particular decomposition of interest (e.g., one-way vs. two-way factor structures). For example, in the data analysis presented in Section 5, for each time series we use a model with two factors, where each factor has two levels, and one idiosyncratic component which tries to capture the individual-specific features that are not described by any of the two factors.
2. Determine the *a priori* spectral properties that characterize each factor and transform this information into kernel parameters $\theta_1^{i_d,d},\ldots,\theta_{K_{i_d}^d}^{i_d,d}$ for the corresponding Dirichlet process. For the idiosyncratic components the parameters can be set all equal to one to represent the prior assumption that these components are white noise. The prior for the remaining parameters $r_{k_d,1}^{i_d,d},\ldots,r_{k_d,L}^{i_d,d}$, $v_{k_d,1}^{i_d,d},\ldots,v_{k_d,L}^{i_d,d}$, $r_{k_h,1}^{\varepsilon_{i_1,\ldots,i_D,h}},\ldots,r_{k_h,L}^{\varepsilon_{i_1,\ldots,i_D,h}}$, and $v_{k_h,1}^{\varepsilon_{i_1,\ldots,i_D,h}},\ldots,v_{k_h,L}^{\varepsilon_{i_1,\ldots,i_D,h}}$ of the Bernstein representation are chosen according to the guidelines provided by Choudhuri et al. (2004).
3. Determine the flexibility of the priors on the factors: the smaller the value of $K_{i_d}^d$ and $K_{i_1,\ldots,i_D,h}$, the smoother the corresponding spectral densities. Notice that this is particularly useful to separate the white-noise effects—which are characterized by flat and smooth spectral densities—from other effects. It is worth to remember that although K_{i_d} and $K_{i_1,\ldots,i_D,h}$ control for the smoothness of the spectra, they do not control for the strength of the prior distributions (this is achieved by changing the values of $M_{i_d}^d$ and $M_{i_1,\ldots,i_D,h}$).
4. Choose the $M_{i_d}^d$ and $M_{i_1,\ldots,i_D,h}$ concentration parameters: large values reduce the variance of the corresponding Dirichlet processes. Note for example that by letting $M_{i_d}^d\to\infty$ the underlying components become fixed (for more details see Macaro, 2010).
5. Determine the prior distributions for the normalization parameters. These parameters are proportional to the variances, therefore inverse gammas seem to be appropriate:

$$\tau_{(.)}\sim\text{InvGamma}\big\{s_{(.)}^{\text{prior}},\tilde{\tau}_{(.)}\big[s_{(.)}^{\text{prior}}-1\big]\big\} \tag{22}$$

with $s_{(.)}^{\text{prior}}$ and $\tilde{\tau}_{(.)}$ chosen using the fact that

$$\text{E}[\tau_{(.)}]=\tilde{\tau}_{(.)}\quad\text{and}\quad\text{var}[\tau_{(.)}]=\frac{\tilde{\tau}_{(.)}^2}{s_{(.)}^{\text{prior}}-2}. \tag{23}$$

3.2. MCMC Algorithm

The posterior distribution associated with the likelihood function (14) and with the prior distributions described above is not available in closed form. The following Metropolis–Hastings MCMC algorithm can be used to obtain samples from the posterior distribution.

1. For each factor indexed by $d=1:D$
   (a) For each level of by $i_d=1:N_d$ of the factor d
      i. **Update** $\mathbf{v}$. For each $l=1:L$
         - Propose a candidate value for $v_{k_d,l}^{i_d,d}$.
         - Evaluate the new set of $p_{k_d,1}^{i_d,d},\ldots,p_{k_d,L}^{i_d,d}$ and the corresponding new set of mixture weights $w_1^{i_d,d},\ldots,w_{K_{i_d}^d}^{i_d,d}$.
         - Evaluate the new spectral density $f_{i_d}^{*,(d)}(\lambda)$ of the corresponding factor component and the likelihood (14).

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

- Decide whether to accept the proposed value of $v_{k_d,l}^{i_d,d}$.

ii. **Update r**. For each $l = 0 : L$

- Propose a candidate value for $r_{k_d,l}^{i_d,d}$.
- Evaluate the new set of mixture weights $w_1^{i_d,d}, \ldots, w_{K_{i_d}^d}^{i_d,d}$.
- Evaluate the new spectral density $f_{i_d}^{*,(d)}(\lambda)$ of the corresponding factor component and the likelihood (14).
- Decide whether to accept the proposed value of $r_{k_d,l}^{i_d,d}$.

iii. **Update τ**. Propose a candidate value of $\tau_{\beta_{i_d}}^{(d)}$ and evaluate the corresponding spectral density and the likelihood (14).

2. Given the updates for the levels i_d of each process d, update the corresponding idiosyncratic component.
   (a) For the set of indices $(i_1, \ldots, i_D, h)$,
   i. **Update v** for the spectral density $f_{\varepsilon_{i_1,\ldots,i_D,h}}^*(\lambda)$ as described in 1(a)i.
   ii. **Update r** for the spectral density $f_{\varepsilon_{i_1,\ldots,i_D,h}}^*(\lambda)$ as described in 1(a)ii.
   iii. **Update τ** for the spectral density $f_{\varepsilon_{i_1,\ldots,i_D,h}}^*(\lambda)$ as described in 1(a)iii.

The above algorithm was implemented in R (R Development Core Team, 2007) using the MCMC package written by Martin and Quinn (2005).[2] The algorithm is run for I iterations with I as large as needed to obtain MCMC convergence. This algorithm is an extension of the MCMC scheme of Macaro (2010) that allows us to consider factorial time series data. From the computational point of view this is rather complex; and, so, the computation time required to do the analysis increases with the number of factors, the number of levels within each factor, and the number of time series within each combination of factors. The proposed candidate values in Steps 1(a)i–ii and 2(a)i–ii are sampled from uniform distributions and those in Steps 1(a)iii and 2(a)iii are sampled from inverse-Gamma distributions whose moments are functions of the old parameter values (see Macaro, 2010 for further details). In addition, the proposed values are accepted or rejected according to Metropolis–Hastings schemes.

4. Simulation Studies

4.1. Baseline Plus Single Factor Model

A data set with a total of four time series of length $T = 500$ was simulated from the one-way model described in Section 2.2.1. The priors are chosen to be Beta kernels. Specifically, we chose a low frequency band-pass kernel for the baseline, $q_{\beta_1^{(1)}}(\lambda) \sim \text{Beta}(1, 20)$, and an approximately periodic kernel for the two levels of the remaining factor $q_{\beta_1^{(2)}}(\lambda) \sim \text{Beta}(50, 20)$ and $q_{\beta_2^{(2)}}(\lambda) \sim \text{Beta}(50, 20)$. Finally, $q_{\varepsilon_{i_1,i_2,h}}(\lambda) \sim \text{Beta}(1, 1)$ implies a white-noise assumption for the error term. The parameters controlling the flexibility of the representation are set to $K_{i_1}^1 = 40$, $K_{i_2}^2 = 40$ and $K_{i_1,i_2,h}^3 = 1$. This prior choice empowers the belief that the main factors must not capture white-noise effects; and, so, their spectra are not expected to be flat a priori. In addition, the parameters which control for the variances of the Dirichlet distributions are set to $M_{i_d}^d = 1$ and $M_{\epsilon_\cdot} = 1$.

[2]The code is available upon request. Please email C. Macaro at christianmacaro@gmail.com.

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

Figure 3 shows the prior and posterior distributions (gray and black areas, respectively), as well as estimators based on the smoothed periodograms for the spectral densities of the observed time series and the unobserved processes $\beta_1^{(1)}(t)$, $\beta_1^{(2)}(t)$, and $\beta_2^{(2)}(t)$. Note that, in spite of the fact that the priors on the spectral densities of $\beta_1^{(2)}(t)$ and $\beta_2^{(2)}(t)$ had modes away from the correct periods, the estimated posterior spectral densities adequately capture the frequencies around 2.0 and 2.6.

4.2. Two-Way Models

We analyze time series data simulated from the two-way model described in Section 2.2.2 with $d = 2$, $i_1 = 1:2$, $i_2 = 1:2$, $h = 1:2$, and $T = 500$. We consider two alternative ways of selecting the prior distributions in this example. First we assume that the spectral density of the first level of the first factor, $\beta_1^{(1)}(t)$, as well as the spectral density of the $\varepsilon_{\cdot,\cdot,\cdot}(t)$ processes are constant over all the frequencies. This is consistent with the assumption that these processes are white-noise processes. In addition, we assume weak and flat priors on the spectral densities of the remaining time series processes. Specifically, we have $q_{\beta_2^{(1)}}(\lambda) \sim \text{Beta}(1.3, 1.3)$, $q_{\beta_1^{(2)}}(\lambda) \sim \text{Beta}(1.3, 1.3)$, $q_{\beta_2^{(2)}}(\lambda) \sim \text{Beta}(1.3, 1.3)$. Notice that we purposely chose these kernels to be nearly flat to contrast the boundary bias (see Zhang & Karunamuni, 2010) which seems to affect weak and flat priors. The parameters controlling the flexibility of the representation are set again to $K_{i_1}^1 = 40$, $K_{i_2}^2 = 40$ and $K_{i_1,i_2,h}^3 = 1$. The parameters that control the variances of the Dirichlet distributions are set to $M_{i_d}^d = 1$ and $M_{\epsilon_\cdot} = 1$.

By fixing the spectral densities of one of the levels of one of the factors and the spectral densities of the error term processes, identification of the remaining components is achieved, as shown in Figures 4 and 5.

The second scenario does not fix the spectra of any of the factor processes. It is assumed that the error processes have white-noise spectra a priori, and slightly more informative priors on the factor processes are considered. In particular we assume band-pass filter types of priors that emphasize smaller frequencies for $\beta_1^{(1)}(t)$ and $\beta_2^{(1)}(t)$ and larger frequencies for $\beta_1^{(2)}(t)$ and $\beta_2^{(2)}(t)$. The graphs in Figures 6 and 7 show that the implementation of a model with these band-pass filter priors produces results similar to those obtained under the white-noise assumption on $\beta_1^{(1)}(t)$ considered above and adequately captures the data structure.

4.3. Further Identification Issues in Two-Way Models

In order to illustrate some additional features related to prior elicitation and corresponding posterior inference, we consider a larger simulated data set. Specifically, a total of 12 time series were simulated from a two-way model with the following structure:

$$y_{i_1,i_2,h}(t) = \alpha_{i_1}^{(1)}(t) + \alpha_{i_2}^{(2)}(t) + \varepsilon_{i_1,i_2,h}(t), \tag{24}$$

for $t = 1:500$, $i_1 = 1:2$, $i_2 = 1:2$, and $l = 1:3$, where

$$\alpha_1^{(1)}(t) \sim \text{AR}\big[2, (0.95, 9), 0.5\big], \qquad \alpha_2^{(1)}(t) \sim \alpha_1^{(1)}(t) + \text{AR}\big[2, (0.95, 3), 0.5\big], \tag{25}$$

$$\alpha_1^{(2)}(t) \sim \text{AR}[1, 0.86, 0.5] + \nu_1(t), \qquad \alpha_2^{(2)}(t) \sim \text{AR}[1, 0.94, 0.5] + \nu_2(t), \tag{26}$$

with $\varepsilon_{i_1,i_2,h}(t) \sim N(0, 0.5)$ for all i_1, i_2 and l, $\nu_1(t) \sim N(0, 0.5)$, and $\nu_2(t) \sim N(0, 0.5)$. In the notation above, AR[2, (0.95, 9), 0.5] denotes a quasiperiodic autoregressive process of order two, or AR(2), with a pair of complex reciprocal roots with modulus 0.95, wavelength 9—or equivalently, frequency 0.698—and standard deviation 0.5. Similarly, AR[2, (0.95, 3), 0.5] denotes a quasiperiodic AR(2) process with a pair of complex reciprocal roots with modulus

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

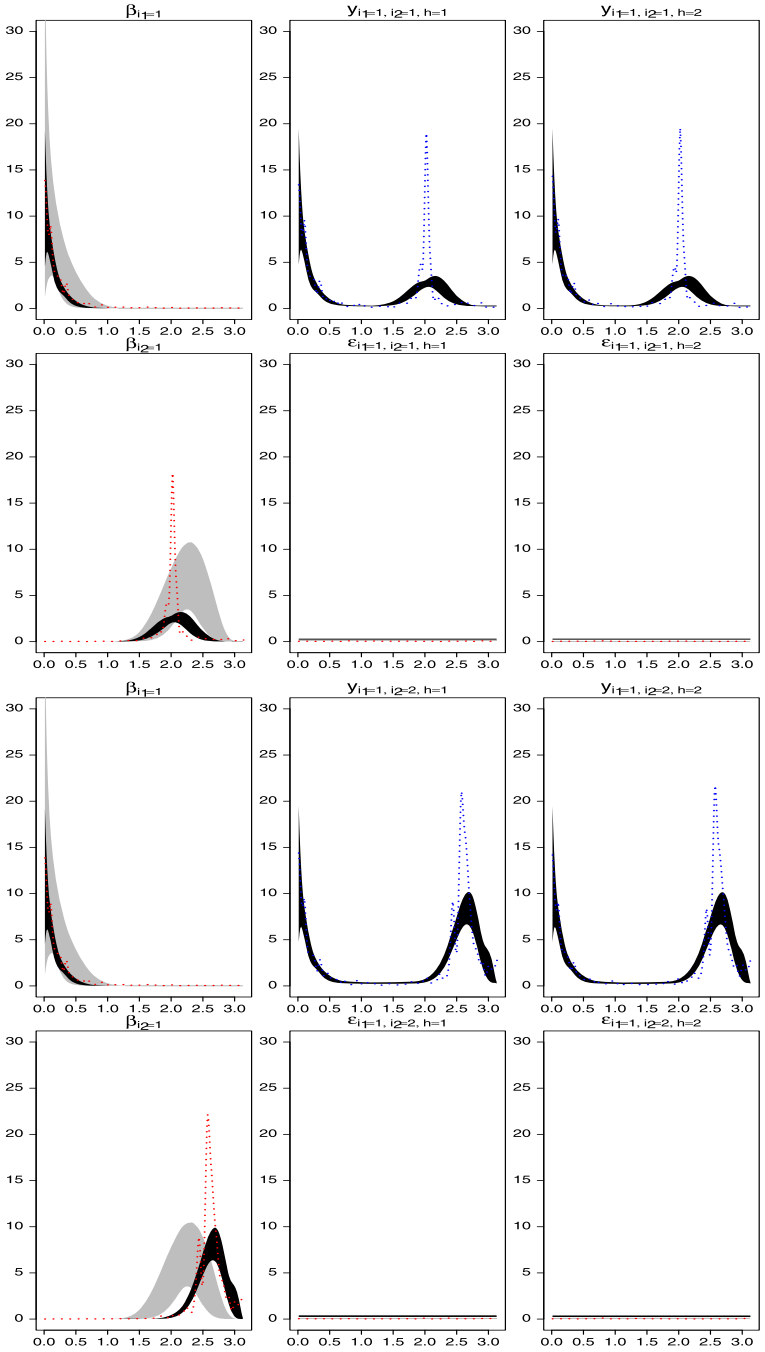


FIGURE 3.

Spectral analysis of time series data simulated from the one-way model described in Section 2.2.1. *Gray areas* are central 95 % spectral prior distributions, while black areas are central 95 % spectral posterior distributions. The *dotted lines* represent estimators of the spectral densities obtained by smoothing the periodograms of the data. *First row*: estimated spectral densities of $\beta_1^{(1)}(t)$, $y_{1,1,1}(t)$, and $y_{1,1,2}(t)$. *Second row*: estimated spectral densities of $\beta_1^{(2)}(t)$, $\varepsilon_{1,1,1}(t)$, and $\varepsilon_{1,1,2}(t)$. *Third row*: estimated spectral densities of $\beta_1^{(1)}(t)$, $y_{1,2,1}(t)$, and $y_{1,2,2}(t)$. *Fourth row*: estimated spectral densities of $\beta_2^{(2)}(t)$, $\varepsilon_{1,2,1}(t)$ and $\varepsilon_{1,2,2}(t)$.

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

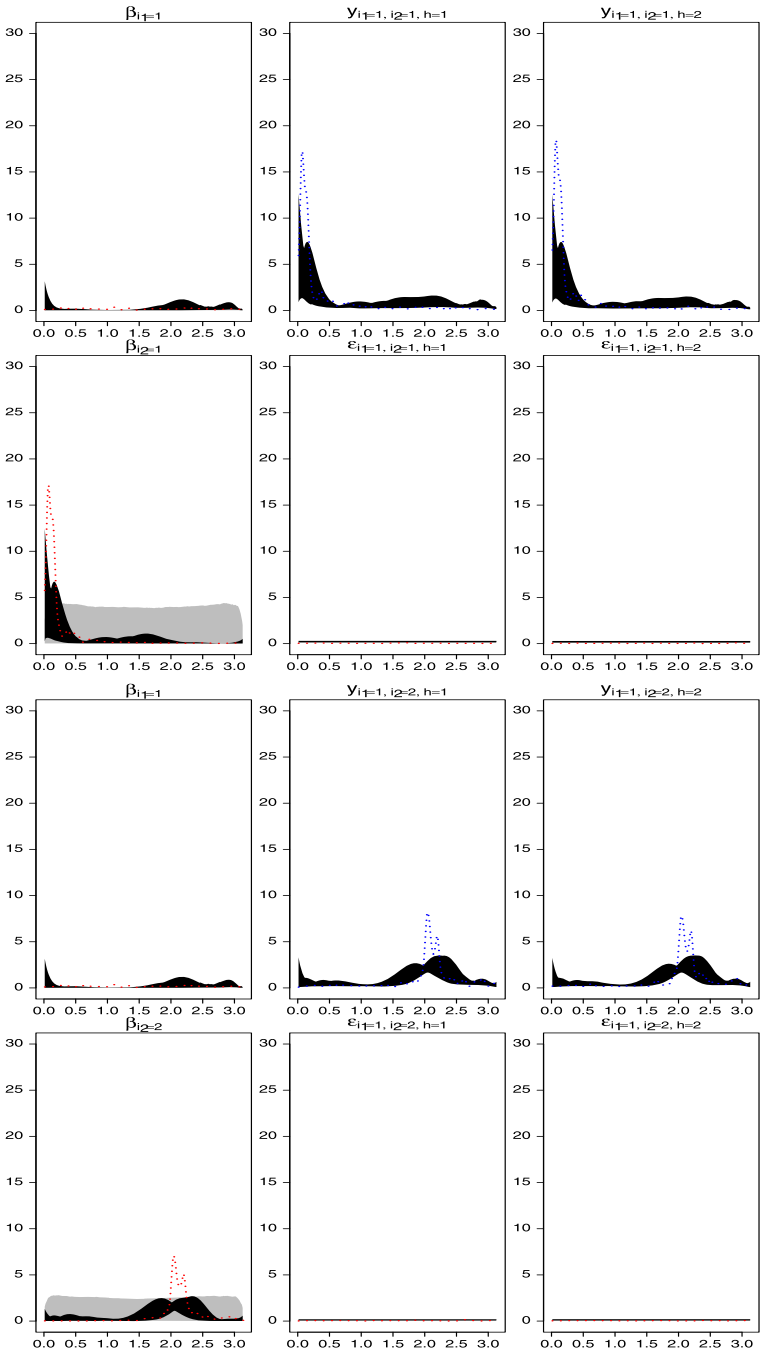


FIGURE 4.
Spectral analysis of time series data simulated from the two-way model described in Section 2.2.2. *Gray areas* are central 95 % spectral prior distributions, while *black areas* are central 95 % spectral posterior distributions. The *dotted lines* represent estimators of the spectral densities obtained by smoothing the periodograms of the data. *First row*: estimated spectral densities of $\beta_1^{(1)}(t)$, $y_{1,1,1}(t)$, and $y_{1,1,2}(t)$. *Second row*: estimated spectral densities of $\beta_1^{(2)}(t)$, $\varepsilon_{1,1,1}(t)$, and $\varepsilon_{1,1,2}(t)$. *Third row*: estimated spectral densities of $\beta_1^{(1)}(t)$, $y_{1,2,1}(t)$, and $y_{1,2,2}(t)$. *Fourth row*: estimated spectral densities of $\beta_2^{(2)}(t)$, $\varepsilon_{1,2,1}(t)$ and $\varepsilon_{1,2,2}(t)$.

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

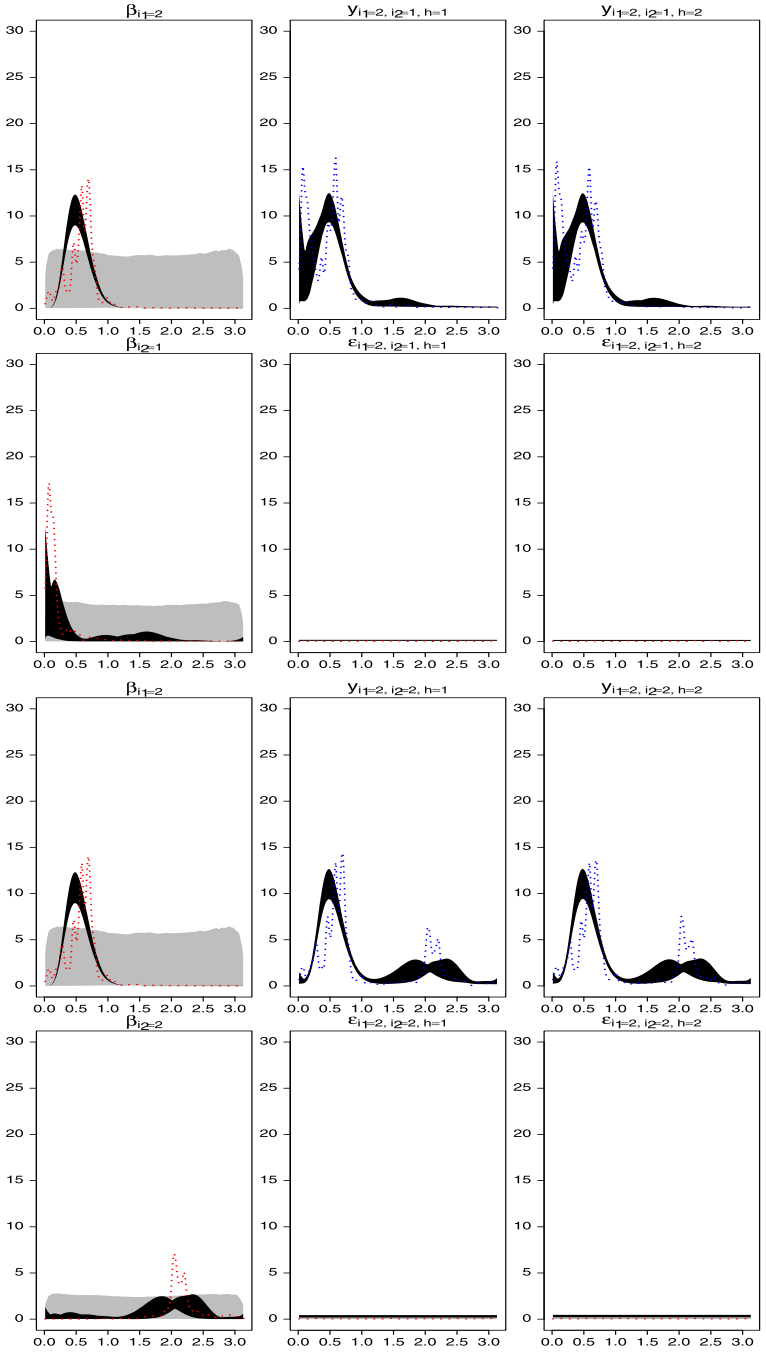


FIGURE 5.

Spectral analysis of time series data simulated from the two-way model described in Section 2.2.2. *Gray areas* represent central 95 % spectral prior distributions, while *black areas* are central 95 % spectral posterior distributions. The *dotted lines* represent estimators of the spectral densities obtained by smoothing the periodograms of the data. *First row*: estimated spectral densities of $\beta_2^{(1)}(t)$, $y_{2,1,1}(t)$, and $y_{2,1,2}(t)$. *Second row*: estimated spectral densities of $\beta_1^{(2)}(t)$, $\varepsilon_{2,1,1}(t)$, and $\varepsilon_{2,1,2}(t)$. *Third row*: estimated spectral densities of $\beta_2^{(1)}(t)$, $y_{2,2,1}(t)$, and $y_{2,2,2}(t)$. *Fourth row*: estimated spectral densities of $\beta_2^{(2)}(t)$, $\varepsilon_{2,2,1}(t)$ and $\varepsilon_{2,2,2}(t)$.

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

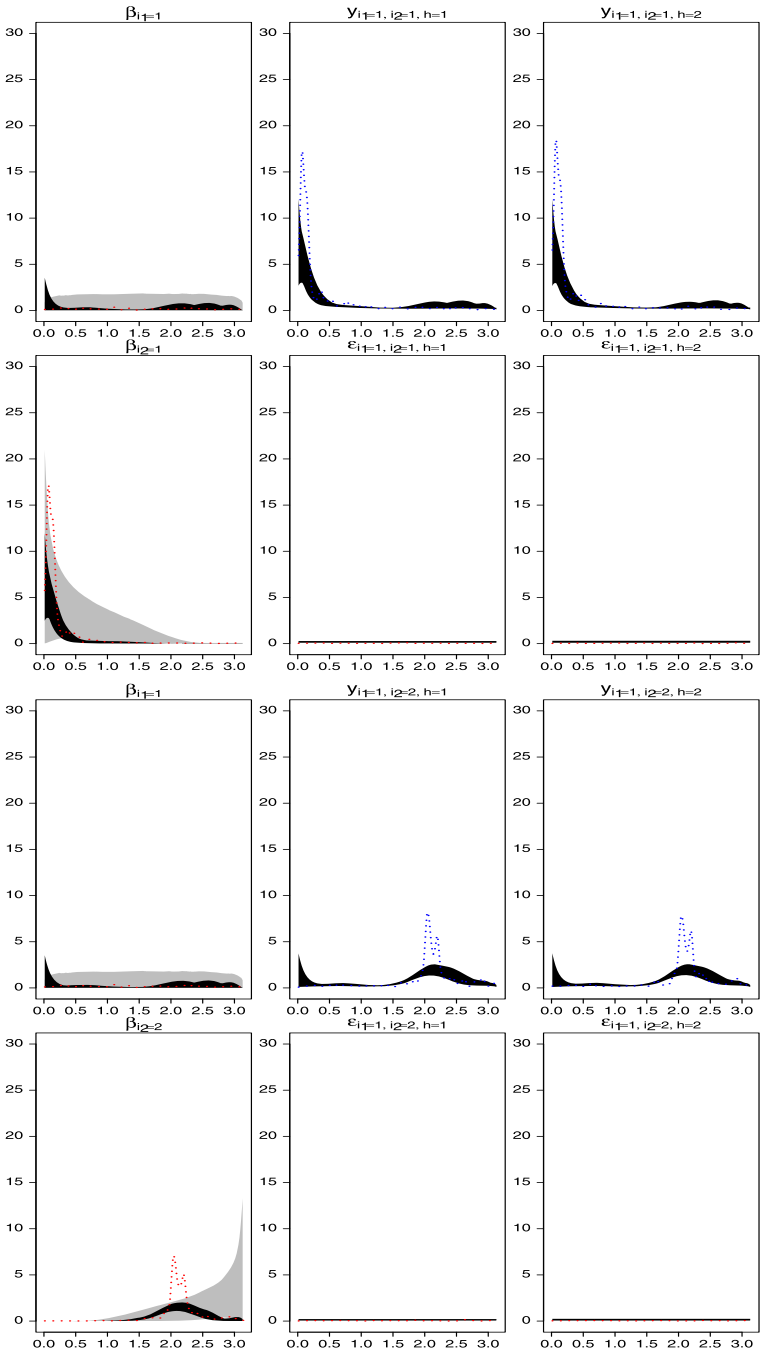


FIGURE 6.

Spectral analysis of time series data simulated from the two-way model described in Section 2.2.2. *Gray areas* represent central 95 % spectral prior distributions, while *black areas* are central 95 % spectral posterior distributions. The *dotted lines* represent estimators of the spectral densities obtained by smoothing the periodograms of the data. *First row*: estimated spectral densities of $\beta_1^{(1)}(t)$, $y_{1,1,1}(t)$, and $y_{1,1,2}(t)$. *Second row*: estimated spectral densities of $\beta_1^{(2)}(t)$, $\varepsilon_{1,1,1}(t)$, and $\varepsilon_{1,1,2}(t)$. *Third row*: estimated spectral densities of $\beta_1^{(1)}(t)$, $y_{1,2,1}(t)$, and $y_{1,2,2}(t)$. *Fourth row*: estimated spectral densities of $\beta_2^{(2)}(t)$, $\varepsilon_{1,2,1}(t)$ and $\varepsilon_{1,2,2}(t)$.

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

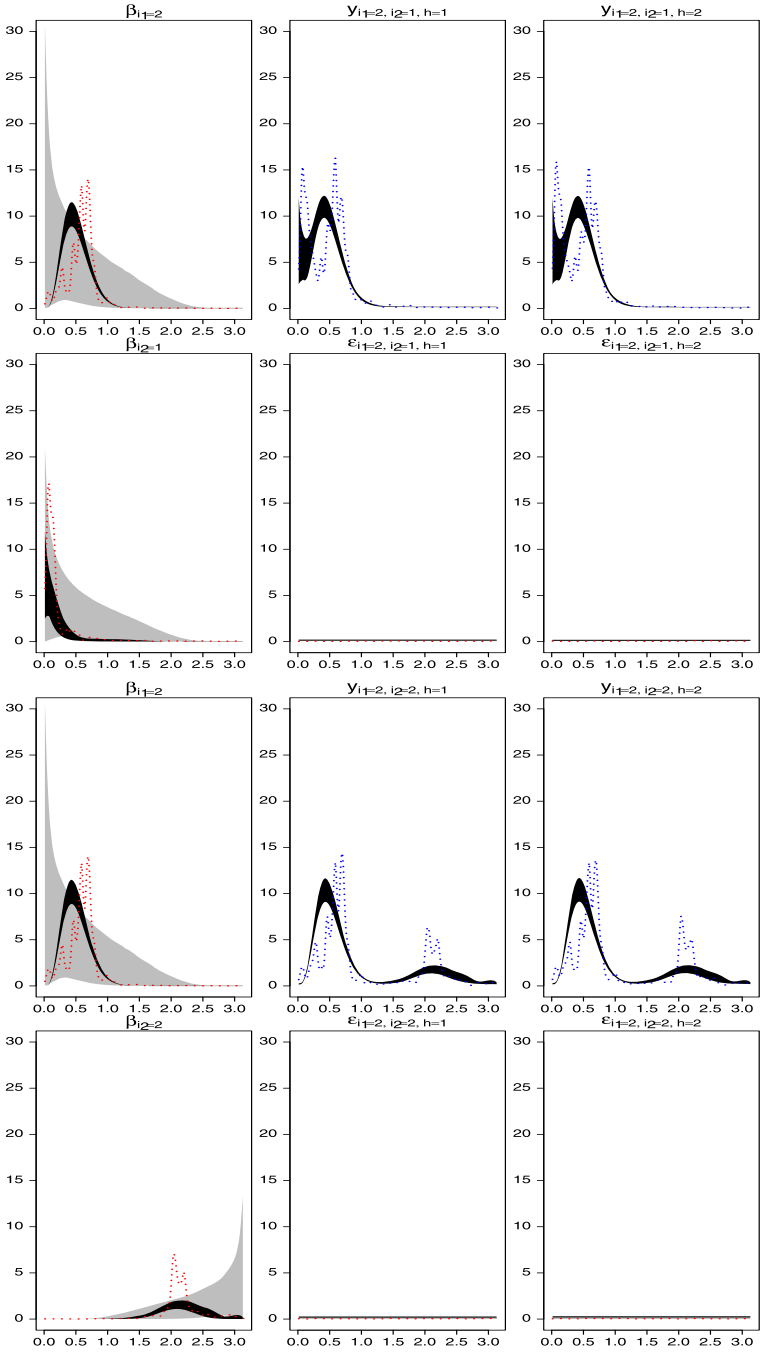


FIGURE 7.

Spectral analysis of time series data simulated from the two-way model described in Section 2.2.2. *Gray areas* represent central 95 % spectral prior distributions, while *black areas* are central 95 % spectral posterior distributions. The *dotted lines* represent estimators of the spectral densities obtained by smoothing the periodograms of the data. *First row*: estimated spectral densities of $\beta_2^{(1)}(t)$, $y_{2,1,1}(t)$, and $y_{2,1,2}(t)$. *Second row*: estimated spectral densities of $\beta_1^{(2)}(t)$, $\varepsilon_{2,1,1}(t)$, and $\varepsilon_{2,1,2}(t)$. *Third row*: estimated spectral densities of $\beta_2^{(1)}(t)$, $y_{2,2,1}(t)$, and $y_{2,2,2}(t)$. *Fourth row*: estimated spectral densities of $\beta_2^{(2)}(t)$, $\varepsilon_{2,2,1}(t)$ and $\varepsilon_{2,2,2}(t)$.

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

0.95, wavelength 3—or frequency 2.094—and standard deviation 0.5; AR[1, 0.86, 0.5] denotes an AR(1) process with a reciprocal root of modulus 0.86 and standard deviation 0.5; and finally, AR[1, 0.94, 0.5] denotes an AR(1) with a reciprocal root of modulus 0.94 and standard deviation 0.5.

The structure above implies that the spectra of the series $y_{1,1,h}(t)$ and $y_{1,2,h}(t)$ will have a peak at the frequency $\omega_1 = 0.698$, in addition to the peak at zero related to the AR(1) components; while those of the series $y_{2,1,h}(t)$ and $y_{2,2,h}(t)$ will have an additional peak at the frequency $\omega_2 = 2.094$. We proceed to analyze these data with the Bayesian spectral non-parametric approach described in the previous sections. In particular, we assume the following underlying structure on the observed data:

$$y_{i_1,i_2,h}(t) = \beta_{i_1}^{(1)}(t) + \beta_{i_2}^{(2)}(t) + \varepsilon_{i_1,i_2,h}(t).$$

Furthermore, we assume that, a priori, the spectra of $\beta_{i_1}^{(1)}(t)$ for $i_1 = 1:2$ have a single peak around the frequency $\omega_1 = 0.698$, the spectra of $\beta_{i_2}^{(2)}(t)$ for $i_2 = 1:2$ have only a peak at zero, and the spectra of $\varepsilon_{i_1,i_2,h}(t)$ are flat.

Figures 8 and 9 show the prior structure (gray areas). Note that this prior is closely describing the structure underlying the processes $y_{1,1,h}(t)$ and $y_{1,2,h}(t)$, but it misspecifies the structure of the processes underlying $y_{2,1,h}(t)$ and $y_{2,2,h}(t)$, completely missing the second peak at $\omega_2 = 2.094$.

Note that the error components, assumed to have flat spectra a priori under a white-noise structure, capture, a posteriori, whatever is left unexplained by the factor structure. Therefore, if an informative prior used on the spectra of a given factor misses certain frequency components present in the data, such components will be captured by the posterior distributions of the spectra of the idiosyncratic components, if relatively noninformative priors were given to these components. This is confirmed by the results shown in Figures 8 and 9. The prior distribution for $\beta_2^{(1)}(t)$ does not include one of the frequency components in the true $\alpha_2^{(1)}(t)$ and the posterior distribution of $\beta_2^{(1)}(t)$ does not capture the peak at $\omega_2 = 2.094$. Note, however, that this peak clearly appears in the estimated spectra of the six idiosyncratic error terms $\varepsilon_{2,1,h}(t)$ and $\varepsilon_{2,2,h}(t)$ for $h = 1:3$, indicating that such structure is missing from the factorial decomposition of the process.

The analysis above that assumes informative priors on some of the unobserved components is performed in order to illustrate a couple of features about the spectral decomposition. First, the use of informative priors allow us to choose a particular decomposition of the series where some of the components may have a specific scientifically interpretable meaning a priori. For example, in this simulation study, we use a strong informative prior on the spectrum of $\beta_2^{(1)}(t)$ that identifies this time series process as one with activity in a particular frequency band. Alternatively, we could have used a more standard time-domain two-way representation of the data by assuming that the spectrum of one of the levels of one of the two factors, say the spectrum of $\beta_1^{(1)}(t)$, is flat (consistent with a white-noise prior assumption on this process). Such representation would also imply strong assumptions on the components that would not lead to the true unobserved representation in (24), (25), and (26), but that would allow us to adequately capture the structure of the spectra of the observed time series a posteriori. Researchers can therefore explore several spectral decompositions of the data by making different assumptions on the unobserved components, as long as such assumptions guarantee identifiability. The second model feature that we want to emphasize is that, regardless of the identifiability conditions used on the spectral characterization of the factors $\beta_{i_d}^{(d)}(t)$, the model will be able to adequately describe the structure of the data a posteriori if the priors on the spectra of $\varepsilon_{i_1,i_2,h}(t)$ are assumed to be flat and diffuse. As seen in the simulation study above, none of the priors on the spectra of $\beta_{i_1}^{(d)}(t)$

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

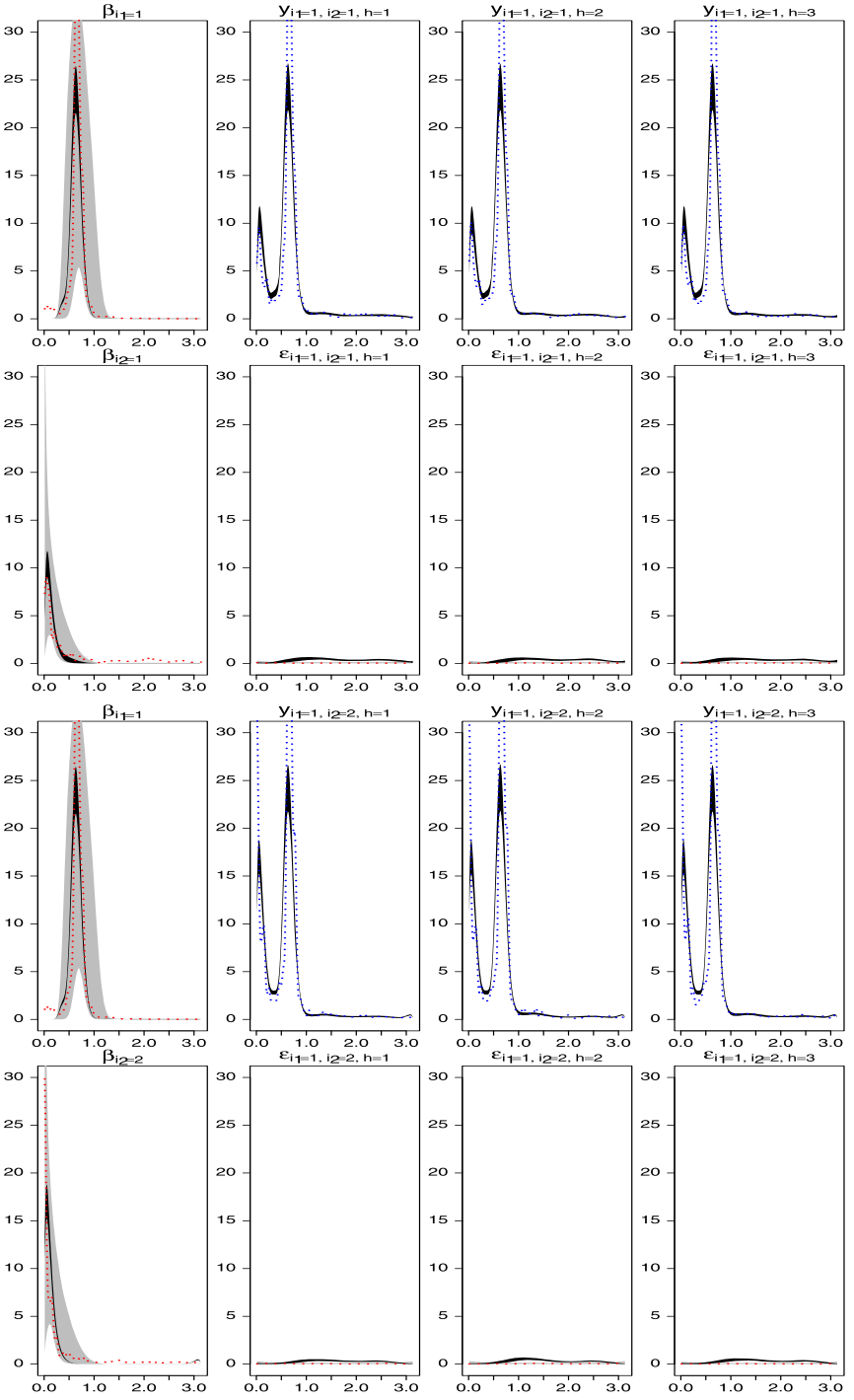


FIGURE 8.
Spectral analysis of the time series data simulated from the model described in Section 4.3. *Gray areas* represent central 95 % of the prior distributions. *Black areas* represent central 95 % of the posterior distributions from the two-way model. The *dotted lines* represent periodogram-based estimators of the spectral densities of the individual series without considering the factor structure.

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

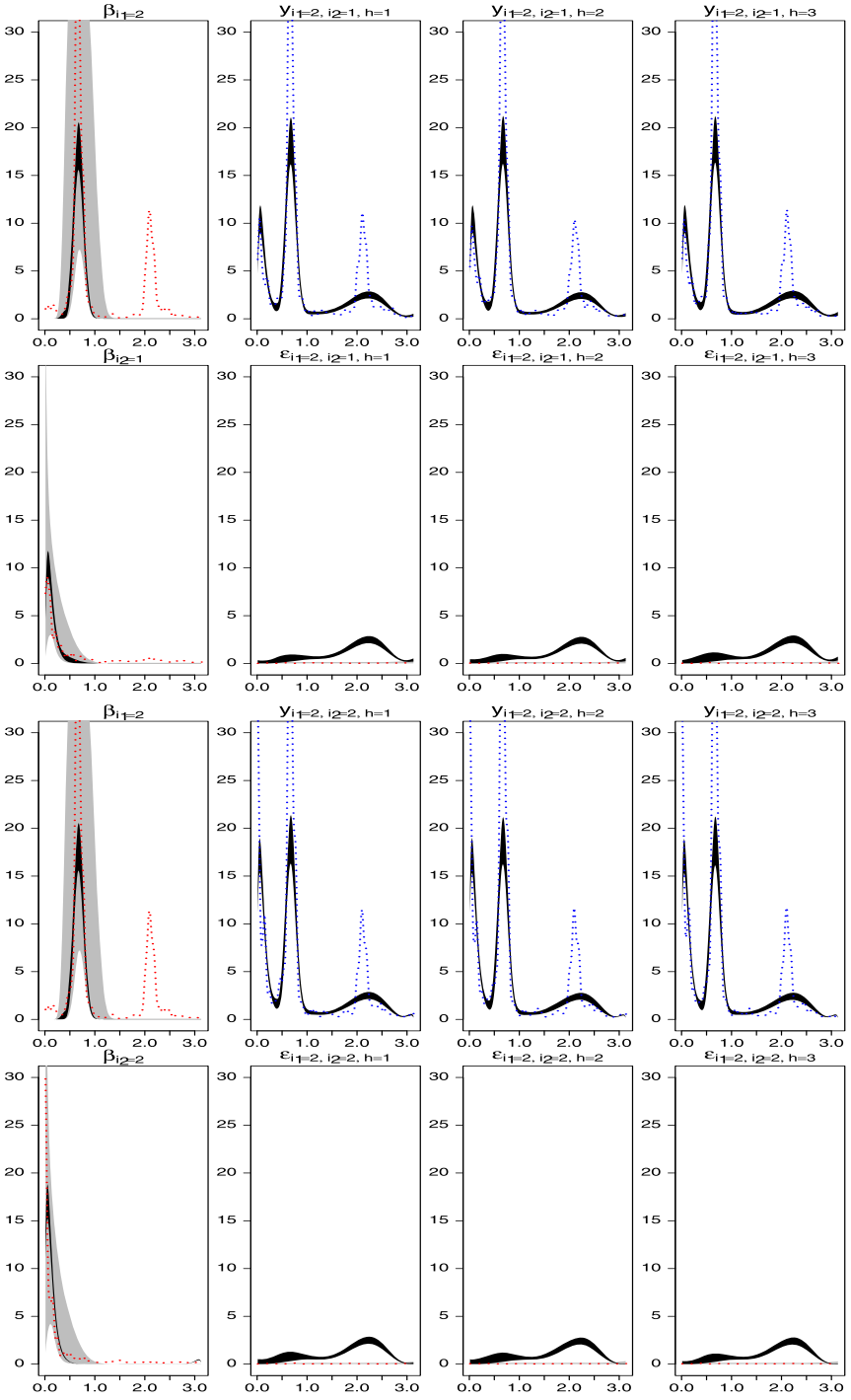


FIGURE 9.
Spectral analysis of the time series data simulated from the model described in Section 4.3. *Gray areas* represent central 95 % of the prior distributions. *Black areas* represent central 95 % of the posterior distributions from the two-way model. The *dotted lines* represent periodogram-based estimators of the spectral densities of the individual series without considering the factor structure.

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

and $\varepsilon_{i_1,i_2,h}(t)$ had peaks on frequencies above 1.5, however, the estimated posterior spectra of $\varepsilon_{2,1,h}(t)$, $\varepsilon_{2,2,h}(t)$, and, therefore, the estimated posterior spectra of $y_{2,1,h}(t)$ and $y_{2,2,h}(t)$, show a peak at the $\omega_2 = 2.094$ frequency that does not appear in any of the spectra of the $y_{i_1,i_2,h}(t)$ series.

5. Functional Magnetic Resonance Imaging Data

We consider the analysis of a data set from an experiment of Antognini, Buonocore, Disbrow, and Carstens (1998) in which functional magnetic resonance imaging (fMRI) was used to examine pain perception in humans. The data, previously analyzed in Stoffer (1999) and Shumway and Stoffer (2006), consist of consecutive measures of blood oxygenation level dependent (BOLD) signal intensities at nine locations in brain. The signals were measured in 26 awake and mildly anesthetized individuals who were presented with three types of periodic stimuli: brushing, heat, and shock. The stimuli were applied alternately for 32 seconds, with a sampling rate of one point every two seconds, and then stopped for 32 seconds.

Data were recorded by several sensors located on different parts of the brain, namely, Cortex 1, 2, 3 and 4, Caudate, Thalamus 1 and 2, and Cerebellum 1 and 2. Here we analyze only those related to the region Cortex 1. The 26 patients were divided into six groups depending on the type of stimulus they received (*Brush*, *Heat*, or *Shock*) and on whether they were awake or mildly anesthetized. Specifically, we have the following: (1) data from three patients in *Low* sedation state who were stimulated with a *Brush*; (2) data from five patients in *Low* sedation state stimulated by *Heat*; (3) data from four patients in *Low* sedation state stimulated by a *Shock*; (4) data from five *Awake* patients recorded while they were stimulated with a *Brush*; (5) data from four *Awake* patients stimulated by *Heat*; and finally, (6) data from five *Awake* patients stimulated by a *Shock*. Given this structure, a two-way model with one factor associated to the level of consciousness (*Low* and *Awake*) and another factor associated to the stimulus type (*Brush*, *Heat*, and *Shock*) was chosen to describe these data. Therefore, we consider a two-way model of the form

$$y_{i_1,i_2,h}(t) = \beta^{(1)}_{i_1}(t) + \beta^{(2)}_{i_2}(t) + \varepsilon_{i_1,i_2,h}(t),$$

where $\beta^{(1)}_{i_1}(t)$ for $i_1 = 1:2$ model the effects related to the consciousness level, $\beta^{(2)}_{i_2}(t)$ for $i_2 = 1:3$ model the effects of the type of stimulus, and each $\varepsilon_{i_1,i_2,h}(t)$ is the idiosyncratic component modeling those effects specific to each patient that are not captured by the factor structure. Since the number of patients varies within each combination of consciousness level and stimulus type, we have $y_{1,1,h}(t)$ (*Low* and *Brush*) for $h = 1:3$; $y_{1,2,h}(t)$ (*Low* and *Heat*) for $h = 1:5$; $y_{1,3,h}(t)$ (*Low* and *Shock*) for $h = 1:4$; $y_{2,1,h}(t)$ (*Awake* and *Brush*) for $h = 1:5$; $y_{2,2,h}(t)$, (*Awake* and *Heat*) for $h = 1:4$; and $y_{2,3,h}(t)$ (*Awake* and *Shock*) for $h = 1:5$.

The priors are chosen to be Beta kernels. Specifically, we chose $q_{\beta^{(1)}_1}(\lambda) \sim \text{Beta}(1.7, 8.6)$; $q_{\beta^{(1)}_2}(\lambda) \sim \text{Beta}(1.7, 10.6)$; $q_{\beta^{(2)}_1}(\lambda) \sim \text{Beta}(1.7, 1.1)$; $q_{\beta^{(2)}_2}(\lambda) \sim \text{Beta}(1.7, 1.1)$; $q_{\beta^{(2)}_3}(\lambda) \sim \text{Beta}(1.7, 1.1)$; and $q_{\varepsilon_{i_1,i_2,h}}(\lambda) \sim \text{Beta}(1, 1)$ for all i_1, i_2 and h. The idea is that, a priori, the $\beta^{(1)}_{i_1}(t)$ processes aim to describe low frequency components, the $\beta^{(2)}_{i_2}(t)$ aim to describe high frequency components and $\varepsilon_{i_1,i_2,h}(t)$ are white-noise processes. For the remaining parameters, we set $K^1_{i_1} = 30$ (cyclical), $K^2_{i_2} = 100$ (long term) and $K^3_{i_1,i_2,h} = 10$ (short-term/white noise) for $i_1 = 1:2$, $i_2 = 1:3$ and $h = [(1:3), (1:5), (1:4), (1:5), (1:4), (1:5)]$. The parameters that control the variances of the Dirichlet distributions were set to $M^d_{i_d} = 1$ and $M_{\epsilon\cdot} = 1$.

Figure 10 shows the prior and posterior distributions (light and dark gray areas, respectively) of the spectra of the factors. Figure 11 shows the prior and posterior distributions (light and

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

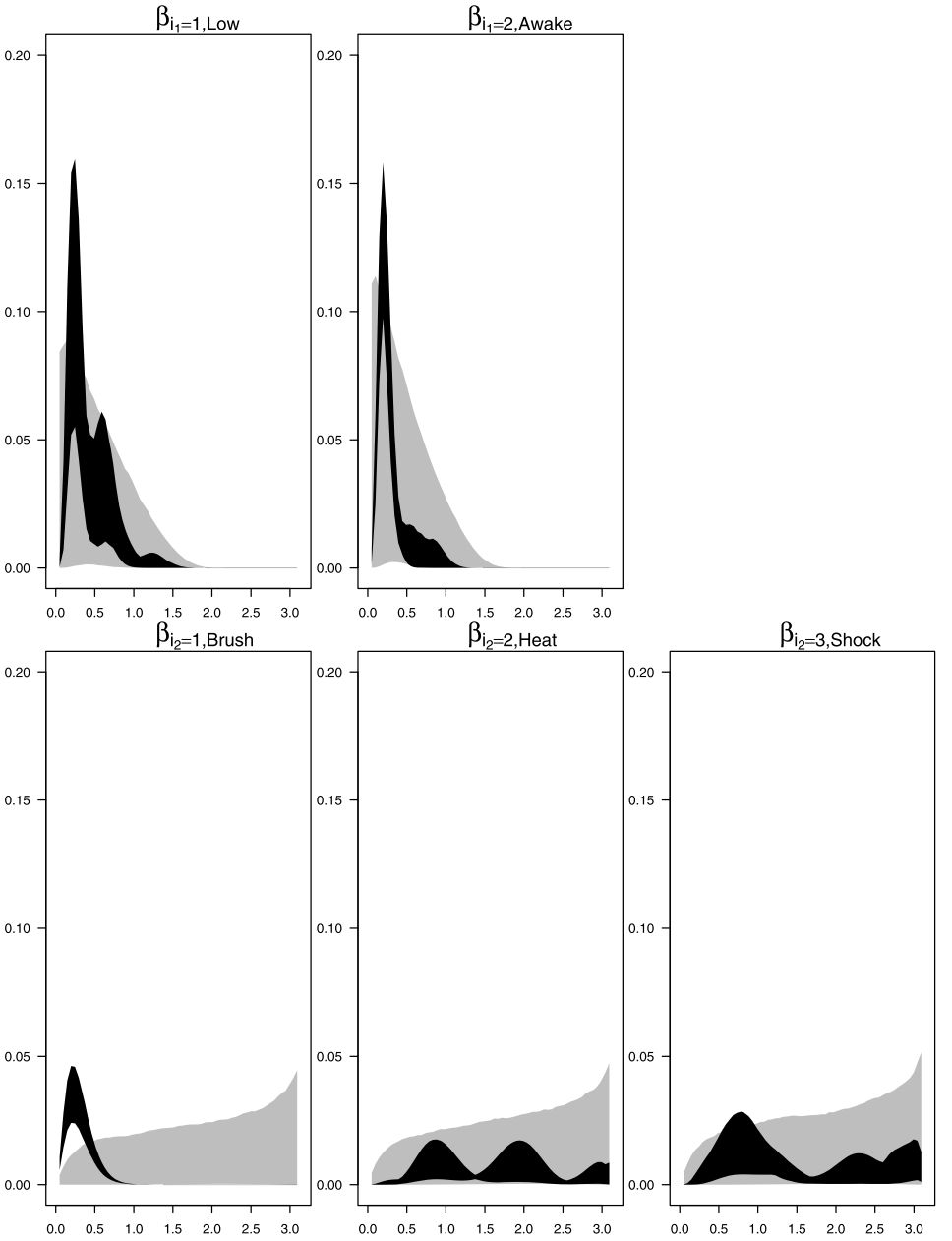


FIGURE 10.
Spectral analysis of the fMRI data. The plots display the spectra of the factor processes. The *light gray areas* represent central 95 % of the prior distributions and the *dark gray areas* represent central 95 % of the posterior distributions.

dark gray areas, respectively) of the spectra of the idiosyncratic processes for each of the 26 individuals.

From the graphs we can see that the posterior distributions of the spectral density for the process $\beta_1^{(1)}(t)$ related to *Low* sedation level and that for the process $\beta_1^{(2)}(t)$ related to *Awake* consciousness level show a marked peak around 0.19–0.22 which, given the sampling rate of the

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

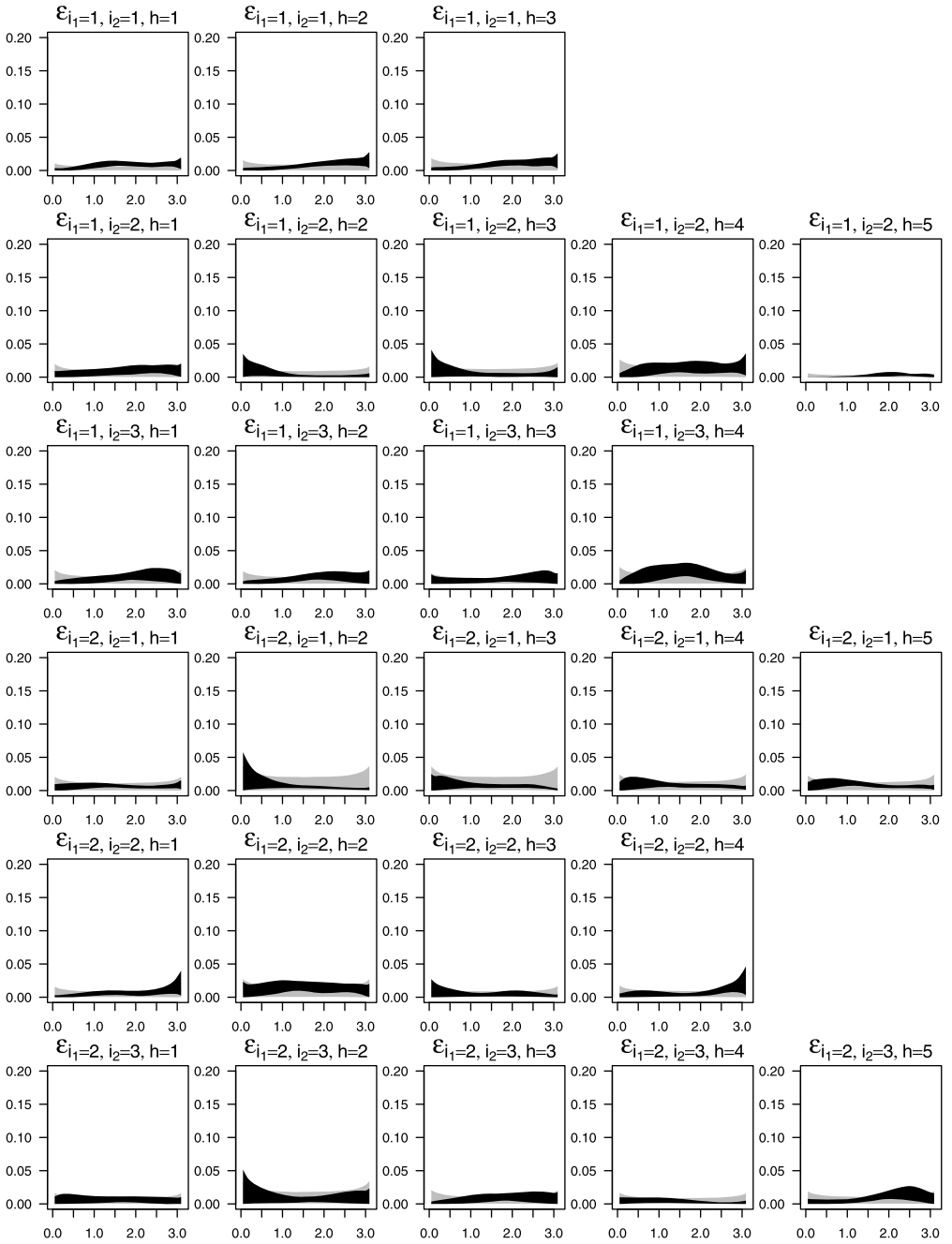


FIGURE 11.
Spectral analysis of the fMRI data. The plots show the spectra of the idiosyncratic components for each of the 26 subjects. The *light gray areas* represent central 95 % of the prior distributions. The *dark gray areas* represent central 95 % of the posterior distributions.

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

data, corresponds to the 64-second period band (or 1/64 cycles per second). The posterior spectral density of $\beta_1^{(1)}(t)$ (*Low*) shows more uncertainty in this frequency band than the posterior spectral density of $\beta_2^{(1)}(t)$ (*Awake*). In addition, the spectral density of $\beta_1^{(1)}(t)$ (*Low*) shows additional activity in higher frequency bands. Similarly, comparing the posterior estimates of the spectral densities of the processes $\beta_1^{(2)}(t)$, $\beta_2^{(2)}(t)$ and $\beta_3^{(2)}(t)$ associated, respectively, with the *Brush*, *Heat* and *Shock* stimuli, we see that the density for *Brush* shows a marked peaked around the 0.19–0.22 (again, the 64-second period band) that increases the power at this frequency in the estimated spectra of the subjects who received this type stimulus. For the other two types of stimuli, rather small peaks appear at higher frequencies. In summary, the posterior results based on the Bayesian non-parametric analysis for the data collected at the Cortex 1 location indicate that there is a significant difference in the power spectra at the 1/64 frequency band for patients that received the *Brush* stimulus.

6. Conclusions

This work presents a Bayesian non-parametric framework for the analysis of multiple time series that are collected in factorial experimental designs. This approach is based on representing the prior distributions on the spectral densities with Bernstein–Dirichlet priors. It is an important extension of the univariate methods of Macaro (2010) in that it allows us to combine data and prior distributions on the unobserved spectral components in the decomposition of multiple time series. In particular, users can explore various decompositions of the observed time series processes and, consequently, various spectral decompositions, which can provide useful representations of the data.

Because of the possibility of extracting unobserved factors from a set of time series, this work is related to the well known literature on dynamic factor models (Geweke, 1977; Sargent & Sims, 1977; Forni, Hallin, Lippi, & Reichlin, 2000; Stock & Watson, 2002, and many others). In our work we emphasize how the priors can be used to guarantee the identifiability of the different model components at the spectral level. Forni et al. (2000) and Stock and Watson (2002), on the other hand, derived their identification conditions from the properties of the principal component analysis.

We illustrate the main features and the performance of the Bayesian non-parametric approach in the analyses of simulated data and in functional magnetic resonance (fMRI) brain responses measured in awake and mildly anesthetized individuals who were presented with three types of stimuli. The posterior distributions on the specific spectral decomposition of the data that we chose—i.e., a decomposition in which each observed signal was represented as a sum of two unobserved components, one related to the level of consciousness and one to the stimulus, plus an idiosyncratic term—reveals differences in the power spectra at certain frequency bands for patients who received a particular stimulus. The spectral domain approaches presented here can be used to study data from a broad range of applications in which multiple time series are collected in the context of a designed experiment.

Acknowledgements

This material was based upon work partially supported by the National Science Foundation under grant DMS-0635449 to the Statistical and Applied Mathematical Sciences Institute. Raquel Prado also was partially supported by NSF grant SES-1060911. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

Antognini, J., Buonocore, M., Disbrow, E.A., & Carstens, E. (1998). Isoflurane anesthesia blunts cerebral responses to noxious and innocuous stimuli: a functional MRI study. *Life Sciences*, *61*, PL349–PL354.

Brillinger, D.R. (1973). The analysis of time series collected in an experimental design. In P.R. Krishnaiah (Ed.), *Multivariate analysis-III* (pp. 241–256). New York: Academic Press.

Brillinger, D.R. (1980). Analysis of variance and problems under time series models. In P.R. Krishnaiah & D.R. Brillinger (Eds.), *Handbook of statistics* (Vol. I, pp. 237–278). Amsterdam: North Holland.

Carter, C.K., & Kohn, R. (1997). Semiparametric Bayesian inference for time series with mixed spectra. *Journal of the Royal Statistical Society. Series B*, *59*, 255–268.

Choudhuri, N., Ghosal, S., & Roy, A. (2004). Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association*, *99*, 1050–1059.

DiMatteo, I., Genovese, C., & Kass, R. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika*, *88*(4), 1055–1071.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized factor model: identification and estimation. *Review of Economics and Statistics*, *82*, 540–554.

Gangopadhyay, A., Mallick, B., & Denison, D. (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.

Gelfand, A., & Kottas, A. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, *11*, 289–305.

Geweke, J. (1977). The dynamic factor analysis of economic time series. In D.J. Aigner & A.S. Goldberger (Eds.), *Latent variables in socio-economic models*. Amsterdam: North-Holland.

Karabatsos, G., & Walker, S. (2009). A Bayesian nonparametric approach to test equating. *Psychometrika*, *74*(2), 211–232.

Liseo, B., Marinucci, D., & Petrella, L. (2001). Bayesian semiparametric inference on long range dependence. *Biometrika*, *88*, 1089–1104.

Macaro, C. (2010). Bayesian non-parametric signal extraction for Gaussian time series. *Journal of Econometrics*, *157*(2), 381–395.

Martin, A., & Quinn, K. (2005). *Markov chain Monte Carlo (MCMCpack) package. R Foundation for Statistical Computing*. R package version 0.6-6.

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

Petrone, S. (1999b). Bayesian density estimation using Bernstein polynomials. *Canadian Journal of Statistics*, *27*, 105–126.

Prado, R. (2010). Multi-state models for mental fatigue. In A. O'Hagan & M. West (Eds.), *The handbook of applied Bayesian analysis*. Oxford: Oxford University Press.

R Development Core Team (2007). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. ISBN: 3-900051-07-0.

Rosen, O., & Stoffer, D. (2007). Automatic estimation of multivariate spectra via smoothing splines. *Biometrika*, *94*, 335–345.

Rosen, O., Stoffer, D., & Wood, S. (2009). Local spectral analysis via a Bayesian mixture of smoothing splines. *Journal of the American Statistical Association*, *104*, 249–262.

Sargent, T., & Sims, C. (1977). Business cycle modeling without pretending to have too much a-priori economic theory. In C. Sims et al. (Eds.), *New methods in business cycle research*. Minneapolis: Federal Reserve Bank of Minneapolis.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, *4*, 639–650.

Shumway, R.H. (1970). Applied regression and analysis of variance for stationary time series. *Journal of the American Statistical Association*, *65*, 1527–1546.

Shumway, R.H., & Stoffer, D. (2006). *Time series analysis and its applications: with R examples* (2nd ed.). New York: Springer.

Stock, J., & Watson, M. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*, 1167–1179.

Stoffer, D. (1999). Detecting common signals in multiple time series using the spectral envelope. *Journal of the American Statistical Association*, *94*, 1341–1356.

Whittle, P. (1957). Curve and periodogram smoothing. *Journal of the Royal Statistical Society. Series B*, *19*, 38–63.

Whittle, P. (1962). Gaussian estimation in stationary time series. *Bulletin of the International Statistical Institute*, *39*, 105–129.

Zhang, S., & Karunamuni, R. (2010). Boundary performance of the beta kernel estimators. *Journal of Nonparametric Statistics*, *22*(1).

Manuscript Received: 22 DEC 2010
Final Version Received: 4 SEP 2013
Published Online Date: 24 OCT 2013

https://doi.org/10.1007/s11336-013-9354-0 Published online by Cambridge University Press