Published in final edited form as:

J Comput Graph Stat. 2009 September 1; 18(3): 675–693. doi:10.1198/jcgs.2009.06109.

A Time-Frequency Functional Model for Locally Stationary Time Series Data

Li Qin [Assistant Member]

Statistical Center for HIV/AIDS Research and Prevention, Fredb Hutchinson Cancer Research Center, Seattle, WA 98109 (lqin@fhcrc.org)

Wensheng Guo [Associate Professor]

Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA 19104 (wguo@cceb.upenn.edu)

Brian Litt [Associate Professor]

Department of Neurology and Bioengineering, University of Pennsylvania, Philadelphia, PA 19104 (littb@mail.med.upenn.edu)

Abstract

Unlike traditional time series analysis that focuses on one long time series, in many biomedical experiments, it is common to collect multiple time series and focus on how the design covariates impact the patterns of stochastic variation over time. In this article, we propose a time-frequency functional model for a family of time series indexed by a set of covariates. This model can be used to compare groups of time series in terms of the patterns of stochastic variation and to estimate the covariate effects. We focus our development on locally stationary time series and propose the covariate-indexed locally stationary setting, which include stationary processes as special cases. We use smoothing spline ANOVA models for the time-frequency coefficients. A two-stage procedure is introduced for estimation. To reduce the computational demand, we develop an equivalent state space model to the proposed model with an efficient algorithm. We also propose a new simulation method to generate replicated time series from their design spectra. An epileptic intracranial electroencephalogram (IEEG) dataset is analyzed for illustration.

Keywords

EEG; Epilepsy; Functional data analysis; Seizure; Smoothing spline; SS ANOVA; State space model

1. INTRODUCTION

Traditional time series analysis emphasizes the estimation and inference mainly for a single univariate or multivariate time series (e.g., Shumway and Stoffer 2000; Wei 2005). However, in biomedical studies, the experimental data usually include many time series and the analysis is usually focused on how the design covariates regulate the patterns of stochastic variation. The analysis of the dependence of the mean response on covariates and experimental treatments has been studied extensively in the framework of longitudinal data analysis (e.g., Diggle, Liang, and Zeger 1994). By contrast, when the focus of the analysis is not on the mean, but on how the covariates impact the stochastic variation of a family of time series over time, the methodological development has been very preliminary. In this article, we propose a time-frequency functional model for a family of time series indexed by a set of covariates. The proposed model is general in that it takes a general design matrix which includes continuous/categorical covariates as well as dummy variables. Because a spectrum uniquely determines the variation of a time series, it is used as the basic unit of the data analysis. To overcome the

heavy computational load in estimating the proposed model, we develop an efficient algorithm through constructing an equivalent state space model. We also propose to generate covariate-indexed time series based on their design spectra. Our development is centered on locally stationary time series (Dahlhaus 1997, 2000) that are most common in biomedical experiments. These include analyses for stationary time series as special cases.

The motivation behind our research is to develop a systematic method to predict epileptic seizures before their clinical onsets using intracranial EEG (IEEG) data. Recent studies suggest that a pre-seizure state may exist measurably on IEEG data of epileptic patients, who may be more vulnerable to seizure if awake (Litt et al. 2001). Therefore, we are interested in comparing the pre-seizure IEEGs with the baseline IEEGs and assessing the impact of the asleep/awake status of the patient on when the seizure starts and how it spreads. This important prognostic information can eventually be used to predict the clinical onset of a seizure in a new patient, which leads to the development of early medical intervention that may prevent seizures' clinical onsets. Cross-sectional IEEG data were collected from implanted electrodes at 200 Hz from a group of patients with medically resistant temporal lobe epilepsy and the sleep status of patients recorded (D'Alessandro et al. 2001). The baseline segments are extracted from the same regions ≥ 4 hr away from seizures, a time shown to likely be outside the usual span of seizure generation. The analysis presents several challenges with existing methodology. Standard analysis (see Litt et al. 2001 and the references therein) breaks the time series into approximately stationary blocks and makes within-block comparisons. Notwithstanding the simplicity of this attempt, it heavily relies on the assumption of stationarity and the blocking is subjective. The inference is usually sensitive to the block size, as too big a block size tends to violate the stationarity assumption whereas too small a block size leads to large variability. Moreover, the inference is only within-block; however, it is usually of more interest to identify the informative regions in terms of the evolutionary spectra. Figure 1 shows 10-min segments of baseline IEEG and pre-seizure IEEG with the seizure onset, as marked by two expert electroencephalographers, at the end of the tenth minute. It can be seen that the patterns of stochastic variation of the time series are different within each of the four groups (baseline-asleep, baseline-awake, pre-seizure-asleep, pre-seizure-awake), whereas the means of the time series are constants. Because the patterns of stochastic variation are uniquely determined by the time-dependent spectra, we propose a time-frequency functional model to systematically analyze the segments on the IEEG in terms of the logarithm of their time-varying spectra (in the rest of the article, we will use $\log$ - to represent the logarithm of a quantity). We use Bayesian confidence intervals (Gu and Wahba 1993b) to identify the regions of significance in both time and frequency domains, which is more informative than a global test. The proposed method allows us to borrow information across different units in making the population-based inference. Ultimately the goal is to develop an online device that can predict seizure using new IEEG segments. This also motivates the development of the equivalent state space model with sequential updating steps in the associated efficient algorithm. Detailed discussion about this application can be found in Section 6.

Functional data analysis (FDA) (Ramsay and Silverman 1997) extends the conventional statistical analysis to the setting where a curve rather than a single data point is the basic analysis unit. Functional linear models (also termed varying-coefficient models) have been proposed for curves (e.g., Gu and Wahba 1993a; Hastie and Tibshirani 1993; Faraway 1997). When the index of the curve is time, functional linear models have been widely applied to longitudinal data analysis (e.g., Hoover et al. 1998; Fan and Zhang 2000; Chiang, Rice, and Wu 2001). The inferential focus of these models is on the covariate effects on the mean, whereas we are interested in modeling the dependence of the stochastic variation on the covariates and treating the mean as the nuisance parameter. We further extend the idea of functional data analysis to locally stationary time series by using time-varying log-spectra as basic analysis units.

The only work that deals with the patterns of stochastic variation in replicated time series is the article by Diggle and al Wasel (1997), in which a generalized linear model was proposed for the periodograms of the time series. This parametric approach is specifically for replicated stationary time series, and does not include the more general setting in which a family of nonstationary time series are indexed by a set of covariates. In this article, we propose a family of covariate-indexed locally stationary time series. We introduce a time-frequency functional model to model the dependency of the time series on the covariates through the time-varying log-spectra. The nonparametric approach allows us to estimate the covariate effects with minimal assumptions.

We propose a two-stage estimation procedure for our model. In the first stage, the initial estimates of the local periodograms are obtained by approximating the locally stationary time series as piecewise stationary. In the second stage, a time-frequency functional model is used to estimate the time-frequency coefficients for the covariates. Unlike most of the existing methods that make their inference based on the segmentation, our first stage is only to obtain the raw estimates of the local periodograms. The second stage allows the information to be borrowed across segments and series in estimating the time-frequency parameters that are infinite dimensional and have global interpretation. As pointed out by Guo et al. (2003), the final estimates are not sensitive to the initial segmentation due to the second-stage smoothing step. We assess the impact of initial segmentation in the simulation study in Section 5.

The covariate-indexed locally stationary framework permits the connection with two-dimensional smoothing techniques. Similarly to Guo et al. (2003), we use the tensor product approach and model the time-frequency coefficients by time-frequency smoothing spline ANOVA (SS-ANOVA) models (Gu 2002). Due to the curse of dimensionality and the huge size of the IEEG data in our application, the estimation of the proposed model would be virtually impossible if we were to extend the existing algorithm for SS-ANOVA (Gu and Wahba 1993a; Guo et al. 2003). In one-dimensional spline smoothing, Wecker and Ansley (1983) proposed an efficient algorithm via the state space model. Qin and Guo (2006) proposed the equivalent state space form for SS-ANOVA models. We extend this approach and develop an efficient estimation procedure for the time-frequency functional model setting. The corresponding state space model is expanded into a model for univariate series. The new algorithm results in more computational savings than does directly extending the algorithm of Qin and Guo (2006), in that the inherent structure of the time-frequency functional model is exploited. This algorithm significantly reduces the computational demand and makes it possible to apply the proposed model to large datasets. It not only yields a unified procedure for estimation and inference, but also enables online implementation of the proposed procedure.

To evaluate the finite sample performance of the proposed method, we propose a new way to simulate covariate-indexed time series based on their design spectra. We show that the square root of the time-varying spectrum can be used as the transfer function to generate a time series with the designed spectrum. This provides a flexible framework for simulation and enables simulation-based inference.

The rest of the article is organized as follows. In Section 2, we define the proposed model. We introduce the two-stage estimation procedure in Section 3. In Section 4, we propose the equivalent state space model and develop the related efficient algorithm. A simulation study is presented in Section 5. The application to the IEEG time series is discussed in Section 6. Some concluding remarks are offered in Section 7. A short proof is deferred to the Appendix.

2. THE MODELS

In this section, we consider a family of independent time series $\{X_{lt}, t = 1, \dots, T, l = 1, \dots, n\}$ indexed by a set of covariates W_{lt} . Similarly to the design matrices in linear models, W_{lt} is a

design matrix in the time-frequency models, and can include continuous covariates and dummy variables. Therefore our models can be viewed as the extension of linear regression, ANOVA, and ANCOVA to the time-frequency setting.

We first introduce the definition of covariate-indexed locally stationary time series. This extends the locally stationary time series of Guo et al. (2003), a slightly modified version of the original proposal by Dahlhaus (1997). We then propose the time-frequency functional model for the time-varying log-spectra and define the corresponding SS-ANOVA models for the time-frequency coefficients.

2.1 Covariate-Indexed Locally Stationary Time Series

Assume the l th time series X_{lt} to be zero-mean locally stationary processes indexed by design matrix W_{lt} ; they are defined as

$$X_{lt} = \int_0^1 A(\omega, t/T, W_{lt}) \exp(i2\pi\omega t) dZ_l(\omega), \quad (2.1)$$

where

1. $Z_l(\omega)$ are mutually independent zero-mean orthogonal-increment processes on $[0, 1]$ with $Z_l(\omega) = \overline{Z_l(1 - \omega)}$, where $Z_l(1 - \omega)$ is the conjugate of $Z_l(1 - \omega)$. The k th-order cumulant $\{dZ_l(\omega_1), \dots, dZ_l(\omega_k)\}$

cumulant $= \Delta\left(\sum_{j=1}^k \omega_j\right) \Lambda_k(\omega_1, \dots, \omega_{k-1}) d\omega_1 \dots d\omega_k$, where

$\Delta(\omega) = \sum_{j=-\infty}^{\infty} \delta(\omega + j)$ is the period- 2π extension of the Dirac delta function, $\Lambda_1 = 0$, $\Lambda_2(\omega)$, and $|\Lambda_k(\omega, \dots, \omega_{k-1})| \leq C_k$ where C_k is a constant.

2. The transfer functions $A(\omega, u, W_{lu})$, $(\omega, u) \in [0, 1] \times [0, 1]$ have continuous up to second-order partial derivatives with respect to ω and u , and

$$\partial^r A(\omega, u, W_{lu})$$

$/\partial\omega^r = \partial^r A(\omega, u, W_{lu}) / \partial\omega^r$ for $r = 0, 1$. $A(\omega, u, W_{lu})$ is assumed to be a smooth function in ω and u .

The covariate-indexed time-varying spectra $f(\omega, u, W_{lu})$ of X_{lt} are given by $f(\omega, u, W_{lu}) = |A(\omega, u, W_{lu})|^2$, where $|A(\omega, u, W_{lu})|$ denotes the norm of $A(\omega, u, W_{lu})$. Thus, $f(\omega, u, W_{lu})$ is smooth in both time and frequency with continuous second-order partial derivatives w.r.t. both ω and u .

2.2 Time-Frequency Functional Model

Because the spectra are connected to the asymptotic distributions of the periodograms through a multiplicative structure (Wahba 1980; Guo et al. 2003), we model the logarithm of the time-varying spectra instead of the spectra themselves. Let $g(\omega, u, W_{lu}) = \log\{f(\omega, u, W_{lu})\}$; we consider the following functional model for $g(\omega, u, W_{lu})$:

$$g(\omega_k, u_j, W_{ij}) = W_{ij} \beta(\omega_k, u_j), \quad (2.2)$$

where $W_{ij} = \{W_{ij}[1], \dots, W_{ij}[P]\}$ is the design matrix that can contain covariates as well as dummy variables, and $\beta(\omega_k, u_j) = \{\beta_1(\omega_k, u_j), \dots, \beta_P(\omega_k, u_j)\}^T$ is a vector of time-frequency coefficients.

The time-frequency coefficients $\beta_p(\omega, u)$, $p = 1, \dots, P$, are periodic functions with period 1 in the frequency domain, and are symmetric around $\omega = 0.5$. For convenience, we only use half

of the frequency domain in model (2.2) to be free of the periodic constraint. In the rest of the article, we omit the dependence of W_{lu} whenever it is obvious.

2.3 SS-ANOVA for the Time-Frequency Coefficients

In model (2.2), the time-frequency coefficients $\beta_p(\omega, u)$ are two-dimensional smooth functions in time and frequency. In this article, we adopt the SS-ANOVA (Gu 2002; Guo et al. 2003) approach to model $\beta_p(\omega, u)$ as a tensor product of two smoothing splines in both domains. We use cubic smoothing splines for simplicity. The reproducing kernel Hilbert space (RKHS) (Aronszajn 1950) H_p of function $\beta_p(\omega, u)$ has the following tensor product decomposition:

$$\begin{aligned} H_p &= 7 \left[\{1\} \oplus \{u\} \oplus H_{p1} \right] \otimes \left[\{1\} \oplus \{\omega\} \oplus H_{p2} \right] \\ &= \{1\} \oplus \{u\} \oplus \{\omega\} \oplus \{u\omega\} \oplus H_{p1} \oplus H_{p2} \oplus \{H_{p1} \otimes \{\omega\}\} \\ &\quad \oplus \{H_{p2} \otimes \{u\}\} \oplus \{H_{p1} \otimes H_{p2}\}, \end{aligned} \quad (2.3)$$

where $\{1, u, \omega, u\omega\}$ spans the null space H_0 of H_p , the reproducing kernels (RKs) for H_{p1} and H_{p2} are

$$\begin{aligned} R_{p1}(s_1, s_2) &= \int_0^1 (s_1 - v)_+ (s_2 - v)_+ dv (s_j - v)_+ = 0 \text{ if } s_j \leq v; (s_j - v)_+ = (s_j - v) \text{ otherwise, } j \\ &= 1, 2. \text{ And the RKs for } H_{p3} = H_{p1} \otimes \{\omega\}, H_{p4} = H_{p2} \otimes \{u\} \text{ and } H_{p5} = H_{p1} \otimes H_{p2} \text{ are} \\ R_{p3}\{(\omega_1, u_1), (\omega_2, u_2)\} &= R_{p1}(u_1, u_2) \omega_1 \omega_2, R_{p4}\{(\omega_1, u_1), (\omega_2, u_2)\} = R_{p2}(\omega_1, \omega_2) u_1 u_2, \text{ and} \\ R_{p5}\{(\omega_1, u_1), (\omega_2, u_2)\} &= R_{p1}(u_1, u_2) \times R_{p2}(\omega_1, \omega_2), \text{ respectively. Therefore, the time-frequency} \\ &\text{coefficients } \beta_p(\omega, u) \text{ can be decomposed as} \end{aligned}$$

$$\begin{aligned} \beta_p(\omega, u) &= d_{p1} + d_{p2}u + d_{p3}\omega + d_{p4}u\omega \\ &\quad + \beta_{p1}(u) + \beta_{p2}(\omega) + \beta_{p3}(\omega, u) + \beta_{p4}(\omega, u) + \beta_{p5}(\omega, u), \end{aligned} \quad (2.4)$$

where $d_{p1} + d_{p2}u + d_{p3}\omega + d_{p4}u\omega$ is the linear trend, $\beta_{p1}(u)$ is the smooth main effect for u , $\beta_{p2}(\omega)$ is the smooth main effect for ω , $\beta_{p3}(\omega, u)$ is linear in ω and smooth in u , $\beta_{p4}(\omega, u)$ is linear in u and smooth in ω , and $\beta_{p5}(\omega, u)$ is the smooth interaction in both u and ω .

3. ESTIMATION

We propose a two-stage estimation procedure for the time-frequency coefficients. The first stage obtains initial estimates of the log-spectra by partitioning the original time series into small blocks and approximating the series within each block as stationary. In the second stage, the log-local periodograms are modeled by a time-frequency functional model, which can be estimated by penalized least squares (PLS).

3.1 Stage I: Calculating the Local Periodograms

We propose to approximate X_{lt} , $t = 1, \dots, T$, by piecewise stationary time series in small time intervals. We first divide the time domain $[0, 1]$ into M_l disjoint blocks $[b_j^l, b_{j+1}^l)$, where $0 = b_1^l \leq b_2^l \leq \dots \leq b_{M_l}^l \leq b_{M_l+1}^l = 1$ are the partition points for the l th time series. In each of the time blocks, the time series X_{lt} is approximated stationary. For simplicity, we use the discrete Fourier transformation (DFT) of the approximated stationary segments to estimate the local periodograms. Other estimates such as smooth localized complex exponential (SLEX) periodograms (Ombao et al. 2001; Guo et al. 2003) can also be adopted.

Denote the j th ($j = 1, \dots, M_l$) time block by $B_j^l = \{b_j^l, b_j^l+1, \dots, b_{j+1}^l - 1\}$. Let u_{lj} be the middle time point of block B_j^l : $u_{lj} = (b_j^l + b_{j+1}^l - 1)/2$ and L_j^l be the length of the block: $L_j^l = b_{j+1}^l - b_j^l$. Similarly to the finite Fourier representation of stationary time series, we have,

$$d_{lj} = (\omega_{lk}, u_{lj}, W_{lj})$$

for block B_j^l of the l th time series, $= \sum_{t=b_j^l}^{b_{j+1}^l-1} X_{lt} \exp(i2\pi\omega_{lk}t)$. Then the local periodograms $\widehat{I}_l(\omega_{lk}, u_{lj}, W_{lj})$ are calculated by

$$\begin{aligned} \widehat{I}_l(\omega_{lk}, u_{lj}, W_{lj}) \\ = (L_j^l)^{-1} |d_{lj}(\omega_{lk}, u_{lj}, W_{lj})|^2. \end{aligned}$$

The local periodograms $\widehat{I}_l(\omega_{lk}, u_{lj}, W_{lj})$ are calculated as a discrete function on the predefined time-frequency grids. The u_{lj} s are the middle points of the time blocks and the ω_{lk} s are chosen to be a subset of the Fourier frequencies. Note that in the partitions, the sizes of time blocks and frequency points are not necessarily equal for each block and for each time series. To ease the presentation, in the following text we assume them to be equal, that is, $M_l = M$, $b^l = b$, $L^l = L$, $u_{lj} = u_j$, and $\omega_{lk} = \omega_k$.

3.2 Stage II: Estimating the Time-Frequency Coefficients

In this section, we propose to estimate the time-frequency coefficients in model (2.2) through fitting the log-local periodograms to a time-frequency functional model. We use the penalized least-squares estimators (PLSE) for the time-frequency coefficients and construct Bayesian confidence intervals (Gu and Wahba 1993b) for the PLSE.

The following lemma summarizes the properties of the local periodograms $\widehat{I}_l(\omega_k, u_j, W_{lj})$ ($k = 1, \dots, S, j = 1, \dots, M, l = 1, \dots, n$):

Lemma 3.1. As $T \rightarrow \infty$, $\min B_j \rightarrow \infty$, and $\max \frac{(B_j)^2}{T} = O(1)$, we have:

1. $\widehat{I}_l(\omega_k, u_j, W_{lj}) \xrightarrow{d} \begin{cases} f(\omega_k, u_j, W_{lj}) \chi_2^2/2 & \text{if } \omega_k \neq 0, 1/2, 1 \\ f(\omega_k, u_j, W_{lj}) \chi_1^2 & \text{if } \omega_k = 0, 1/2, 1 \end{cases}$
2. $\widehat{I}_{l1}(\omega_{k1}, u_j, W_{lj})$ and $\widehat{I}_{l2}(\omega_{k2}, u_j, W_{lj})$ are mutually independent for any given ω_{k1} , ω_{k2} , and $l_1 \neq l_2 \in \{1, \dots, n\}$.
3. $\widehat{I}_l(\omega_{k1}, u_j, W_{lj})$ and $\widehat{I}_l(\omega_{k2}, u_j, W_{lj})$ are asymptotically independent for fixed l and $k_1 \neq k_2 \in [0, 1]$.

4. Local periodograms are asymptotically unbiased for the spectra:

$$\begin{aligned} E\left(\widehat{I}_l(\omega_k, u_j, W_{lj})\right) \\ = f(\omega_k, u_j, W_{lj}) \\ + O\left(\frac{1}{\min B_j}\right) \\ + O\left(\frac{\max B_j}{T}\right). \end{aligned}$$

The proof is similar to that in Guo et al. (2003) and is omitted here. Using the results from this lemma, we have the following time-frequency functional model for the log-local periodograms:

$$\log\left(\widehat{I}_l(\omega_k, u_j, W_{lj})\right) \approx g(\omega_k, u_j, W_{lj}) + \delta_{ljk}, \quad (3.1)$$

where δ_{ljk} are asymptotically independent with mean $C_{\omega k}$ and variance $\sigma^2 = \pi^2/6$, where $C_{\omega k}$ is the Euler Mascherni constant, $-C_{\omega k} = \gamma = 0.57721\dots$ for $\omega_k \neq 0, 1/2$, and $C_0 = C_{1/2} = -(\ln 2 + \gamma)/\pi$.

Let $y_{ljk} = \log(\widehat{I}_l(\omega_k, u_j, W_{lj})) - C_{\omega k}$. By plugging in (2.2), model (3.1) becomes

$$y_{ljk} \approx W_{lj}\beta(\omega_k, u_j) + \epsilon_{ljk}, \quad (3.2)$$

where ϵ_{ljk} are asymptotically independent with mean zero and variance $\pi^2/6$.

Denote a time-frequency point $\phi_h = (\omega_k, u_j)$ with a one-to-one correspondence between h and (j, k) : $h = (j-1)S + k$, for $j = 1, \dots, M$, $k = 1, \dots, S$. Then denote the predefined time-frequency grid for the l th time series by $\Phi_l = \{\Phi_1, \dots, \Phi_m\}^T$ with $m = MS$, and the total time-frequency grids by $\Phi = \{\Phi_1^T, \dots, \Phi_n^T\}^T$. Correspondingly, let $y_{lh} = \{y_{ljk}\}$ for $j = 1, \dots, M$, $k = 1, \dots, S$, $\mathbf{y}_l = \{y_{lh}, h = 1, \dots, m\}$, $\mathbf{Y} = \{\mathbf{y}_1^T, \dots, \mathbf{y}_n^T\}^T$, $\beta(\Phi) = \{\beta_1^T(\Phi), \dots, \beta_p^T(\Phi)\}^T$, and $\mathbf{W}[P] = \text{diag}\{\mathbf{W}_1[p], \dots, \mathbf{W}_n[p]\}$, where $\mathbf{W}_l[p] = \text{diag}\{W_{l1}[p], \dots, W_{lm}[p]\}$ and $W_{lh}[p] = W_{ljk}[p]$ with $h = (j-1)S + k$ ($k = 1, \dots, S$) in model (2.2). An estimate of the p th time-frequency coefficient $\beta_p(\Phi)$ is the minimizer of the following penalized least squares (PLS):

$$\begin{aligned} 1/(nm) \left\| \mathbf{Y} - \sum_{p=1}^P \mathbf{W}[p] \beta_p(\Phi) \right\|^2 \\ + \sum_{p=1}^P \sum_{r=1}^5 \lambda \theta_{pr}^{-1} \|P_{pr} \beta_p\|^2, \end{aligned} \quad (3.3)$$

where λ and θ_{pr} are reparameterizations of $\lambda_{pr} = \lambda/\theta_{pr}$ and P_{pr} denote the orthogonal projections onto the RKHS H_{pr} in the tensor product decomposition (2.3). For fixed λ and θ_{pr} , by the property of RKHS with $W_{lh}[p]$ the bounded linear functional, the PLSE of $\beta_p(\Phi)$ is given by (Wahba 1990, p. 127)

$$\widehat{\beta}_p(\Phi) = U_p \mathbf{d}_p + Q_p \mathbf{c}_p, \quad (3.4)$$

where $U_p = (1, \omega_h, u_h, \omega_h u_h)_{h=1}^{nm}$ correspond to the null space H_0 , $Q_p = \sum_{r=1}^5 \theta_{pr} \times \mathbf{R}_{pr}(\Phi, \Phi)$ span $\Sigma_{j=1}^r \oplus H_{pj}$, $\mathbf{R}_{pr}(\Phi, \Phi) = \{R_{pr}(\phi_h, \phi_l)\}_{h=1, l=1}^{nm, nm}$ with $R_{pr}(\phi_h, \phi_l)$ defined in (2.3) and details be found in Gu (2002, pp. 40–42). The coefficients $\mathbf{d}_p = \{d_{p1}, d_{p2}, d_{p3}, d_{p4}\}^T$ correspond to a parametric model with an intercept d_{p1} , slopes d_{p2}, d_{p3} , and an interaction d_{p4} for a linear trend in u and ω ; and $\mathbf{c}_p$ represent the smoothness derivation from the parametric model. They can be solved from

$$\begin{aligned} (Q_p + nm\lambda \mathbf{I}) \mathbf{c}_p + U_p \mathbf{d}_p &= \mathbf{Y}, \\ U_p^T \mathbf{c}_p &= 0. \end{aligned}$$

Casting in matrix form, the PLSE of the whole time-frequency coefficient vector $\beta(\Phi) : \widehat{\beta}(\Phi) = \mathbf{T}\mathbf{D} + \mathbf{Q}\mathbf{C}$ can be estimated from the PLS (Wahba 1990)

$$1/(nm) \|\mathbf{Y} - \mathbf{W}\mathbf{T}\mathbf{D} - \mathbf{W}\mathbf{Q}\mathbf{C}\|^2 + \lambda \mathbf{C}^T \mathbf{Q}\mathbf{C},$$

where $\mathbf{W} = \{\mathbf{W}[1], \dots, \mathbf{W}[P]\}$, $\mathbf{T} = \text{diag}\{U_1, \dots, U_P\}$, $\mathbf{Q} = \text{diag}\{Q_1, \dots, Q_P\}$. The minimizing coefficient vectors $\mathbf{D} = \{\mathbf{d}_1^T, \dots, \mathbf{d}_P^T\}^T$ and $\mathbf{C} = \{\mathbf{c}_1^T, \dots, \mathbf{c}_P^T\}^T$ are given by

$$\begin{aligned} \mathbf{D} &= \{T^T \mathbf{W}^T \mathbf{M}^{-1} \mathbf{W}\mathbf{T}\}^{-1} T^T \mathbf{W}^T \mathbf{M}^{-1} \mathbf{Y}, \\ \mathbf{C} &= \mathbf{W}\mathbf{M}^{-1} \left\{ \mathbf{I} - \mathbf{W}\mathbf{T} (T^T \mathbf{W}^T \mathbf{M}^{-1} \mathbf{W}\mathbf{T})^{-1} T^T \mathbf{W}^T \mathbf{M}^{-1} \right\} \mathbf{Y}, \end{aligned}$$

where $\mathbf{M} = \mathbf{W}\mathbf{Q}\mathbf{W}^T + nm\lambda \mathbf{I}$.

The consistency of the PLSE for a single log-spectrum has been proved by Guo et al. (2003), which can be easily extended in the setting of a time-frequency functional model.

The parameters λ and θ_{pr} can be chosen by generalized cross-validation (GCV) or generalized maximum likelihood (GML) (Wahba 1985). We focus on the GML criterion in this article:

$$V(\lambda, \theta_{pr}) = \frac{\mathbf{Y}^T \mathbf{F}_2 (\mathbf{F}_2^T \mathbf{W}\mathbf{Q}\mathbf{W}^T \mathbf{F}_2 + nm\lambda \mathbf{I})^{-1} \mathbf{F}_2^T \mathbf{Y}}{[\det(\mathbf{F}_2^T \mathbf{W}\mathbf{Q}\mathbf{W}^T \mathbf{F}_2 + nm\lambda \mathbf{I})]^{1/(nm-4)}}, \quad (3.5)$$

where $\mathbf{F}_2$ is given by the QR decomposition of $\mathbf{W}\mathbf{T} : \mathbf{W}\mathbf{T} = (\mathbf{F}_1, \mathbf{F}_2) \begin{pmatrix} \mathbf{G} \\ \mathbf{0} \end{pmatrix}$.

Following (3.4), we have the PLSE at any given time-frequency point ϕ_0 :

$$\widehat{\beta}_p(\phi_0) = U_{p0} \mathbf{d}_p + Q_{p0} \mathbf{c}_p,$$

where $U_{p0} = (1, \omega_0, u_0, \omega_0 u_0)$, $Q_{p0} = \sum_{r=1}^5 \theta_{pr} \mathbf{R}_{pr}(\phi_0, \Phi)$, and $\mathbf{R}_{pr}(\phi_0, \Phi) = \{R_{pr}(\phi_0, \phi_h)\}_{h=1}^{nm}$, for $p = 1, \dots, P$. Similarly to Wahba (1983), assuming $\mathbf{d}_p \sim N(\mathbf{0}, \kappa \mathbf{I})$ with $\kappa \rightarrow \infty$, $\widehat{\beta}_p(\phi_0)$ can be shown to be the posterior mean of $\beta_p(\phi_0)$ given $\mathbf{Y}$.

Conditioning on the estimates of the smoothing parameters, we have the following posterior variance of $\widehat{\beta}_p(\phi_0)$:

$$\text{var} \left\{ \widehat{\beta}_p(\phi_0) | \mathbf{Y} \right\} = \sigma^2 / (nm\lambda) \left[U_{p0} \Gamma_p A^{-1} \left(\Gamma_p^T U_{p0}^T - U^T M^{-1} W Q_{p0} \right) + Q_{p00} - Q_{p0} W^T M^{-1} \left\{ W Q_{p0} + U A^{-1} \left(U^T M^{-1} W Q_{p0} + \Gamma_p^T U_{p0}^T \right) \right\} \right],$$

where $\sigma^2 = \pi^2/6$, $\Gamma_p = \{\Gamma_{p1}, \dots, \Gamma_{pp}, \dots, \Gamma_{pP}\}$ is a $4 \times 4P$ matrix with $\Gamma_{pp} = I_4$ and $\Gamma_{pq} = \mathbf{0}$, for $q \neq p$, $A = (U^T M^{-1} U)^{-1}$, $U = WT$, and $Q_{p00} = \sum_{r=1}^5 \theta_{pr} R_{pr}(\phi_0, \phi_0)$.

We then construct Bayesian confidence intervals for $\beta(\phi_0)$:

$$\left[\widehat{\beta}(\phi_0) - z_{1-\alpha/2} \sqrt{\text{var} \left\{ \widehat{\beta}(\phi_0) | \mathbf{Y} \right\}}, \widehat{\beta}(\phi_0) + z_{1-\alpha/2} \sqrt{\text{var} \left\{ \widehat{\beta}(\phi_0) | \mathbf{Y} \right\}} \right], \quad (3.6)$$

where $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$ percentile of the standard Normal distribution. The Bayesian confidence intervals have curvewise interpretation as averaging “confidence intervals” across the whole time-frequency grid (Nychka 1988). They can be used to identify the informative regions pertaining to design covariates in the time and frequency domains.

4. ESTIMATION THROUGH THE STATE SPACE APPROACH

In Section 3.2, we presented the estimation of the time-frequency coefficients through their PLSEs. However, this estimation procedure requires inverting high-dimensional matrices and is extremely computationally intensive. In this section, we construct an equivalent state space model for the proposed time-frequency functional model. This approach not only allows practical utility of the proposed method, but also gives unified treatment for estimating the time-frequency coefficients, choosing the model parameters via GML, constructing Bayesian confidence intervals, and implementing the procedure online.

We propose to derive the equivalent state space model by making connections between models (2.2), (3.2), and the Bayesian stochastic models of the time-frequency coefficients. We introduce the following Bayesian model as a prior distribution for (2.4):

$$\begin{aligned} \beta_p(\omega, u) = & d_{p1} + d_{p2}u + d_{p3}\omega + d_{p4}u\omega + \lambda_{p1}^{1/2} \int_0^u (u-v) dW_{p1}(v) \\ & + \lambda_{p2}^{1/2} \int_0^\omega (\omega-v) dW_{p2}(v) + \lambda_{p3}^{1/2} \omega \int_0^u (u-v) dW_{p3}(v) \\ & + \lambda_{p4}^{1/2} u \int_0^\omega (\omega-v) dW_{p4}(v) \\ & + \lambda_{p5}^{1/2} \int_0^u (u-v) dW_{p5}(v) \int_0^\omega (\omega-v) dW_{p6}(v), \end{aligned} \quad (4.1)$$

where $\mathbf{d}_p = [d_{p1}, d_{p2}, d_{p3}, d_{p4}]^T$ has diffuse Normal priors, $\{\lambda_{pr}, r = 1, \dots, 5\}$ are smoothing parameters, and $\{W_{pk}(v), k = 1, \dots, 6\}$ are Wiener processes.

$$\begin{aligned} & \tilde{\beta}_p(\omega_k, u_j) \\ &= \left\{ \beta_p(\omega_k, u_j) \right. \\ & \quad, \frac{\partial \beta_p(\omega_k, u_j)}{\partial u} \\ & \quad, \frac{\partial \beta_p(\omega_k, u_j)}{\partial \omega} \\ & \quad, \left. \frac{\partial \beta_p^2(\omega_k, u_j)}{\partial u \partial \omega} \right\}^T, \quad v_{jk}^p = (v_{jk1}^p, v_{jk2}^p, v_{jk3}^p, v_{jk4}^p)^T, \end{aligned}$$

We define

$$H_{jk}^p = \begin{pmatrix} 1 & \delta u_j & \delta \omega_k & \delta u_j \delta \omega_k \\ 0 & 1 & 0 & \delta \omega_k \\ 0 & 0 & 1 & \delta u_j \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and $\Sigma_{jk}^p = \{\sigma_{\gamma\rho}^{pjk}\}_{\gamma,\rho=1,1}^{4,4}$, with

$$\begin{aligned} \sigma_{11}^{pjk} &= (\lambda_{p1} + \lambda_{p3} \omega_k^2) \frac{\delta u_j^3}{3} \\ &+ (\lambda_{p2} + \lambda_{p4} u_j^2) \frac{\delta \omega_k^3}{3} \\ &+ \lambda_{p5} \frac{u_j^3 \delta \omega_k^3 + \delta u_j^3 \omega_k^3 - \delta u_j^3 \delta \omega_k^3}{9}, \end{aligned}$$

$$\begin{aligned} \sigma_{21}^{pjk} &= \sigma_{12}^{pjk} = \lambda_{p1} \frac{\delta u_j^2}{2} \\ &+ \lambda_{p3} \omega_k^2 \frac{\delta u_j^2}{2} \\ &+ \lambda_{p4} u_j \frac{\delta \omega_k^3}{3} \\ &+ \lambda_{p5} \frac{u_j^2 \delta \omega_k^3 + \delta u_j^2 \omega_k^3 - \delta u_j^2 \delta \omega_k^3}{6}, \end{aligned}$$

$$\begin{aligned} \sigma_{31}^{pjk} &= \sigma_{13}^{pjk} = \lambda_{p2} \frac{\delta \omega_k^2}{2} \\ &+ \lambda_{p3} \omega_k \frac{\delta u_j^3}{3} \\ &+ \lambda_{p4} u_j^2 \frac{\delta \omega_k^2}{2} \\ &+ \lambda_{p5} \frac{u_j^3 \delta \omega_k^2 + \delta u_j^3 \omega_k^2 - \delta u_j^3 \delta \omega_k^2}{6}, \end{aligned}$$

$$\begin{aligned}\sigma_{23}^{pjk} &= \sigma_{32}^{pjk} = \sigma_{41}^{pjk} = \sigma_{14}^{pjk} = \lambda_{p3} \omega_k \frac{\delta u_j^2}{2} \\ &+ \lambda_{p4} u_j \frac{\delta \omega_k^2}{2} \\ &+ \lambda_{p5} \frac{u_j^2 \delta \omega_k^2 + \delta u_j^2 \omega_k^2 - \delta u_j^2 \delta \omega_k^2}{4},\end{aligned}$$

$$\begin{aligned}\sigma_{22}^{pjk} &= \lambda_{p1} \delta u_j \\ &+ \lambda_{p3} \omega_k^2 \delta u_j \\ &+ \lambda_{p4} \frac{\delta \omega_k^3}{3} \\ &+ \lambda_{p5} \frac{\delta u_j \omega_k^3 + u_{j-1} \delta \omega_k^3}{3},\end{aligned}$$

$$\begin{aligned}\sigma_{42}^{pjk} &= \sigma_{24}^{pjk} = \lambda_{p3} \omega_k \delta u_j \\ &+ \lambda_{p4} \frac{\delta \omega_k^2}{2} \\ &+ \lambda_{p5} \frac{u_{j-1} \delta \omega_k^2 + \omega_k^2 \delta u_j}{2},\end{aligned}$$

$$\begin{aligned}\sigma_{33}^{pjk} &= \lambda_{p2} \delta \omega_k \\ &+ \lambda_{p3} \frac{\delta u_j^3}{3} \\ &+ \lambda_{p4} u_j^2 \delta \omega_k \\ &+ \lambda_{p5} \frac{u_j^3 \delta \omega_k + \delta u_j^3 \omega_{k-1}}{3},\end{aligned}$$

$$\begin{aligned}\sigma_{43}^{pjk} &= \sigma_{34}^{pjk} = \lambda_{p3} \frac{\delta u_j^2}{2} \\ &+ \lambda_{p4} u_j \delta \omega_k \\ &+ \lambda_{p5} \frac{u_j^2 \delta \omega_k + \delta u_j^2 \omega_{k-1}}{2},\end{aligned}$$

$$\begin{aligned}\sigma_{44}^{pjk} &= \lambda_{p3} \delta u_j \\ &+ \lambda_{p4} \delta \omega_k \\ &+ \lambda_{p5} (u_j \omega_k - u_{j-1} \omega_{k-1}),\end{aligned}$$

where $\delta u_j = u_j - u_{j-1}$ ($u_{j-1} \leq u_j$) and $\delta \omega_k = \omega_k - \omega_{k-1}$ ($\omega_{k-1} \leq \omega_k$), $p = 1, \dots, P$. In the above definitions of H_{jk}^p and Σ_{jk}^p , two adjacent points (ω_{k-1}, u_{j-1}) and (ω_k, u_j) in the time-frequency grid share the same frequency, that is, $\omega_{k-1} = \omega_k$ and $\delta \omega_k = 0$, or the same time, that is, $u_{j-1} = u_j$ and $\delta t_j = 0$. Details of the derivation for (4.1) are similar to those in Qin and Guo (2006) and thus are omitted to save space.

The Markovian structure for the state space approach requires that $\delta u_j \geq 0$ and $\delta \omega_k \geq 0$ in the associated Kalman filter algorithm. As was defined in Section 3.2 for PLS (3.3), we denote a point on the time-frequency grid as $\phi_h = (\omega_k, u_j)$ and define a one-to-one correspondence between h and (j, k) : $h = (j-1)S + k$, for $j = 1, \dots, M$, $k = 1, \dots, S$, $h = 1, \dots, m$. Correspondingly, we denote $\widehat{\beta}_p(\omega_k, u_j)$ by $\widehat{\beta}_p(\phi_h)$, $\mathbf{v}_{jk}^p$ by $\mathbf{v}_h^p$, H_{jk}^p by H_h^p , Σ_{jk}^p by Σ_h^p , and let $\widehat{\beta}(\phi_h) = \{\widehat{\beta}_1(\phi_h)^T, \dots, \widehat{\beta}_p(\phi_h)^T\}^T$, $\widehat{H}_h = \text{diag}\{H_h^1, \dots, H_h^p\}$, $\widehat{\mathbf{v}}_h = \{\mathbf{v}_h^{1T}, \dots, \mathbf{v}_h^{pT}\}^T$, and $\widehat{\Sigma}_h = \text{diag}\{\Sigma_h^1, \dots, \Sigma_h^p\}$. Under the one-to-one mapping, the observations y_{ljk} , $k = 1, \dots, S$, $j = 1, \dots, M$, $l = 1, \dots, n$, in the two-dimensional time-frequency grids are converted to a univariate time series $y_{11}, \dots, y_{n1}, \dots, y_{1m}, \dots, y_{nm}$. Thus we denote $ylh = y_{ljk}$ and $\epsilon_{lh} = \epsilon_{ljk}$.

Define a new state vector $\widetilde{\beta}(\omega, u_h) = \left\{ \widetilde{\beta}(\omega_1, u_h), \dots, \widetilde{\beta}(\omega_S, u_h) \right\}^T$. We have the following state space model equivalent to model (3.2):

$$\begin{aligned} y_{lh} &= \widetilde{W}_{lh} \widetilde{\beta}(\omega, u_h) + \epsilon_{lh}, \\ \widetilde{\beta}(\omega, u_h) &= \begin{cases} \widetilde{H}_h \widetilde{\beta}(\omega, u_{h-1}) + \widetilde{\mathbf{v}}_h, \\ \widetilde{\mathbf{v}}_h \sim \mathcal{N}(0, \widetilde{\Sigma}_h), & h = jS + 1 \\ \widetilde{\beta}(\omega, u_{h-1}), & h = (j-1)S + 2, \dots, jS, \end{cases} \end{aligned} \quad (4.2)$$

where ϵ_{lh} are zero-mean uncorrelated errors with common variance σ^2 ; the new design matrices $\widetilde{W}_{lh}$ are defined through augmenting W_{lh} in model (2.2) by zeros for the columns corresponding to the derivatives:

$$\begin{aligned} \widetilde{W}_{lh}[4p-3] &= W_{lh}[p], \\ \widetilde{W}_{lh}[4p] &= \widetilde{W}_{lh}[4p-1] = \widetilde{W}_{lh}[4p-2] = 0, \quad p = 1, \dots, P; \end{aligned}$$

$\widetilde{H}_h = \text{diag}\{\widetilde{H}_{j1}, \dots, \widetilde{H}_{jS}\}$; $\widetilde{\mathbf{v}}_h = \left\{ \widetilde{\mathbf{v}}_{j1}^T, \dots, \widetilde{\mathbf{v}}_{jS}^T \right\}^T$, and $\widetilde{\Sigma}_h$ is a $4S \times 4S$ block matrix with the (j, k) th block $\widetilde{\Sigma}_{jk}$ ($k = 1, \dots, S$), and $\widetilde{\Sigma}_{kj} = \widetilde{\Sigma}_{jk}^T$.

Model (4.2) is equivalent to a multivariate state space model (Koopman and Durbin 2000).

The state vector $\widetilde{\beta}(\omega, u_h)$ is a latent process underlying the univariate series y_{lh} , $h = 1, \dots, m$, $l = 1, \dots, n$. Hence this model can be fitted by a univariate Kalman filter and smoother (KFS) (Anderson and Moore 1979), which significantly reduces the computational load. The proposed algorithm is different from that in Qin and Guo (2006) due to the Markovian nature in the time-frequency functional model, resulting in further computational savings. Conditioning on $\widetilde{\beta}(\omega, u_h)$, the observations from different time series $\{y_{lh}, l = 1, \dots, n\}$ are mutually independent. Thus in the Kalman filter, only one prediction step is required for each h . The state vector $\widetilde{\beta}(\omega, u_h)$ is then sequentially filtered for all the replicates, whereas in Qin and Guo

(2006), the state vectors $\tilde{\beta}(\omega, u_h)$ are distinct for each y_{lh} , therefore are of higher dimensions (4nPS) than those (4PS) in our model.

The unknown parameters in model (4.2) are the smoothing parameters $\{\lambda_{pr}, r = 1, \dots, 5, p = 1, \dots, P\}^T$. Let $\eta = \{\lambda_{pr}, r = 1, \dots, 5, p = 1, \dots, P\}^T$; η can be chosen by the following restrictive maximum likelihood (REML):

$$L(\eta|Y) = p(\tilde{Y}_M | \tilde{Y}_{M-1}, \eta) \cdots p(\tilde{Y}_{M-1} | \tilde{Y}_{M-2}, \eta) \cdots p(\tilde{Y}_2 | \tilde{Y}_1, \eta) p(\tilde{Y}_1 | \eta), \quad (4.3)$$

where $\tilde{Y}_j$ denotes the observations from all the time series at time points $u_j, \dots, u_J$. The conditional densities $p(\tilde{Y}_j | \tilde{Y}_{j-1}, \eta)$ can be computed from the proposed KFS algorithm. With diffuse Normal priors for the initial state $\tilde{\beta}(\omega, u_1)$, criterion (4.3) is equivalent to the GML (3.5) in Section 3.2.

Given the estimates of the model parameters, the posterior means $E\{\tilde{\beta}(\omega, u_j) | Y\}, j = 1, \dots, M$ are equivalent to the PLSEs (3.4), and the posterior variances $\text{var}\{\tilde{\beta}(\omega, u_j) | Y\}$ can be used to construct Bayesian confidence intervals (3.6).

For ease of presentation, we have assumed equal time-frequency grids for all the time series $i = 1, \dots, n$. In general, the time blocks can be different with different lengths for each time series; so are the frequencies. This can be handled easily in the state space model framework, where the total time-frequency grids are defined including all the different time points and frequency points from all the time series. As already stated, the initial blocking on the time domain (and the frequency domain) only influences the raw estimates; the extra efforts are not necessary due to the second-stage smoothing.

Because of the recursive nature of the state space model, the estimation procedure can be implemented online by extending the proposed KFS for data collected at real time. This is particularly useful for our application in online seizure prediction.

5. SIMULATION

We conduct a simulation study to evaluate the empirical performance of the proposed method. The following lemma gives a guideline to simulate covariate-indexed time series $\{X_{lt}, t = 1, \dots, T\}$ with log-spectra $g_d(\omega, u, W_{lu})$ equal to the design log-spectra $g(\omega, u, W_{lu})$ at a finite grid of time-frequency points $(k/T, k/T), k = 1, \dots, T$.

Lemma 5.1. Let $g(\omega, u, W_{lu}) = W_{lu}\beta(\omega, u)$ be the logarithm of the design spectrum $f(\omega, u, W_{lu})$ with continuous up to second-order partial derivatives with respect to ω and u , where (ω, u) is defined on $[0, 1] \times [0, 1]$, $W_{lu} = \{W_{lu}[1], \dots, W_{lu}[p]\}$ are the design covariates, and $\beta(\omega, u) = \{\beta_1(\omega, u), \dots, \beta_p(\omega, u)\}^T$ are the corresponding time-frequency coefficients. We have a finite sample $\{X_{lt}, t = 1, \dots, n, t = 1, \dots, T\}$ of a locally stationary time series:

$$X_{lt} = \sum_{k=1}^T \exp\left\{\frac{W_{lt}}{2}\beta\left(\frac{k}{T}, \frac{t}{T}\right)\right\} \exp\left(\frac{2\pi k t}{T}i\right) Z_l(k), \quad (5.1)$$

where i is the imaginary unit: $i^2 = -1$; $Z_l(k)$, $l = 1, \dots, n$, are mutually independent zero-mean random variables distributed as complex Normal with variance $1/T$ and $z(k) = z(T - k)$, when $k/T \neq 0, 0.5$, or 1 ; or as real Normal with variance $1/T$, otherwise. The spectrum of X_{lt} , $f_d(\omega, u, W_{lu})$ satisfies (1) $f_d(k/T, t/T, W_{lt}) = f(k/T, t/T, W_{lt}) = \exp\{W_{lt}\beta(k/T, t/T)\}$ for $k, t = 1, \dots, T$ and (2) $|f_d(\omega, u, W_{lu}) - f(\omega, u, W_{lu})| = O(1/T)$ for $\forall (\omega, u) \in [0, 1] \times [0, 1]$.

The proof of the lemma is deferred to the Appendix.

The simulation takes three steps:

- Step 1: simulation of replicated time series from given log-spectra;
- Step 2: calculation of local periodograms by piecewise stationary approximation;
- Step 3: estimation of the true log-spectra through fitting the time-frequency functional model.

Based on Lemma 5.1, we simulated replicated time series from three log-spectra settings: $\beta(\omega, t) = 5 + A \sin(2\pi t) + 2 \log(1.25 - \cos(2\pi\omega))$ with $A = 0, 1, 2$. These settings represent different levels of local stationarity. The case $A = 0$ is equivalent to a stationary process, whereas $A = 1$ and $A = 2$ correspond to two locally stationary processes with increased local variations. We assume that the replicated time series X_{lt} are indexed by a continuous covariate W_l : $W_l \sim U(0.5, 1.5)$. We obtained 40 replicates ($l = 1, \dots, 40$) for time series of length $T = 6,000$. To investigate the impact of different segmentation sizes, we use three equally spaced time-frequency grids $\{(\omega_k, t_j), k = 1, \dots, S, j = 1, \dots, M\}$ for calculating the local periodograms: (1) $(S \times M) = (16 \times 25)$, (2) $(S \times M) = (20 \times 20)$ and (3) $(S \times M) = (25 \times 16)$. One hundred datasets are generated for each setting. Figures 2 and 3 show the boxplots of the mean squared errors (MSEs) of the posterior estimates and the 95% nominal coverage of the Bayesian confidence intervals. It can be seen that our method gives small MSEs as compared to the scales of the signal and noise. The coverage of the 95% Bayesian confidence intervals is slightly lower than the nominal level. This is because the penalized least-square method only uses the first two moments of the log-periodograms in our estimation. Within each log-spectrum setting, both the MSEs and the coverage results are not sensitive to the initial segmentation. As A becomes large, the MSEs tend to be bigger as the same number of approximated stationary blocks are used to accommodate increased nonstationarity. The estimates of the error variance were around 1.65 for all the simulation runs, which approaches the asymptotic variance $\sigma^2 = \pi^2/6$ pretty well.

6. APPLICATION

We apply model (2.2) to the IEEG time series data shown in Figure 1. We started from the full model for the 70 time series:

$$g_l(\omega_k, u_j) = \beta_1(\omega_k, u_j) + I(\text{pre-seizure})\beta_2(\omega_k, u_j) + I(\text{awake})\beta_3(\omega_k, u_j) + I(\text{pre-seizure})I(\text{awake})\beta_4(\omega_k, u_j), \quad (6.1)$$

where $\beta_1(\omega, u)$ denotes the log-spectrum for the baseline segments collected from asleep patients, $\beta_2(\omega, u)$ ($\beta_3(\omega, u)$) denotes the difference in the log-spectra between pre-seizure and baseline (awake and asleep) segments, and $\beta_4(\omega_k, u_j)$ denotes the interaction effect. $I(\cdot)$ is the indicator function and $k = 1, \dots, 20, j = 1, \dots, 40, l = 1, \dots, 70$.

The interaction term $\beta_4(\omega_k, u_j)$ dropped off due to lack of significance and our final model fit to the main effects only. Figure 4 presents the estimated log-spectra and difference log-spectra with their 95% Bayesian confidence intervals. It shows that when patients are asleep, the baseline segments are approximately stationary as the time-varying spectrum reduces to a

smooth curve in frequency. The log-spectra difference estimates between the awake and asleep segments are presented on the second row of Figure 4. The confidence intervals exclude zero, confirming that epileptic patients generally have higher brain energy when they are awake. The difference between the pre-seizure log-spectrum and the baseline log-spectrum displays an overall increasing trend over time, implying that power of the IEEG builds up during the pre-seizure state. This agrees with previous studies, which demonstrate an increase in energy derived from the IEEG signal before seizure onset, appearing as an eruption of energy across the whole frequency domain (Litt et al. 2001). As shown in the (3, 2) panel of Figure 4, even 10 min ahead of the seizure onset, the pre-seizure epochs already have significantly higher energy at the lower frequencies (0–25 Hz). At the higher frequencies (25–100 Hz), the differences between the two group-average log-spectra do not become significant until around 1 min prior to seizure onset. We interpret this as meaning that the difference of the power distribution between the pre-seizure and the baseline segments mainly locates in the lower frequency end (≤ 20 Hz), which then abruptly spreads to higher frequencies as it moves closer to seizure onset. It indicates that changes between the baseline and pre-seizure EEG time series are already prominent in their spectra 10 min before the seizure's clinical onset, which is useful information for developing seizure prediction tools. It also suggests that the lower frequency end contains the most important information about the seizure, because their lower 95% Bayesian confidence intervals of the difference log-spectrum are consistently above the zero plane. This information allows us to design new experiments to study low-frequency activities, and offers potential clues to mechanisms underlying seizure generation.

7. CONCLUSION

In this article, we have extended functional data analysis to the setting where a time-varying log-spectrum is the basic unit in the analysis. The relationship between the log-spectra and covariates is described by a time-frequency functional model. This model sets out a general framework to make inference on a family of time series indexed by a set of covariates, when the interest is in the stochastic variation of the time series.

To overcome the computational intensity, we have developed an efficient algorithm through constructing an equivalent state space model for the time-frequency functional model. This approach provides an efficient way to fit a general functional linear model.

We have extended the Bayesian confidence intervals to the time-frequency setting. This inferential tool conveys important information that would not be effectively obtained via other methods such as global tests on the time-frequency coefficients. In the IEEG time series example, the use of Bayesian confidence intervals has successfully identified regions of the spectrum important to seizure generation, which may help identify specific neuronal populations in the mesial temporal lobe during this process.

We have assumed a linear relationship between the log-spectra and the covariates. If the linearity assumption does not hold, certain transformations are needed to fit the proposed setting, whereas a more viable computational method needs to be developed.

Acknowledgments

The authors thank the editor, associate editor, two referees, and Ming Dai for their constructive comments and suggestions that have significantly improved this article. The research was partially supported by NIH grant R01 CA 84438.

APPENDIX: PROOF OF LEMMA 5.1

From (5.1), we have the following covariance expression for a given $l \in \{1, \dots, n\}$, and for $t, s = 1, \dots, T$:

$$\begin{aligned} \text{cov}(X_{lt}, X_{ls}) = & \frac{1}{T} \sum_{k=1}^T \exp \left[W_{lt} \left\{ \beta \left(\frac{k}{T}, \frac{t}{T} \right) + \beta \left(\frac{k}{T}, \frac{s}{T} \right) \right\} \right] \\ & \times \exp \left\{ \frac{2\pi k(t-s)}{T} i \right\}. \end{aligned} \quad (\text{A.1})$$

Let $U_T(\omega)$ be a count measure on $[0, 1]$ with jump $1/T$ at each $k = 1, \dots, T$ and let $\beta_d(\omega, u)$ be a smooth function connecting $\{k/T, t/T, \beta(k/T, t/T)\}$. Equation (A.1) can be rewritten as

$$\begin{aligned} \text{cov}(X_{lt}, X_{ls}) &= \int_0^1 \exp \left[W_{lt} \left\{ \beta_d \left(\omega, \frac{t}{T} \right) + \beta_d \left(\omega, \frac{s}{T} \right) \right\} \right] \exp \{2\pi\omega(t-s)i\} dU_T(\omega). \end{aligned}$$

It then follows from the Karhunen Theorem (Grenander and Rosenblatt 1984, p. 29) that there exists an orthogonal stochastic process $Z_l(\omega)$ on $[0, 1]$, such that

$$X_{lt} = \int_0^1 \exp \left\{ \frac{W_{lt}}{2} \beta_d \left(\omega, \frac{t}{T} \right) \right\} \exp(2\pi\omega t i) dZ_l(\omega).$$

This also shows that X_{lt} is a locally stationary process with spectrum $f_d(\omega, u, W_{lu})$ being $\exp \{W_{lu} \beta_d(\omega, u)\}$. The condition (1) follows from the definition of $\beta_d(\omega, u)$. By the smoothness assumption of $f(\omega, u, W_{lu})$, $|f_d(\omega, u, W_{lu}) - f(\omega, u, W_{lu})| < c(1/T)$ with c the Lipschitz constant on $[0, 1] \times [0, 1]$, which concludes the proof for condition (2).

REFERENCES

- Anderson, BO.; Moore, JB. Optimal Filtering. Prentice-Hall; Englewood Cliffs: 1979.
- Aronszajn N. Theory of Reproducing Kernels. Transactions of the American Mathematical Society 1950;68:337–404.
- Chiang CT, Rice JA, Wu CO. Smoothing Spline Estimation for Varying Coefficient Models With Repeatedly Measured Dependent Variables. Journal of the American Statistical Association 2001;96:605–619.
- Dahlhaus R. Fitting Time Series Models to Nonstationary Processes. The Annals of Statistics 1997;25:1–37.
- Dahlhaus R. A Likelihood Approximation for Locally Stationary Processes. The Annals of Statistics 2000;28:1762–1794.
- D'Alessandro, M.; Vachtsevanos, G.; Esteller, R.; Echaz, J.; Litt, B. A Generic Approach to Selecting the Optimal Feature for Epileptic Seizure Prediction. IEEE International Meeting of the Engineering in Medicine and Biology Society; Istanbul, Turkey. 2001.
- Diggle PJ, al Wasel I. Spectral Analysis of Replicated Biomedical Time Series. Applied Statistics 1997;46:31–71.

- Diggle, PJ.; Liang, KY.; Zeger, SL. Analysis of Longitudinal Data. Oxford University Press; Oxford: 1994.
- Fan J, Zhang JT. Functional Linear Models for Longitudinal Data. *Journal of the Royal Statistical Society* 2000;62:303–322.Ser. B
- Faraway JJ. Regression Analysis for a Functional Response. *Technometrics* 1997;39:254–261.
- Grenander, U.; Rosenblatt, M. Statistical Analysis of Stationary Time Series. Chelsea; New York: 1984.
- Gu, C. Smoothing Spline ANOVA Models. Springer; New York: 2002.
- Gu C, Wahba G. Semiparametric Analysis of Variance With Tensor Product Thin Plate Splines. *Journal of the Royal Statistical Society* 1993a;55:353–368.Ser. B
- Gu C, Wahba G. Smoothing Spline ANOVA With Components-Wise Bayesian Confidence Intervals. *Journal of Computational and Graphical Statistics* 1993b;2:97–117.
- Guo W, Dai M, Ombao H, von Sachs R. Smoothing Spline ANOVA for Time-Dependent Spectral Analysis. *Journal of the American Statistical Association* 2003;98:643–652.
- Hastie T, Tibshirani R. Varying-Coefficient Models” (with discussions). *Journal of the Royal Statistical Society* 1993;55:757–796.Ser. B
- Hoover DR, Rice JA, Wu CO, Yang LP. Nonparametric Smoothing Estimates of Time-Varying Coefficient Models With Longitudinal Data. *Biometrika* 1998;85:809–822.
- Koopman SJ, Durbin J. Fast Filtering and Smoothing for Multivariate State Space Models. *Journal of Time Series Analysis* 2000;21:281–296.
- Litt B, Esteller R, Echauz J, D'Alessandro M, Shor R, Henry T, Rennell P, Epstein C, Bakay R, Dichter M, Vachtsevanos G. Epileptic Seizures May Begin Hours in Advance of Clinical Onset: A Report of Five Patients. *Neuron* 2001;30:51–64. [PubMed: 11343644]
- Nychka D. Bayesian Confidence Intervals for Smoothing Spline. *Journal of the American Statistical Association* 1988;83:1134–1143.
- Ombao H, Raz J, von Sachs R, Malow B. Automatic Statistical Analysis of Bivariate Nonstationary Time Series. *Journal of the American Statistical Association* 2001;96:543–560.
- Qin L, Guo W. State Space Representation for Smoothing Spline ANOVA Models. *Journal of Computational and Graphical Statistics* 2006;15:830–847.
- Ramsay, JO.; Silverman, BW. Functional Data Analysis. Springer; Berlin: 1997.
- Shumway, RH.; Stoffer, DS. Time Series Analysis and Its Applications. Springer; New York: 2000.
- Wahba G. Automatic Smoothing of the Log Periodogram. *Journal of the American Statistical Association* 1980;75:122–132.
- Wahba G. Bayesian Confidence Intervals for the Cross-Validated Smoothing Spline. *Journal of the Royal Statistical Society* 1983;45:133–150.Ser. B
- Wahba G. A Comparison of GCV and GML for Choosing the Smoothing Parameters in the Generalized Spline Smoothing Problem. *The Annals of Statistics* 1985;13:1378–1402.
- Wahba, G. Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics. Vol. Vol. 59. SIAM; Philadelphia: 1990.
- Wecker WE, Ansley CF. The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing. *Journal of the American Statistical Association* 1983;78:81–89.
- Wei, WES. Time Series Analysis: Univariate and Multivariate Methods. 2nd ed.. Addison Wesley; Reading, MA: 2005.

Figure 1.

Ten-min IEEG time series from 36 baseline segments (13 asleep patients and 23 awake patients) and 34 pre-seizure segments (12 asleep patients and 22 awake patients). The sample rate is 200 Hz. The x -axis is time and the y -axis is amplitude. It shows that the four groups of time series have constant means but different patterns of stochastic variation over time.

Figure 2. Mean squared errors (MSEs). Panels from left to right: $A = 0, 1, 2$. For each setting, the boxplots correspond to three time-frequency grids $S \times M = 25 \times 16, 20 \times 20$, and 16×25 , respectively.

Figure 3.

Coverage of the 95% Bayesian confidence intervals. Panels from left to right represent the settings with $A = 0, 1$, and 2 each with boxplots for three time-frequency grids $S \times M = 25 \times 16, 20 \times 20$, and 16×25 , respectively.

Figure 4.

Posterior estimates of the time-frequency coefficients (middle panels) with the 95% Bayesian confidence intervals (CIs) (lower CI: left panels; upper CIs: right panels). The three rows correspond to estimates of the baseline log-spectrum of asleep patients (row 1), the log-fold spectral difference between awake and asleep patients (row 2), and the log-fold spectral difference between pre-seizure and baseline segments (row 3). The estimates are obtained for 40 time points in the time domain (x -axis) and 20 frequency points in the frequency domain (y -axis). The colors range from cold color to warm color indicating increased log-spectrum amplitude (z -axis). The baseline log-spectrum estimate of asleep patients does not vary much with time, implying these time series are approximately stationary. The confidence intervals of the log-fold spectral difference between awake and asleep patients exclude zero, which shows that in general higher brain power is accumulated when patients are awake. The log-fold spectral difference between the pre-seizure and baseline segments demonstrates the electrical activity of the brain when it approaches to the seizure onset. It shows that significantly higher power concentrates at the low frequency end (≤ 20 Hz) 10 min prior to the seizure onset, then abruptly spreads to higher frequencies at around 1 min pre-seizure.