Statistica Sinica **16**(2006), 825-845

MULTIVARIATE TIME-DEPENDENT SPECTRAL ANALYSIS USING CHOLESKY DECOMPOSITION

Wensheng Guo and Ming Dai

University of Pennsylvania and University of North Carolina at Charlotte

Abstract: In this paper, we propose a nonparametric method to estimate the spectrum of a multivariate locally stationary process. The time-varying spectrum is assumed to be smooth in both time and frequency. In order to ensure that the final estimate of the multivariate spectrum is positive definite while allowing enough flexibility in estimating each of its elements, we propose to smooth the Cholesky decomposition of an initial spectral estimate and to reconstruct the final spectral estimate from the smoothed Cholesky elements. We propose a two-stage estimation procedure. The first stage approximates the locally stationary process by a piecewise stationary time series to obtain the local estimate of the time varying spectrum and its Cholesky decomposition on discrete time-frequency points. The second stage uses a smoothing spline ANOVA to jointly smooth each Cholesky element in both time and frequency, and reconstructs the final estimate of the time varying multivariate spectrum for any time-frequency point. The final estimate is a smooth function in time and frequency, has a global interpretation, and is consistent and positive definite. We show that the Cholesky decomposition of a time varying spectrum can be used as a transfer function to generate a locally stationary time series with the designed spectrum. This not only provides us flexibility in simulations, but also allows us to construct bootstrap confidence intervals on the time varying multivariate spectrum. A simulation is conducted to investigate its performance and an application to an EEG data set recorded during an epileptic seizure is used as an illustration.

Key words and phrases: Bootstrap, Cholesky decomposition, locally stationary time series, smoothing spline, spectral analysis.

1. Introduction

Multivariate spectral analysis plays an important role in studying relationships between time series. Traditional methods are based on the assumption of second-order stationarity that seldom holds in real applications. Extensions to non-stationary settings are still very preliminary. Dahlhaus (2000) extended the univariate locally stationary process (Dahlhaus (1997)) to a multivariate setting and proposed a parametric approach for estimation. Ombao, Raz, von Sachs and Malow (2001) proposed a nonparametric approach for a bivariate locally stationary time series by using a piecewise stationary approximation. The bivariate

spectrum is only defined within a segment and, in order to ensure that the final estimate is positive definite, the same smoothing parameter is used to smooth the two spectra and the cross-spectrum. The main difficulties in the nonparametric approach are: 1) while the calculation of the initial local estimate of the time-varying spectrum often requires some segmentation, it is desired that the final estimate of the multivariate spectrum has a global interpretation; 2) it is difficult to ensure that the final estimate of the multivariate spectrum is positive definite while allowing optimal smoothing for each of its elements. In this paper, we propose to estimate the spectrum of a multivariate locally stationary process by extending the Cholesky decomposition approach (Dai and Guo (2004)) using a two-stage estimation procedure (Guo, Dai, Ombao and von Sachs (2003)). This approach not only avoids the aforementioned difficulties, but also provides much flexibility in simulations.

Our motivation comes from an interest in studying the initialization and propagation of seizures in epileptic patients by analyzing the spectrum of the electroencephalograms(EEG) data. Our eventual goal is to develop statistical methods that can predict the onset of a seizure and can localize its initiation based on the multi-channel EEG data, so that treatments can be given to prevent its onset in an early stage. Our first step is to develop a flexible method that can characterize the time varying multivariate spectrum of the EEG data recorded before, during and after a seizure. This will help us understand the mechanism on how seizures are generated and spread. Figure 1.1 shows the EEG data recorded during a seizure at the left temporal lobe (T3 channel) and the left parietal lobe (P3 channel). On these two channels, the energy increases until the seizure erupts and then gradually decreases. Litt, Esteller, Echauz, D'Alessandro, Shor, Henry, Pennell, Epstein, Bakay, Dichter and Vachtsevanos (2001) show that seizures are not generated abruptly, there are precursors in the EEG data minutes before the clinical onset of the seizure. Guo, Dai, Ombao and von Sachs (2003) show that the time varying spectrum of an epileptic EEG time series can be approximated by a two dimensional smooth function in time and frequency. We adopt the same approach in the multivariate setting and assume that the time-varying multivariate spectrum changes slowly in time and can be approximated by a smooth function in time.

In this paper, we consider a multivariate locally stationary process whose spectrum is assumed to be continuous and smooth in both time and frequency. This is a modified version of the Dahlhaus (2000) model, in which the multivariate transfer function is assumed to be a two-dimensional smooth function in time and frequency. This assumption essentially turns a locally stationary spectral estimation problem into a two-dimensional surface estimation problem

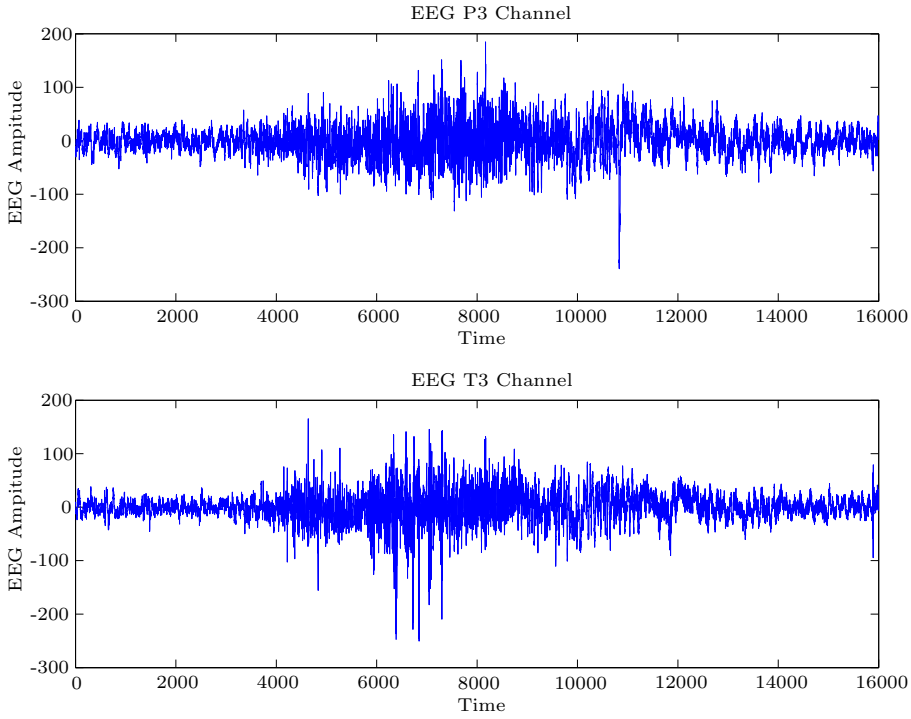

Figure 1.1. Bivariate EEG data at Channel T3 and P3 during a seizure from an epileptic patient. Number of time points $T = 16,000$, sampling rate =100 Hz.

(Guo, Dai, Ombao and von Sachs (2003)). Because of the smoothness assumption on the underlying spectrum, once we have estimated the spectrum at a given time-frequency grid, we can calculate the estimate for any time-frequency point, which leads to high computational efficiency. To ensure that the final estimate is positive definite while allowing optimal smoothing for each element, we propose to smooth the Cholesky decomposition of an initial local estimate, and reconstruct the final estimate from the smoothed Cholesky elements. The first stage of our estimation procedure approximates the locally stationary process by a piecewise stationary time series to obtain the local estimate of the time varying spectrum and its Cholesky decomposition on discrete time-frequency points. The second stage uses a smoothing spline ANOVA (Gu and Wahba (1993)) to jointly smooth each Cholesky element in both time and frequency, and reconstructs the final estimate of the time varying multivariate spectrum for any time-frequency point. The estimation procedure consists of the following five steps: (1) partition the locally stationary time series into small segments that are approximately stationary; (2) obtain a non-singular and asymptotically unbiased estimate for the spectrum using the multitaper estimate (Thomson (1982)) for each segment;

(3) calculate the initial estimate of the Cholesky decomposition of the spectrum from the raw spectral estimate within each segment; (4) jointly smooth each Cholesky decomposition element across segments using a smooth spline ANOVA which allows optimal smoothing in both time and frequency; (5) reconstruct the estimate of the spectrum from the smoothed Cholesky components. The first three steps are the first stage in obtaining a raw estimate, the last two steps are the second stage. The final estimate of the time-varying multivariate spectrum is a smooth function in both time and frequency and has a global interpretation, compared with the estimate in Ombao, Raz, von Sachs and Malow (2001) that is only defined conditional on the segmentation. It is consistent and positive definite.

Dai and Guo (2004) showed that the Cholesky decomposition of the spectrum can be used as a transfer function in stationary settings. In this paper, we show that this is also true in locally stationary settings. Using the Cholesky decomposition of the given spectrum, we can generate a locally stationary time series whose spectrum is identical to the designed spectrum at the time-frequency grid, and is close to the true one at other points. This provides great flexibility in simulations. Moreover, using the Cholesky decomposition of the estimated spectrum, we can generate bootstrap samples and construct confidence intervals for the estimated spectral components then take into account the variations in all the estimation steps.

The rest of the paper is organized as follows. In Section 2, we introduce the estimation procedure. The properties of the proposed procedure are given in Section 3. In Section 4, we propose a bootstrap procedure to construct confidence intervals for the spectrum. A simulation is offered in Section 5 and a data example is given in Section 6. Section 7 concludes the paper with some discussions. The proofs of the theorems are in the Appendix.

2. The Proposed Estimation Procedure

In this section, we describe the estimation procedure for the spectrum of a multivariate locally stationary time series. First we introduce notation: T denotes the sample size of the time series; I_k is the $k \times k$ identity matrix; unless stated otherwise, other capital symbols, such as P, F, L, A denote matrix or vectors with elements p_{kj}, f_{kj}, l_{kj}, a_{kj} or p_k, f_k, l_k, b_k; P_k denotes the sub-matrix of P consisting of the first k rows and columns of P; L' is the transpose of L, $\overline{L}$ is the conjugate of L and L^* denotes the conjugate and transpose of L.

We consider a modified Dahlhaus (2000) multivariate locally stationary time series model, where the original finite sample transfer function $A^0_{t,T}(\omega)$ is replaced by a smooth transfer function $A(t/T, \omega)$.

Definition 2.1. A real zero-mean n-dimensional stochastic process $\{X_t, t = 1, \ldots, T\}$ is called locally stationary if $X_t = \int_0^1 A(\frac{t}{T}, \omega) \exp(i2\pi\omega t) dZ(\omega)$, where the following hold.

1. $Z(\omega)$ is a zero-mean n-dimensional orthogonal increment process on $[0, 1]$ with $Z(\omega) = \overline{Z(1-\omega)}$, and $cum\{dZ(\omega_1), \ldots, dZ(\omega_k)\} = \Delta(\sum_{j=1}^{k} \omega_j)\Lambda_k(\omega_1, \cdots, \omega_{k-1})d\omega_1 \cdots d\omega_k$, where $cum\{\cdot\}$ denotes the cumulant of k-th order; $\Lambda_1 = 0$, $\Lambda_2(\omega) = I_n$, $|\Lambda_k(\omega_1, \ldots, \omega_{k-1})| \leq C_k$, C_k is a constant and $\Delta(\omega) = \sum_{j=-\infty}^{\infty} \delta(\omega+j)$ is the period 2π extension of the Dirac delta function; Λ_4 is continuous.
2. The transfer function $A(u, \omega)$, $u, \omega \in [0, 1] \times [0, 1]$, has continuous second partial derivatives w.r.t u and ω, and $\partial^l A(u, \omega)/\partial\omega^l = \overline{\partial^l A(u, 1-\omega)/\partial\omega^l}$ for $l = 0, 1$.
3. The time-dependent spectral function is $F(u, \omega) = A(u, \omega)A(u, \omega)^*$ for $(u, \omega) \in [0, 1] \times [0, 1]$.

In the original Dahlhaus model, $A^0_{t,T}(\omega)$ is only asymptotically tied to the true transfer function $A(t/T, \omega)$, because the representation in Definition 2.1 does not allow some parametric time domain models, such as multivariate time varying AR models. For a finite T, the transfer function $A^0_{t,T}(\omega)$ is only defined at t/T, for $t = 1, \ldots, T$, and is an approximation to an underlying object $A(u, \omega)$ that is only defined in the limit. In this paper, our focus is not on estimating the parameters in a time domain model, but rather in estimating the multivariate time-varying spectrum using a purely nonparametric approach. By assuming $A^0_{t,T}(\omega) = A(t/T, \omega)$ and a smooth transfer function, we can estimate $A(u, \omega)$ for any given time-frequency point (u, ω). This essentially turns a time-frequency spectral estimation problem into a 2-dimensional surface estimation problem.

From the definition, for a fixed t, there exists an underlying stationary process with the transfer function $A(t/T, \omega)$ and X_t can be viewed as one realization from this underlying stationary process. While it is impossible to estimate the spectrum from a single observation, the identifiability of the locally stationary time series model relies on the smoothness condition of the transfer function which enables one to borrow neighborhood information in estimating the spectrum.

We propose a two-stage estimation approach. In the first stage, the time series is partitioned into small segments that are approximately stationary, thus the initial spectral estimate as well as its Cholesky decomposition can be calculated within each segment. In the second stage, each Cholesky decomposition element is treated as a bivariate smooth function of time and frequency and a smoothing spline ANOVA (Gu and Wahba (1993)) is used to jointly smooth each element in time and frequency. The final spectral estimate is then reconstructed

from the smoothed Cholesky decomposition elements. The proposed procedure consists of the following five steps.

2.1. Piecewise stationary approximation

In order to obtain the local estimate of the spectrum, we first approximate the locally stationary time series $\{X_t,\, t = 1, \ldots, T\}$ as a piecewise stationary time series:

$$X_t \approx \sum_{k=1}^{\tau} \int A_k(\omega) \exp(i2\pi\omega t) dZ(\omega),$$

where $A_k(\omega) = A(u_k,\, \omega)$ if $b_k \le t < b_{k+1}$ and 0 otherwise, $u_k = (b_k + b_{k+1})/(2T)$ and $\{[b_k,\, b_{k+1}),\, k = 1, \ldots, \tau\}$ is a partition of $[1, \ldots, T]$. Write $B_k = b_{k+1} - b_k$. The initial blocking in our two-stage analysis is only used to obtain a local estimate, and optimal smoothing is performed by the smoothing spline ANOVA in the second stage, which reduces the variances of the final estimate by borrowing information across segments. Optimal segmentation in the first stage is not important to the final estimate, as long as the blocks are small enough to be approximately stationary. For $t = b_k, \ldots, b_{k+1} - 1$, the time series can be approximated by a stationary time series with the transfer function $A_k(\omega)$. Then the initial spectral estimate can be calculated by treating each segment as stationary; approximation error of the initial spectral estimate and the spectrum is given in Lemma 2.1. From our experience, when the block lengths are around $\sqrt{T}$, different partitions produce very similar average mean squared errors (see the results in Section 5, for example).

2.2. Multitaper spectral estimate

The second step is to obtain a non-singular and asymptotically unbiased initial estimate $P_k(\omega)$ for the multivariate spectrum $F(u_k,\, \omega)$ from the kth segment. The non-singularity is required to perform the Cholesky decomposition. The asymptotic unbiasedness and the smoothing step together lead to the consistency of the final estimate. In this paper, we use the multitaper estimate (Thomson (1982)).

The multitaper estimate is defined as follows. On $[b_k,\, b_{k+1})$, take a set of m tapers $\{h_k(t; j),\, t = b_k, \ldots, b_{k+1} - 1\}$, $j = 0, \ldots, m - 1$, so that

$$\sum_{t=b_k}^{b_{k+1}-1} \{h_k(t; j)\}^2 = 1, \text{ and } \sum_{t=b_k}^{b_{k+1}-1} h_k(t; j) h_k(t; v) = 0, \text{ if } j \neq v.$$

Then the $n \times n$ multitaper spectral estimate at the frequency $\omega \in [0, 1]$ on the

block is given by

$$P_k(\omega) = \frac{1}{m}\sum_{j=0}^{m-1} S_k(\omega; j)S_k(\omega; j)^*,$$

where $S_k(\omega; j)$ is the $n \times 1$ tapered Fourier transformation on the block:

$$S_k(\omega; j) = \sum_{t=b_k}^{b_{k+1}-1} h_k(t; j)X_t \exp(i2\pi\omega t).$$

The following lemma gives the properties of the multitaper spectral estimate.

Lemma 2.1. *Assume that $B_k = O(T^{1/2})$ as $T \to \infty$. The multitaper spectral estimates $P_k(\omega)$ are asymptotically independent at different time or frequency points, and the asymptotic distribution of $P_k(\omega)$ is $1/mW_c\{m, F(u_k, \omega)\}$ for $\omega \neq 0, 0.5, 1$ and $1/mW_r\{m, F(u_k, \omega)\}$ for $\omega = 0, 0.5, 1$, where $W_c(m, F)$ and $W_r(m, F)$ denote the complex and real Wishart distributions with m degrees freedoms and covariance matrix F respectively. Moreover $E\{P_k(\omega)\} = F(u_k, \omega) + O\{1/B_k\} + O\{B_k/T\}$.*

The proof is given in Appendix. The above asymptotic result does not mean that a fixed continuous time process is discretized on a finer and finer grid as $T \to \infty$. It implies that as more data are collected, there are more data locally to obtain a better local estimate within a segment, which is reflected by the increase of the block size B_k of the approximately stationary segment. The block size going to infinity at a slower rate than T ensures that the piecewise approximation is close to the underlying continuous object.

To ensure that $P_k(\omega)$ is positive definite, the number of the tapers m must be equal to or greater than n, the dimension of the time series. Unlike Thomson (1982) and Walden and McCoy (1995), who used multitapers to obtain the final estimate of the spectrum, we only use the multitapers to obtain an initial spectral estimate. Thus the choice of the tapers and the optimal number of the tapers do not have much impact on the final estimate because of the second stage smoothing. In our bivariate time series application, we use two tapers given by

$$h_k(t; j) = \sqrt{2/(B_k + 1)}\sin\{\pi(j + 1)(t - b_k + 1)/(B_k + 1)\},\ t = b_k, \ldots, b_{k+1} - 1,$$

for $j = 0, 1$. For more details about multitaper estimates, see Thomson (1982), Percival (1994), Walden and McCoy (1995) and Walden (2000).

The multitaper spectral estimate $P_k(\omega)$ is an asymptotically unbiased estimate for $F(u_k, \omega)$. In order to obtain a consistent estimate for the spectrum at any time-frequency, we need to jointly smooth the initial estimates across segments in both time and frequency. In this paper we propose to smooth the

Cholesky decomposition components instead of the raw spectral estimates directly. This guarantees that the final estimate is positive definite, while allowing flexible smoothing for each component.

2.3. Estimate Cholesky decomposition of the spectrum

At this step, a raw estimate to the Cholesky decomposition of the spectrum is obtained from $P_k(\omega)$. The estimate is similar to that in Eaton and Olkin (1987), in which they dealt with the real Wishart distribution.

When $\omega = 0, 0.5, 1$, the asymptotic distribution of $P_k(\omega)$ is $1/mW_r\{m, F(u_k, \omega)\}$. Thus the unique unbiased estimate for the Cholesky decomposition of $F(u_k, \omega)$ can be obtained as (Eaton and Olkin (1987)) $Q_k(\omega) = L_k(\omega)\Delta^{-1}$, where $L_k(\omega)$ is the Cholesky decomposition to $P_k(\omega)$ and Δ is a diagonal matrix with $\delta_{jj} = \sqrt{2}\Gamma(m/2 - j/2 + 1)\}/\{\sqrt{m}\Gamma(m/2 - j/2 + 1/2)\}$ for $j = 1, \ldots, n$.

The unbiased estimate to the Cholesky decomposition of the spectrum at other frequency points is given by the following lemma.

Lemma 2.2. *When $\omega \neq 0, 0.5, 1$, let the distribution of $P_k(\omega)$ be $1/mW_c\{m, F(u_k, \omega)\}$. Then the unbiased estimate for the Cholesky decomposition of the spectrum is given by $Q_k(\omega) = L_k(\omega)\Delta^{-1}$, where Δ is a diagonal matrix with $\delta_{jj} = \Gamma(m - j + 3/2)\}/\{\sqrt{m}\Gamma(m - j + 1)\}$.*

The two diagonal matrices for the real and complex Wishart distributions are different, since for the complex Wishart distribution $1/mW_c(m, F)$, the distribution of the diagonal is the corresponding spectrum times $1/(2m)\chi^2_{2m}$ (See Brillinger (1981, p.109, Exercise 4.8.4)); for the real Wishart distribution, the distribution of the diagonal is the corresponding spectrum times $1/m\chi^2_m$, where χ^2_k denotes the chi-square distribution with k degree of freedom.

2.4. Smoothing the Cholesky decomposition

To reconstruct a global consistent estimate for the time varying spectrum, we use a smoothing spline ANOVA (Gu and Wahba (1993)) to jointly smooth each estimate of the Cholesky decomposition element in time and frequency. We smooth $l_{\mu\mu}(\cdot, \cdot), \mu = 1, \ldots, n$, the real part of $l_{\mu\nu}(\cdot, \cdot)$: $\mathrm{Re}\{l_{\mu\nu}(\cdot, \cdot)\}$ and imaginary part of $l_{\mu\nu}(\cdot, \cdot)$: $\mathrm{Im}\{l_{\mu\nu}(\cdot, \cdot)\}$ for $\mu > \nu$ separately with their own smoothing parameters. Since the smoothing steps for these elements are the same, we suppress the subscripts in this subsection in describing the smoothing steps.

In the following, $l(\cdot, \cdot) \in [l_{\mu\mu}(\cdot, \cdot), \mathrm{Re}\{l_{\mu\nu}(\cdot, \cdot)\}, \mathrm{Im}\{l_{\mu\nu}(\cdot, \cdot)\}, \mu > \nu, \mu, \nu = 1, \ldots, n]$. We can write $l(u_k, \omega_{kj}) = q(u_k, \omega_{kj}) + e(u_k, \omega_{kj})$, where $q(u, \omega)$ is the corresponding Cholesky decomposition element of the spectrum $q(u, \omega) = \mathrm{E}\{l(u, \omega)\}$, the $e(u_k, \omega_{kj})$ are independent errors with zero mean and the variances depending on the time-frequency point (u_k, ω_{kj}). To deal with the curse of

dimensionality, we model $q(u, \omega)$ as a smoothing spline ANOVA (Gu and Wahba (1993)). For simplicity, we do not explicit model the heterogeneous variances. This is equivalent to using the average of the variances, which produces satisfactory results in our applications. Let N be the total number of the selected time-frequency points, usually determined by computational feasibility because the estimation of a smoothing spline ANOVA requires $O(N^3)$ steps and $O(N^2)$ memory, computationally expensive for large N.

To define the smoothing spline ANOVA model, we first need to define the corresponding reproducing kernels (RKs) for the time and frequency domains. In the frequency domain, the reproducing kernel Hilbert space (RKHS) $\mathcal{W}_1$ for ω can be decomposed as $\mathcal{W}_1 = 1 \oplus \mathcal{H}_1$. The RK for $\mathcal{H}_1$ is $R_1(\omega_1, \omega_2) = -\mathcal{K}_4(|\omega_1 - \omega_2|)/24$, where $\mathcal{K}_k(\cdot)$ is the k-th order Bernoulli polynomial. In the time domain, the RKHS $\mathcal{W}_2$ can be decomposed as $\mathcal{W}_2 = 1 \oplus \{u - 0.5\} \oplus \mathcal{H}_2$, where the RK for $\mathcal{H}_2$ is $R_2(u_1, u_2) = \mathcal{K}_2(u_1)\mathcal{K}_2(u_2)/4 - \mathcal{K}_4(|u_1 - u_2|)/24$.

The full tensor product RKHS for $\{u, \omega\}$ is given by

$$\begin{aligned}\mathcal{W} &= \mathcal{W}_1 \otimes \mathcal{W}_2 = [\{1\} \oplus \mathcal{H}_1] \otimes [\{1\} \oplus \{u - 0.5\} \oplus \mathcal{H}_2] \\ &= \{1\} \oplus \{u - 0.5\} \oplus \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \{\mathcal{H}_1 \otimes (u - 0.5)\} \oplus \{\mathcal{H}_1 \otimes \mathcal{H}_2\}.\end{aligned}$$

The RK for $\mathcal{H}_3 = \mathcal{H}_1 \otimes (u-0.5)$ is $R_3((u_1, \omega_1), (u_2, \omega_2)) = R_1(\omega_1, \omega_2)(u_1-0.5)(u_2-0.5)$, the RK for $\mathcal{H}_4 = \mathcal{H}_1 \otimes \mathcal{H}_2$ is $R_4((u_1, \omega_1), (u_2, \omega_2)) = R_1(\omega_1, \omega_2)R_2(u_1, u_2)$.

Correspondingly, $q(u, \omega)$ has the ANOVA decomposition

$$q(u, \omega) = d_1 + d_2(u - 0.5) + q_1(\omega) + q_2(u) + q_3(u, \omega) + q_4(u, \omega),$$

where $d_1 + d_2(u - 0.5)$ is the linear trend, $q_1(\omega)$ is the smooth main effect for frequency, $q_2(u)$ is the smooth main effect for time, $q_3(u, \omega)$ is smooth in frequency and linear in time, $q_4(u, \omega)$ is the interaction that is smooth in both time and frequency.

Remark. when smoothing the imaginary parts of the cross-spectrum, because the mean is zero in the frequency domain, $\mathcal{W}_1 = \mathcal{H}_1$, and therefore the final tensor product model is $q(u, \omega) = q_1(\omega) + q_3(u, \omega) + q_4(u, \omega)$.

The estimate $\hat{q}(u, \omega)$ is found by minimizing the penalized least squares

$$\frac{1}{N} \sum_{k,j} \{l(u_k, \omega_{kj}) - \hat{q}(u_k, \omega_{kj})\}^2 + \sum_{v=1}^{4} \lambda_v ||J_v \hat{q}||^2,$$

where J_v is the orthogonal projection operator onto $\mathcal{H}_v$. The estimate at time-frequency point (u, ω) is $\hat{(u, \omega)} = D\Phi + C\xi$, where $\Phi = \binom{1}{u-0.5}$, $\xi = \sum_{v=1}^{4} \theta_v R_v$ $\{(u, \omega), (\boldsymbol{u}, \boldsymbol{\omega})\}$, with $\boldsymbol{u}$ being the collection of all time points and $\boldsymbol{\omega}$ being the

collection of all frequency points, $\theta_v = \lambda/\lambda_v$. Here C, D are the solutions to the following equations:

$$\Sigma C + WD = \mathbf{l}, \qquad W'C = 0,$$

where $\Sigma = \sum_{v=1}^{4} \theta_v R_v\{(\boldsymbol{u}, \boldsymbol{\omega}), (\boldsymbol{u}, \boldsymbol{\omega})\} + N\lambda I_N$, $W = (\mathbf{1}, \boldsymbol{u} - 0.5)$ is a $N \times 2$ matrix, $\mathbf{l} = \{l(\boldsymbol{u}, \boldsymbol{\omega})\}$ is an $N \times 1$ vector that contains the values of $l(u_k, \omega_{kj})$ evaluated at all design points, λ and θ_k are smoothing parameters which can be chosen by generalized cross validation (GCV) or by generalized maximum likelihood (GML). In our simulations, GML outperforms GCV and is recommended. The GML criterion is

$$M(\lambda, \theta_v, v = 1, \ldots, 4) = \frac{\mathbf{1}'G\mathbf{1}}{\det^+(G)^{(N-1)^{-1}}},$$

where $G = N\lambda\{\Sigma^{-1} - \Sigma^{-1}W(W'\Sigma^{-1}W)^{-1}W'\Sigma^{-1}\}$, and $\det^+$ is the product of the nonzero eigenvalues.

2.5. Reconstruction

After we obtain the smoothed estimates $\hat{Q}(u, \omega) = \{\hat{q}_{kj}(u, \omega)\}_{k,j=1}^{n}$ for the Cholesky decomposition of the spectrum, the estimate of the spectrum can be reconstructed as $\hat{F}(u, \omega) = \hat{Q}(u, \omega)\{\hat{Q}(u, \omega)\}^*$.

3. Properties

The following theorem shows that the Cholesky decomposition of multivariate time varying spectrum can be used as a transfer function to generate a multivariate locally stationary time series with the designed spectrum. This also allows us to construct bootstrap confidence intervals, which we discuss in Section 4.

Theorem 3.1. *Let $F(u, \omega)$, $(u, \omega) \in [0, 1] \times [0, 1]$, be a positive definite spectrum with continuous up-to-second order partial derivatives w.r.t u and ω, and let $\Psi(u, \omega)$ be the Cholesky decomposition of $F(u, \omega)$. Define $\{Y_t, t = 1, \ldots, T\}$ by*

$$Y_t = \sum_{k=1}^{T} \Psi(\frac{t}{T}, \frac{k}{T}) \exp(i\frac{2\pi kt}{T}) Z(k), \tag{1}$$

where $Z(k)$, $k = 1, \ldots, T$, are independent. For $k/T \neq 0, 0.5, 1$, the distribution of $Z(k)$ is complex normal with mean zero and covariance matrix $1/TI_n$ and $Z(k) = \overline{Z(T-k)}$; for $k/T = 0, 0.5, 1$, the distribution of $Z(k)$ is real normal with mean zero and covariance matrix $1/TI_n$. The series $\{Y_t, t = 1, \ldots, T\}$ is locally stationary, and satifies

1. *$F_T(t/T, k/T) = F(t/T, k/T)$ for $t, k = 1, \ldots, T$, where $F_T(u, \omega)$ is the spectrum of Y_t;*

2. $|F_T(u, \omega) - F(u, \omega)| = O(T^{-1})$ *for any* $(u, \omega) \in [0, 1] \times [0, 1]$.

The next theorem shows that the final estimate $\hat{F}(u, \omega)$ is a consistent estimate for $F(u, \omega)$.

Theorem 3.2. *Let* $T, N \to \infty$, *and* $\min(B_k) \to \infty$, *and* $\max(B_k) = O(T^{1/2})$. *Then the following hold for* $k, j = 1, \ldots, n$.

1. *The optimal convergence rate of the estimate for the spectrum in terms of the integrated mean squared error is given by*

$$\int_0^1 \int_0^1 \mathrm{E}\{\hat{f}_{kj}(u, \omega) - f_{kj}(u, \omega)\}^2 du d\omega$$
$$= \mathrm{O}(\frac{N}{\log N})^{-4/5} + \mathrm{O}\{\frac{1}{\min(B_k)^2}\} + \mathrm{O}\{\frac{\max(B_k)^2}{T^2}\}.$$

2. *When* $N = T$ *and* $B_k = B$, *the optimal convergence rate in term of mean square integrated errors is obtained by choosing* $B = O(T^{1/2})$ *and*

$$\int_0^1 \int_0^1 \mathrm{E}\{\hat{f}_{kj}(u, \omega) - f_{kj}(u, \omega)\}^2 du d\omega = \mathrm{O}\Big(\frac{T}{\log T}\Big)^{-\frac{4}{5}}.$$

The proof is given in the Appendix. According to this result, the rate of convergence is determined more by the number of points used in the second stage than by the block size.

4. Inference for the Spectrum

As the estimation of the multivariate time varying spectrum includes multiple steps that can introduce variabilities, it is desirable to construct confidence intervals that can take this into account. In this section, we propose a bootstrap procedure to construct confidence intervals on the time varying spectrum. The method is an application of Theorem 3.1 except that we generate bootstrap samples using the Cholesky decomposition of the estimated time varying spectrum. The procedure is summarized into the following five steps.

Step 1. Obtain a consistent estimate $\hat{Q}(t/T, k/T)$, $t, k = 1, \ldots, T$ for the Cholesky decomposition of the spectrum using the procedure proposed in Section 2.
Step 2. Generate a new time series using $\hat{Q}(t/T, k/T)$ as the transfer function:

$$X_t^v = \sum_{k=1}^{T} \hat{Q}(\frac{t}{T}, \frac{k}{T}) \exp(i\frac{2\pi kt}{T}) Z^v(k),$$

where $Z^v(k), k = 1, \ldots, T$, are independent, and for $k/T \neq 0, 0.5, 1$, the distribution is complex normal with zero mean and covariance matrix $1/TI_n$; for

$k/T = 0, 0.5, 1$, the distribution is real normal with zero mean and covariance matrix $1/TI_n$.

Step 3. Compute the estimate of the spectrum of X_t^v using the procedure in Section 2, and denote the estimate as $\hat{F}^v(u, \omega)$.

Step 4. Continue Step 2, Step 3 and obtain the estimates $\{\hat{F}^v(u,\omega),\ v = 1,\ldots,M\}$, from which the $1-\alpha$ bootstrap confidence intervals can be formed.

5. Simulations

In this section, we conduct a simulation to investigate the performance of our proposed method. From Theorem 3.1, we can generate a multivariate locally stationary time series with any given spectrum. We simulate bivariate time series with the following designed spectra, specified by a parameter a controlling the rate of the change of the spectrum over time.

$$
\begin{aligned}
f_{11}(u,\omega) &= \{[1.2\cos(\pi\omega)]^2 + a\sin(2.0\pi uT) + 0.7\}^2;\\
f_{21}(u,\omega) &= \{0.6\cos(2\pi\omega) + a\cos(2\pi uT) + 1 + ia\sin(2\pi\omega)[4(uT-0.5)^2 + 0.5]\}\\
&\qquad \{[1.2\cos(\pi\omega)]^2 + a\sin(2.0\pi uT) + 0.7\};\\
f_{22}(u,\omega) &= \{[1.3\cos(2\pi\omega)]^2 + a\sin(2.0\pi uT) + 0.8\}^2 + [0.6\cos(2\pi\omega)\\
&\qquad + a\cos(2\pi uT) + 1]^2 + \{a\sin(2\pi\omega)[4(uT-0.5)^2 + 0.5]\}^2.
\end{aligned}
$$

We chose $a = 0.1, 0.4$ and 1, where $a = 0.1$ corresponds to a nearly stationary time series and $a = 1.0$ results in a rapidly changing time series. The Cholesky decomposition $\Psi(u, \omega)$ of the spectrum is, respectively,

$$
\begin{aligned}
\psi_{11}(u,\omega) &= \{1.2\cos(\pi\omega)\}^2 + a\sin(2.0\pi uT) + 0.7,\\
\psi_{21}(u,\omega) &= 0.6\cos(2\pi\omega) + a\cos(2\pi uT) + 1 + ia\sin(2\pi\omega)\{4(uT-0.5)^2 + 0.5\},\\
\psi_{22}(u,\omega) &= \{1.3\cos(2\pi\omega)\}^2 + a\sin(2.0\pi uT) + 0.8.
\end{aligned}
$$

From $\Psi(u, \omega)$, we generate the time series for $T = 1,024, 2,048$ given by

$$
X_t = \begin{Bmatrix} x_1(t) \\ x_2(t) \end{Bmatrix} = \sum_{k=1}^{T} \Psi(\frac{t}{T}, \frac{k}{T}) \exp(i\frac{2\pi kt}{T}) Z(k),
$$

where $\{Z(k),\ k = 1,\ldots,T\}$ are independent. For $k/T \neq 0, 0.5, 1$, the distribution of $Z(k)$ is bivariate complex normal with zero mean and covariance $1/TI_2$; for $k/T = 0, 0.5, 1$, the distribution of $Z(k)$ is bivariate real normal with zero mean and covariance $1/TI_2$.

We partitioned the time series in equal length blocks. Under each setting, we used two different partitions: for $T = 1,024$, we first used 16 blocks, and 32 frequency points in each block; we then used 16 block with 64 frequency points.

For $T = 2,048$, the first setting used 16 blocks with 64 frequency points, and the second setting used 32 blocks with 32 time points. For each case, we simulated 100 time series, and calculated the average mean squared errors. Each case takes about 20 minutes to finish on a typical PC. The results are given in Figure 5.1.

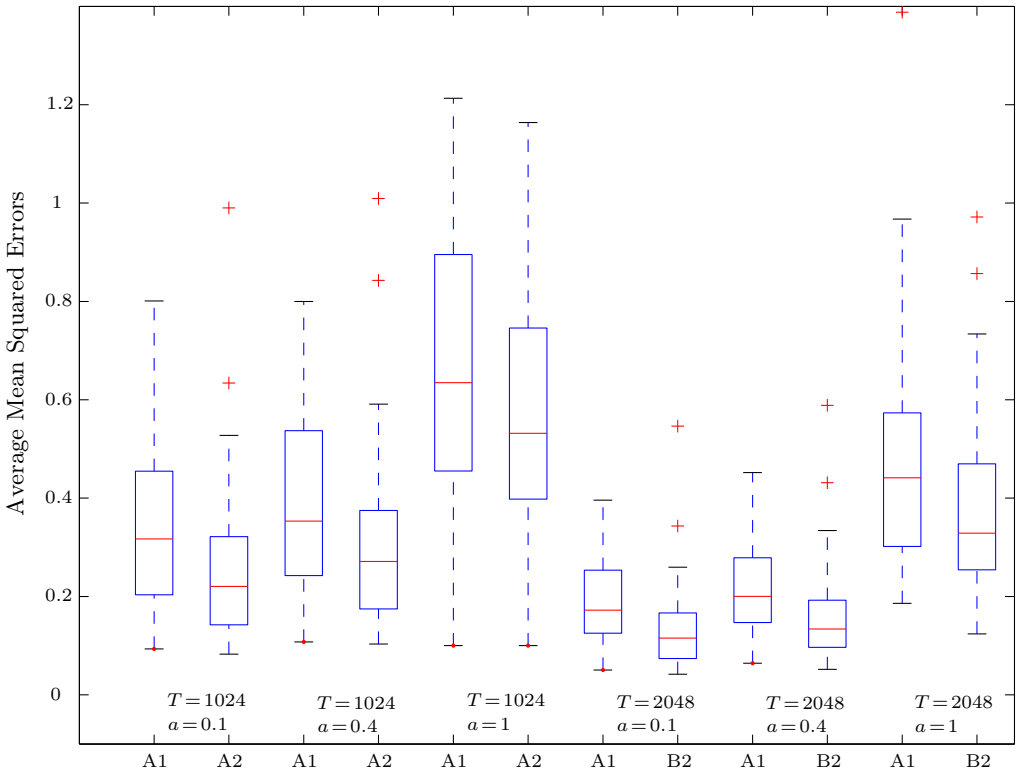

Figure 5.1. Boxplot of the average mean squared errors. A1: 16 blocks, 32 frequency points; A2: 16 blocks, 64 frequency points; B2: 32 blocks, 32 frequency points.

From a plot, we can see that that the average mean squared errors for $T = 2,048$ are smaller than those for $T = 1,024$, even if the total number of time-frequency points are the same. For the same time series with different partitions, the more time-frequency points used, the smaller the average mean squared errors. This confirms the results from Theorem 3.2 that the average mean squared errors are determined by both the length of the time series and the total number of time-frequency points. In comparing the three choices of a, we find that our method gives similar results for $a = 0.1$ and $a = 0.4$, where the spectrum varies slowly over time. The mean squared errors are substantially

larger when $a = 1$. This means that when the spectrum changes rapidly over time, a locally stationary model may not be a good approximation. A new model, a faster sampling rate, or a longer time series may be needed.

Because of heavy computational demands, it is unreasonable to calculate the bootstrap confidence intervals for all the simulations. We only calculate the 95% bootstrap confidence intervals using 100 bootstrap draws for the first simulated series under each setting. The percentages of the coverage under the 12 settings are close to the nominal level: 97.66, 95.26, 97.45, 96.79, 96.44, 93.12, 97.9, 96.12, 97.22, 95.83, 96.22, 95.75 and 94.24.

6. EEG Data

We apply our proposed method to the bivariate EEG time series shown in Figure 1.1. The sampling rate is 100 Hz and the length of the time series $T = 16,000$. The time series was partitioned into 40 equal-sized blocks and, on each block, 40 frequency points were selected to calculate the initial and final spectral estimates. Figure 6.1 shows the estimates of the spectrum and cross-spectrum, together with their 95% bootstrap confidence intervals using 500 draws. The first time series is the P3 channel and the second is the T3 channel. Before the seizure the energy is mainly concentrated at the lower frequencies. During the seizure the energy spreads to all the frequencies, but the power at the lower frequencies is still much greater than that at the high frequencies. It is interesting to see that the energy at all frequencies increases by about the same magnitude in both channels. In examining the cross spectrum, the imaginary part is essentially zero because of the associated wide confidence intervals. This indicates that the spread of the seizure is almost instantaneously and it is difficult to detect the lag between the two time series although, from clinical knowledge, it is known that in this patient P3 is the leading channel. The real part of the cross spectrum takes a similar shape as do the two spectra, but only at about half the magnitude. This indicates that the two channels synchronize mainly in the lower frequencies before the seizure while this synchrony spreads to the higher frequencies during the seizure. This is consistent with current knowledge on seizure propagation.

7. Discussion

Computational intensity of the smoothing spline ANOVA prevents us from directly applying the proposed method to a long time series without choosing a very coarse initial time-frequency grid. We are developing an $O(N)$ algorithm for the smoothing spline ANOVA model by constructing equivalent state space models. The recursive nature of the state space model will also enable the method to be implemented online.

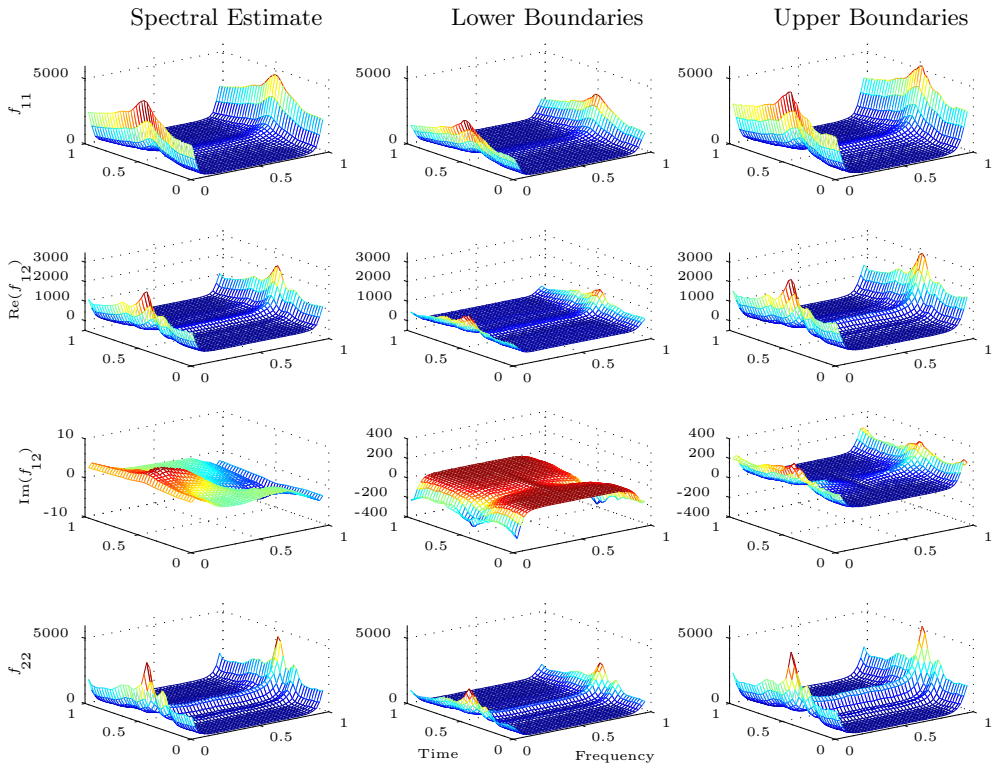

Figure 6.1. Left: The spectral estimates for the EEG data. Middle: lower confidence intervals. Right: upper confidence intervals.

We have not explicitly modeled the heterogeneous variances of the Cholesky elements in this paper. This is equivalent to using the average of the variances. Because of the robustness of smoothing spline ANOVA models, the proposed method usually produces satisfactory results in our simulations and applications. However, we expect that explicitly modeling the heteroscedasticity may improve the efficiency of the estimates, and we will pursue this in our future research. The smoothness assumption of the underlying transfer function can also be relaxed to include a finite numbers of abrupt changes in time using an approach similar to that of Ombao, Raz, von Sachs and Malow (2001). The main difference is that we need to replace the smoothing spline ANOVA by a wavelet based method (e.g. von Sachs and MacGibbon (2000)) in smoothing the Cholesky elements. The other steps in the estimation procedure remain unchanged.

Acknowledgement

This research was supported by NIH R01 grant CA84438. We would like to

thank the editor and the two referees for helpful comments.

Appendix. Proof of Results

Proof of Lemma 2.1. First we have that

$$\begin{aligned} S_k(\omega; j) &= \sum_{t=b_k}^{b_{k+1}-1} h_k(t; j) X_t \exp(-i2\pi\omega t) \\ &= \sum_{t=b_k}^{b_{k+1}-1} h_k(t; j) \int_0^1 A(\frac{t}{T}, v) \exp\{i2\pi(v-\omega)t\} dZ(v). \end{aligned}$$

If $\tilde{S}_k(\omega; j) = \sum_{t=b_k}^{b_{k+1}-1} h_k(t; j) \int_0^1 A(u_k, v) \exp\{i2\pi(v-\omega)t\} dZ(v)$. Then

$$S_k(\omega; j) - \tilde{S}_k(\omega; j) = \sum_{t=b_k}^{b_{k+1}-1} h_k(t; j) \int_0^1 \{A(\frac{t}{T}, v) - A(u_k, v)\} \exp\{i2\pi(\omega - v)t\} dZ(v).$$

Since $\mathrm{E}S_k(\omega; j) = \mathrm{E}\tilde{S}_k(\omega; j) = 0$, and $A(u, \omega)$ is Lipschitz continuous w.r.t. u, we have that

$$\begin{aligned} &\mathrm{var}\Big\{ S_k(\omega; j) - \tilde{S}_k(\omega; j) \Big\} = \sum_{t,\, s=b_k}^{b_{k+1}-1} h_k(t; j) h_k(s; j) \int_0^1 \Big[\Big\{ A(\frac{t}{T}, v) - A(u_k, v) \Big\} \\ &\qquad \exp\{i2\pi(v-\omega)(t-s)\} \Big\{ A(\frac{s}{T}, v)^* - A(u_k, v)^* \Big\} \Big] dv \\ &\le \sum_{t,\, s=b_k}^{b_{k+1}-1} |h_k(t; j) h_k(s; j)| \int_0^1 \Big| \Big\{ A(\frac{t}{T}, v) - A(u_k, v) \Big\} \Big\{ A(\frac{s}{T}, v)^* - A(u_k, v)^* \Big\} \Big| dv \\ &= \mathrm{O}(\frac{B_k}{T})^2 \sum_{t,\, s=b_k}^{b_{k+1}-1} |h_k(t; j) h_k(s; j)| \le \mathrm{O}(\frac{B_k}{T})^2 \sum_{t,\, s=b_k}^{b_{k+1}-1} [\{h_k(t; j)\}^2 + \{h_k(s; j)\}^2] \\ &= \mathrm{O}\{\frac{(B_k)^3}{T^2}\}. \end{aligned}$$

Since $B_k^2 = \mathrm{O}(T)$, $\mathrm{O}(B_k^3/T^2) = \mathrm{O}(B_k/T)$. So we have that $S_k(\omega; j) - \tilde{S}_k(\omega; j) \to 0$ in probability. From Walden (2000) and Dahlhaus (1997), $\tilde{S}_k(\omega; j)$ are asymptotically independent at different frequency, different block and different k, the asymptotic distribution is normal with mean zero and covariance matrix $F(u_k, \omega)$. So we have that the $S_k(\omega; j)$ are asymptotically independent, the asymptotic distribution is normal with mean zero and covariance matrix $F(u_k, \omega)$. The asymptotic distribution of $P_k(\omega)$ then follows.

To find the order of $\mathrm{E}\{S_k(\omega; j)S_k(\omega; j)^*\} - F(u_k, \omega)$, we have

$$\begin{aligned}
&\mathrm{E}\{S_k(\omega; j)S_k(\omega; j)^*\} \\
&= \sum_{t,\, s=b_k}^{b_{k+1}-1} h_k(t; j)h_k(s; j) \int_0^1 \Big[A(\frac{t}{T}, v)A(\frac{s}{T}, v)^* \exp\{i2\pi(v-\omega)(t-s)\}\Big]dv.
\end{aligned}$$

Since $A(u, v)$ has second continuous partial derivative w.r.t. u, we can write $A(t/T, v) = A(u_k, v) + \partial A(u_t^*, v)/\partial u(t/T - u_k)$, where u_t^* is some point between u_k and t/T. Thus the above expansion is

$$\begin{aligned}
&\sum_{t,\, s=b_k}^{b_{k+1}-1} h_k(t; j)h_k(s; j) \int_0^1 \Big[\Big\{A(u_k, v) + \frac{\partial A(u_t^*, v)}{\partial u}(\frac{t}{T} - u_k)\Big\} \\
&\qquad \times \Big\{\overline{A(u_k, v)} + \frac{\partial A(u_s^*, v)^*}{\partial u}(\frac{s}{T} - u_k)\Big\} \exp\{i2\pi(v-\omega)(t-s)\}\Big]dv \\
&= \sum_{t,\, s=b_k}^{b_{k+1}-1} h_k(t; j)h_k(s; j) \int_0^1 F(u_k, v) \exp\{i2\pi(v-\omega)(t-s)\}dv + \mathrm{O}(\frac{B_k^3}{T^2}) \\
&\quad + \sum_{t=b_k}^{b_{k+1}-1} h_k(t; j) \int_0^1 \frac{\partial A(u_t^*, v)}{\partial u}(\frac{t}{T} - u_k)A(u_k, v)^* \\
&\qquad\qquad \sum_{s=b_k}^{b_{k+1}-1} h_k(s; j) \exp\{i2\pi(v-\omega)(t-s)\}dv \\
&\quad + \sum_{s=b_k}^{b_{k+1}-1} h_k(s; j) \int_0^1 \frac{\partial A(u_s^*, v)^*}{\partial u}(\frac{s}{T} - u_k)A(u_k, v) \\
&\qquad\qquad \sum_{t=b_k}^{b_{k+1}-1} h_k(t; j) \exp\{i2\pi(v-\omega)(t-s)\}dv \\
&= F(u_k, \omega) + \mathrm{O}(\frac{1}{B_k}) + \mathrm{O}(\frac{B_k}{T}) \\
&\quad + \sum_{t=b_k}^{b_{k+1}-1} \frac{h_k(t; j)}{B_k}\Big\{\frac{\partial A(u_t^*, \omega)}{\partial u}(\frac{t}{T} - u_k)A(u_k, \omega)^* + \mathrm{O}(\frac{1}{B_k})\Big\} \\
&\quad + \sum_{s=b_k}^{b_{k+1}-1} \frac{h_k(s; j)}{B_k}\Big\{\frac{\partial A(u_s^*, \omega)^*}{\partial u}(\frac{s}{T} - u_k)A(u_k, \omega) + \mathrm{O}(\frac{1}{B_k})\Big\} \\
&= F(u_k, \omega) + \mathrm{O}(\frac{1}{B_k}) + \mathrm{O}(\frac{B_k}{T}).
\end{aligned}$$

So $E\{P_k(\omega)\} = F(u_k, \omega) + \mathrm{O}(1/B_k) + \mathrm{O}(B_k/T)$.

Proof of Theorem 2.2. Since the distribution of P is $1/mW_c(m, F)$, there exists D_q, $q = 1, \ldots, m$, which are independently distributed as complex normal with zero mean and covariance matrix F such that $P = 1/m \sum_{q=1}^{m} D_q D_q^*$. The result follows by induction.

The distribution of p_{11} is $f_{11}\chi^2_{2m}/(2m)$, where χ^2_k denotes the chi-square distribution with k degrees of freedoms, (Brillinger (1981, Exercise 4.8.4)). The probability density function of χ^2_{2m} is

$$g(x) = \frac{1}{\Gamma(m)2^m} x^{m-1} \exp(-x/2),\ x > 0.$$

The first Cholesky decomposition $\gamma_{11} = (p_{11})^{1/2}$. It is easy to find that $\mathrm{E}(\gamma_{11}) = \delta_1 (f_{11})^{1/2}$ where $\delta_1 = \Gamma(m+1/2)/\{(m)^{1/2}\Gamma(m)\}$. Thus the result holds for γ_{11}.

Suppose the result holds for k, i.e., for $\Gamma_k = \{\gamma_{sj}\}^k_{s,\,j=1}$ $L_k = \{l_{sj}\}^k_{s,\,j=1}$ $\Delta_k = \{\delta_{sj}\}^k_{s,\,j=1}$ and $L_k = \Gamma_k \Delta_k^{-1}$, $\mathrm{E}(L_k) = Q_k$, where $Q_k = \{q_{sj}\}^k_{s,\,j=1}$. Since Δ_k is a constant matrix, we have that $\mathrm{E}(\Gamma_k) = Q_k \Delta_k$.

Rewrite

$$\Gamma_{k+1} = \begin{pmatrix} \Gamma_k & 0 \\ \Gamma_{k+1,1} & \gamma_{k+1,k+1} \end{pmatrix},\ P_{k+1} = \begin{pmatrix} P_k & P^*_{k+1,1} \\ P_{k+1,1} & p_{k+1,k+1} \end{pmatrix},\ F_{k+1} = \begin{pmatrix} F_k & F^*_{k+1,1} \\ F_{k+1,1} & f_{k+1,k+1} \end{pmatrix},$$
$$Q_{k+1} = \begin{pmatrix} Q_k & 0 \\ Q_{k+1,1} & q_{k+1,k+1} \end{pmatrix}.\ L_{k+1} = \begin{pmatrix} L_k & 0 \\ L_{k+1,1} & l_{k+1,k+1} \end{pmatrix},\ \Delta_{k+1} = \begin{pmatrix} \Delta_k & 0 \\ 0 & \delta_{k+1,k+1} \end{pmatrix}.$$

To calculate the expectation for $L_{k+1,1}$, for $s = 1, \ldots, m$, let H_s be the $k \times 1$ vector consisting of the first k elements of D_s and h_s be the $(k+1)th$ element of D_s. Thus the distribution of $\begin{pmatrix} H_s \\ h_s \end{pmatrix}$ is normal with zero mean and covariance matrix F_{k+1}. Let $\tilde{h}_s = h_s - F_{k+1,1} F_k^{-1} H_s$, $s = 1, \ldots, m$. Then $\tilde{h}_s$ is independent of H_s and the mean is zero. Thus from the definition of the Cholesky decomposition, we have

$$\begin{aligned} \Gamma_{k+1,1} = P_{k+1,1}(\Gamma_k^*)^{-1} &= \frac{1}{m} \sum_{s=1}^{m} h_s H_s^* (\Gamma_k^*)^{-1} \\ &= \frac{1}{m} \sum_{s=1}^{m} (\tilde{h}_s + F_{k+1,1} F_k^{-1} H_s) H_s^* (\Gamma_k^*)^{-1}, \\ \mathrm{E}(\Gamma_{k+1,1}) = \frac{1}{m} \sum_{s=1}^{m} \mathrm{E}\{\tilde{h}_s H_s^* (\Gamma_k^*)^{-1}\} &+ \mathrm{E}\{\frac{1}{m} \sum_{s=1}^{m} F_{k+1,1} F_k^{-1} H_s H_s^* (\Gamma_k^*)^{-1}\}. \end{aligned}$$

Since $\mathrm{E}(\tilde{h}_s) = 0$ for $s = 1, \ldots, m$, $\{\tilde{h}_s,\ s = 1, \ldots, m\}$ are independent of $\{H_s,\ s = 1, \ldots, m\}$, and Γ_k only depends on $\{H_s,\ s = 1, \ldots, m\}$, the last expression is

$$\begin{aligned} F_{k+1,1} F_k^{-1} E\{\frac{1}{m} \sum_{s=1}^{m} H_s H_s^* (\Gamma_k^*)^{-1}\} &= F_{k+1,1} F_k^{-1} E\{P_k (\Gamma_k^*)^{-1}\} \\ = F_{k+1,1} F_k^{-1} E\{\Gamma_k \Gamma_k^* (\Gamma_k^*)^{-1}\} &= F_{k+1,1} F_k^{-1} E(\Gamma_k) = F_{k+1,1} F_k^{-1} Q_k \Delta_k. \end{aligned} \quad \text{(A.1)}$$

From the definition of the Cholesky decomposition, we have that $\mathrm{E}(L_{k+1,1}) = \mathrm{E}(\Gamma_{k+1,1})\Delta_k^{-1} = Q_{k+1,1}$.

Since $\gamma^2_{k+1,k+1} = p_{k+1,k+1} - P_{k+1,1}P_k^{-1}P^*_{k+1,1}$ and the distribution of P_{k+1} is $1/mW_c(m, F_{k+1})$, the distribution of $\gamma^2_{k+1,k+1}$ is $1/mW_c(m-k, f_{k+1,k+1} - F_{k+1,1}F_k^{-1}F^*_{k+1,1})$, Brillinger (1981, Exercise 4.8.8). From Brillinger (1981) Exercise 4.8.4, the distribution of $\gamma^2_{k+1,k+1}$ is $1/(2m)(f_{k+1,k+1} - F_{k+1,1}F_k^{-1}F^*_{k+1,1})$ $\chi^2_{2(m-k)}$. As with γ_{11}, it is easy to find that

$$\mathrm{E}(\gamma_{k+1,k+1}) = \delta_{k+1}q_{k+1,k+1}, \tag{A.2}$$

where $\delta_{k+1} = \Gamma(m-k+1/2)/\{(m)^{1/2}\Gamma(m-k)\}$.

From Equations (A.1) and (A.2), we have $\mathrm{E}(L_{k+1}) = Q_{k+1}$. This proves the result.

Proof of Theorem 3.1. From (1), we have that for $t,\ s = 1, \ldots, T$,

$$\mathrm{Cov}\,(Y_t,\, Y_s) = \frac{1}{T}\sum_{k=1}^{T} \Psi(\frac{t}{T},\, \frac{k}{T})\Psi(\frac{s}{T},\, \frac{k}{T})^* \exp\{i\frac{2\pi k(t-s)}{T}\}. \tag{A.3}$$

Let $V_T(u,\, \omega)$ be a smooth $n \times n$ matrix passing through $\{t/T,\, k/T,\, \Psi(t/T,\, k/T)\}$, $t,\, k = 1, \ldots, T$, and let $U_T(\omega)$ be the countting measure on $[0,1]$ that has jump $1/T$ at $(k/T,\, k = 1, \ldots, T)$. Then (A.3) can be written as

$$\mathrm{Cov}\,(Y_t,\, Y_s) = \int_0^1 V_T(\frac{t}{T},\, \omega)V_T(\frac{s}{T},\, \omega)^* \exp\{i2\pi\omega(t-s)\}dU_T(\omega).$$

From Karhunen's Theorem (Grenander and Rosenblatt (1984)), there exists an orthogonal process $Z(\omega)$ on $[0,1]$ such that $Y_t = \int_0^1 V_T(t/T,\, \omega)\exp(i2\pi\omega t)dZ(\omega)$, Thus Y_t is locally stationary and the spectrum $F_T(u,\, \omega)$ of the time series Y_t is $V_T(u,\, \omega)V_T(u,\, \omega)^*$. From the construction of the function $V_T(u,\, \omega)$, we have that for $t,\, k = 1, \ldots, T$, $F_T(t/T,\, k/T) = F(t/T,\, k/T)$. This proves the first part of the theorem.

The second part of the theorem follows from the Lipschitz continuity of $F_T(u,\omega)$ and $F(u,\omega)$, which is implied by the smoothness condition on $F(u,\omega)$.

The proof of Theorem 3.2 is based on the following lemma and the consistency of the smoothed Cholesky decomposition of the initial spectral estimate.

Lemma A.1. *For a given constant c, $0 < c < 1$, if for all $k,\, j = 1, \ldots, n$, $\int_0^1\int_0^1 E\{\hat{l}_{kj}(u,\, \omega) - q_{kj}(u,\omega)\}^2 dud\omega = O(T^{-c})$, then $\int_0^1\int_0^1 E(\hat{f}_{kj}(u,\omega) - f_{kj}(u,\omega))^2$ $dud\omega = O(T^{-c})$ for all $k,\, j$.*

Proof. The proof of this Theorem is based on a Taylor expansion. It is obvious that that the spectrum $F(u,\, \omega) = \{f_{kj}(u,\, \omega)\}_{k,\,j=1}^n$ is a known function of $\{q_{kj}(u,\, \omega))\}_{k,\,j=1}^n$, the Cholesky decomposition of the spectrum, so let

$f_{kj}(u, \omega) = f_{kj}(q_{vs}(u, \omega), v, s = 1, \ldots, n)$. From a Taylor expansion, there exist $(b_{vs}, v, s = 1, \ldots, n)$ such that

$$\hat{f}_{kj}(u, \omega) = f_{kj}(\hat{l}_{vs}(u, \omega), v, s = 1, \ldots, n) = f_{kj}(q_{vs}(u, \omega), v, s = 1, \ldots, n) + \sum_{v, s=1}^{n} \frac{\partial f_{kj}(u, \omega)}{\partial q_{vs}(u, \omega)}\Big|_{b_{vs}, v, s=1,\ldots,n} \{\hat{l}_{vs}(u, \omega) - q_{vs}(u, \omega)\}.$$

Then we have $\mathrm{E}\{\hat{f}_{kj}(u, \omega) - f_{kj}(u, \omega)\}^2 \leq n^2 C \sum_{v, s=1}^{n} \mathrm{E}\{\hat{l}_{vs}(u, \omega) - q_{vs}(u, \omega)\}^2$, where $C = \max_{v, s} \sup\{\partial f_{kj}(u, \omega)/\partial q_{vs}(u, \omega)\}^2$. The result follows.

Proof of Theorem 3.2. The convergence rate for the smoothed Cholesky decomposition was given by Lin (2000). This can be stated as

$$\int_0^1 \int_0^1 \mathrm{E}|\hat{q}_{kj}(u, \omega) - q_{kj}(u, \omega)|^2 du d\omega = \mathrm{O}\{(\frac{N}{\log N})^{-\frac{4}{5}}\}, \tag{A.4}$$

for any $k, j = 1, \ldots, n$. The result then follows. The second part of the theorem follows from Lemma A.1, Lemma 2.1 and Equation (A.4).

When $N = T$ and $B_k \equiv B$, then

$$\int_0^1 \int_0^1 \mathrm{E}|\hat{f}_{kj}(u, \omega) - f_{kj}(u, \omega)|^2 du d\omega = \mathrm{O}\{(\frac{T}{\log T})^{-\frac{4}{5}}\} + \mathrm{O}(\frac{1}{B^2}) + \mathrm{O}(\frac{B^2}{T^2}).$$

Thus the optimal convergence rate occurs when $B = \mathrm{O}(T^{1/2})$, and the rate is $\mathrm{O}\{T/\log T)^{-4/5}\}$.

References

Brillinger, D. (1981). *Time Series: Data Analysis and Theory.* Holden-Day, Oakland, California.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25**, 1-37.

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist.* **28**, 1762-1794.

Dai, M. and Guo, W. (2004). Multivariate spectral analysis using Cholesky decomposition. *Biometrika* **91**, 629-643.

Eaton, M. and Olkin, I. (1987). Best equivariant estimators of a Cholesky decomposition. *Ann. Statist.* **15**, 639-650.

Grenander, U. and Rosenblatt, M. (1984). *Statistical Analysis of Stationary Time series.* Chelsea, New York.

Gu, C. and Wahba, G. (1993). Semiparametric analysis of variance with tensor product thin plate splines. *J. Roy. Statist. Soc. Ser. B* **55**, 353-368.

Guo, W., Dai, M., Ombao, H., and von Sachs (2003). Smoothing spline anova for time-dependent spectral analysis. *J. Amer. Statist. Assoc.* **98**, 643-652.

Lin, Y. (2000). Tensor product space ANOVA models. *Ann. Statist.* **28**, 734 - 755.

Litt, B., Esteller, R., Echauz, J., D'Alessandro, M., Shor, R., Henry, T., Pennell, P., Epstein, C., Bakay, R., Dichter, M. and Vachtsevanos, G. (2001). Epileptic seizures may begin hours in advance of clinical onset: A report of five patients. *Neuron* **30**, 51-64.

Ombao, H., Raz, J., von Sachs, R. and Malow, B. (2001). Automatic statistical analysis of bivariate nonstationary time series. *J. Amer. Statist. Assoc.* **96**, 543-560.

Percival, D. B. (1994). Spectral analysis of univariate and bivariate time series. *Statistical Methods for Physical Science*, 313-348. Academic Press, New York.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proc. IEEE* **70**, 1055-1096.

von Sachs, R. and MacGibbon, B. (2000). Nonpametric spectral density estimation using penalized Whittle likelihood. *Scand. J. Statist.* **27**, 475-499.

Walden, A. T. (2000). A unified view of multitaper spectral estimation. *Biometrika* **87**, 767-788.

Walden, A. T. and McCoy, E. J. (1995). The effective bandwidth of a multitaper spectral estimator. *Biometrika* **82**, 201-214.

Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

E-mail: wguo@cceb.upenn.edu

Department of Mathematics & Statistics, University of North Carolina at Charlotte, 9201 University City BLVD Charlotte, NC 28223, U.S.A.

E-mail: mdai@email.uncc.edu

(Received June 2003; accepted July 2005)