Automatic Statistical Analysis of Bivariate Nonstationary Time Series

In memory of Jonathan A. Raz

Hernando C. Ombao, Jonathan A. Raz, Rainer von Sachs, and Beth A. Malow

We propose a new method for analyzing bivariate nonstationary time series. The proposed method is a statistical procedure that automatically segments the time series into approximately stationary blocks and selects the span to be used to obtain the smoothed estimates of the time-varying spectrum and coherence. It is based on the smooth localized complex exponential (SLEX) transform, which forms a library of orthogonal complex-valued transforms that are simultaneously localized in time and frequency. We show that the smoothed SLEX periodograms are consistent estimators, report simulation results, and apply the method to a two-channel electroencephalogram dataset recorded during an epileptic seizure.

KEY WORDS: Kernel Smoothing; Nonstationary time series; SLEX periodogram; SLEX transform; Time-frequency analysis; Time-varying spectrum and coherence.

1. INTRODUCTION

Electroencephalograms (EEG's) or brain wave patterns are brain electrical potentials that are recorded by placing electrodes on the scalps of subjects. Examples of EEG's recorded during an epileptic seizure by two electrodes, namely the left temporal lobe (T3 channel) and the left parietal lobe (P3 channel), are given in Figure 1. These EEG's were recorded from a female patient who was diagnosed with left temporal lobe epilepsy. Each EEG was recorded for a total of approximately 5 min and 28 s. The sampling rate was set at 100 Hz and the total number of time points per EEG is $T = 32,768$. The neurologist stated that seizure for this epileptic episode started at about $t = 185$ s. Note that the amplitude (or variability) of the EEG's changes over time. In particular, the amplitude of the EEG's during seizure is greater than the amplitude before or at the end of seizure. This suggests that the spectrum of the epileptic seizure EEG's changes over time and hence these EEG's should be modeled as nonstationary processes.

It is well known that as a patient progresses from the preictal to the postictal state, changes in the electrical activity of the brain are manifested in changes in the estimated spectral power. In addition to estimating the power spectra, neurologists are interested in the estimate of coherence between EEG's recorded at different channels. An estimate of coherence is important because it gives an indication of how a seizure spreads from a local area to other regions of the brain.

In this article, we propose a new method that automatically segments a bivariate time series into segments that are approximately stationary. Our method also automatically chooses a smoothing parameter that is used to give smoothed kernel estimates of the spectra of the two time series and the coherence between the two time series. We motivate our approach by first giving a review of the principles of the existing methods of analyzing nonstationary EEG time series.

One parametric method for analyzing a univariate EEG recorded during an epileptic seizure is given in Jansen, Hasman, and Lenten (JHL) (1981). The JHL method uses autoregressive models. For completeness, a random process $Y_t$ is said to be autoregressive of order $p$, denoted AR($p$), with parameters $\theta_1, \ldots, \theta_p$, if it is generated by the equation

$$Y_t = \theta_1 Y_{t-1} + \cdots + \theta_p Y_{t-p} + \xi_t,$$

where $\xi_t$ are independent random variables with mean 0 and variance $\sigma^2$. In addition, it is said to be a Gaussian AR($p$) if the $\xi_t$ are independent $N(0, \sigma^2)$ random variables. The JHL method estimates the time-varying spectrum by sequential steps: (1) adaptively segmenting the EEG into approximately stationary blocks, (2) fitting an AR model to these blocks, and (3) finally giving a parametric estimate of the spectrum at a stationary block to be $f(\omega) = \hat{\sigma}^2/(1 - \hat{\theta} \exp(-2\pi\omega) - \cdots - \hat{\theta}_p \exp(-2\pi p\omega))^2$, where $\omega \in [-1/2, 1/2]$. The JHL method, although conceptually easy to understand, is heavily computational and has several limitations. The choice of the length of the initial stationary segment as well as the order $p$ of the AR model is subjective. Moreover, it is not clear how this method can be extended to multivariate time series.

Kitagawa and Gersch (1996) proposed using time-varying parameter autoregressive (TVAR) models to analyze EEG's. This approach can be considered as a generalization of the JHL method, because the time-varying parameters are allowed to change abruptly in time. The spectral estimate at time $t$ can be obtained by plugging in the AR parameter estimates at that time. West, Prado, and Krystal (WPK) (1999) also used TVAR models in conjunction with time domain decomposition methods to compare the latent structure of EEG's recorded from depressed patients during induced brain seizures. The WPK method decomposes an EEG time series into the different
components that are assumed to represent high frequency activities, low frequency activities, and noise. One limitation of the WPK method is that it is only able to examine one time series at a time. Extensions to multivariate time series are not easy. In addition, the method is also heavily computational and needs an efficient algorithm to handle very long time series datasets. In WPK (1999), the computational burden is eased by subsampling a long time series. A consequence of subsampling is that information on high frequencies is lost. This can be problematic, especially when the main interest lies in the detection of onset of high frequency activity and the evolution of high frequency activity over time (Fisher 1992). Our neurologist, in particular, is interested in extracting the following information from the EEG dataset (Figure 1): (1) the band of frequency activity at the onset of seizure and (2) the evolution over time of the high frequency activity.

The preceding AR-based methods give parametric estimates of the spectrum. The frequency domain methods generally give nonparametric estimates. The frequency domain methods for analyzing EEG’s are built on the premise that EEG’s can be modeled approximately by piecewise stationary processes. Hence, the time-varying spectra of EEG’s can be approximated by the piecewise constant spectra of the stationary processes. The standard approach of estimating the time-varying spectrum is first to divide the time series into blocks of equal sizes and then compute the tapered Fourier periodograms at each of the blocks (Cohen 1989; Williams 1997). A drawback to this methodology is that the length of the blocks is fixed and is chosen subjectively. When the length of the block is small, it is often difficult to resolve the components of the signal that oscillate at frequencies that are nearly identical. A large block, on the other hand, permits good frequency resolution, but causes difficulty in determining where in time the various frequency components act.

Adak (1998) proposed a methodology that allows a data-dependent choice of the block (or window) lengths. The lengths are adapted to the rate of change of the spectrum; that is, larger windows are generated for longer periods of stationarity. Adak’s methodology divides the time series in a dyadic manner, up to some prespecified scale \( J \), into possibly overlapping blocks and then computes the windowed Fourier transform at each of the blocks. At a particular scale \( j \) (where \( j = 0, 1, \ldots, J \)), the time series is divided into \( 2^j \) blocks. Adjacent blocks at every scale are combined to form one (larger) stationary block if their estimated spectra are close according to some distance measure; otherwise, they are left as distinct blocks.

Donoho, Mallat, and von Sachs (DMvS) (1998), following the ideas of Mallat, Papanicolaou, and Zhang (MPZ) (1998), introduced a model of a locally stationary process and then developed a procedure for obtaining an adaptively smoothed and consistent estimator of the time-varying covariance under this model. Intuitively, a process is said to be locally stationary if an interval can be formed around each time point so that the process is approximately stationary at that interval. Dahlhaus (1997) gave a slightly different definition of this local stationarity, which we use as a model to validate the properties of our estimation procedure from the asymptotic point of view of consistency.

As in MPZ, the DMvS method uses the cosine packet transform (CPT) that forms a library of orthonormal transforms (collection of orthonormal cosine packet bases). Cosine packets are functions that form the building blocks of the CPT. They are obtained by applying a smooth and compactly supported window to the cosine vectors. Thus, they are localized in time and frequency. The DMvS method first computes the cosine packet coefficients, which are the inner products between the time series and the cosine packets. Next, the coefficients are arranged in a cosine packet table and squared to form the cosine packet periodograms. The time series is automatically segmented by applying the best basis algorithm (BBA) of Coifman and Wickerhauser (1992) on the cosine packet periodograms. Different variants of their algorithm, operating with different cost functions in the BBA, are suggested. One possibility is that at each of the blocks included in the best segmentation chosen by the BBA, wavelet-denoised estimators for the log spectrum are obtained and then exponentiated to form the spectral estimate. The inverse CPT is then applied on the spectral estimate to obtain the estimate of the time-varying covariance.

The primary limitation of the Adak method is that it is based on a nonorthogonal transform. As a consequence, establishing consistency of the method is not tractable (S. Adak, personal communication, 1997). Moreover, the Adak method is for univariate processes and does not lend itself to a natural extension to multivariate processes. Hence, it cannot give

Figure 1. Bivariate EEG Time Series at Channels T3 and P3. Number of time points \( T = 32,768 \); sampling rate \( = 100 \) H. Subject: Female patient diagnosed with left temporal lobe epilepsy. Dataset collected by the Department of Neurology, University of Michigan.
estimates of the time lag and coherence between components of a multivariate time series.

The main goal of the DMvS method is to give a consistent estimator of the time-varying covariance of the DMvS model of a locally stationary process. This approach does not directly address the need of neurologists to characterize changes in the electrical activity of the brain during an epileptic seizure. These changes can be described directly by the evolutionary spectrum of a nonstationary process. Another limitation of the DMvS method, as well as its predecessor MPZ, is that it uses the CPT, which is a real-valued transform. It does not have a phase and hence is not able to model lag between pairs of components in a multivariate time series. It is developed only for univariate processes and, therefore, is not able to give estimates of coherence.

In this article, we propose a methodology, which we call the Auto-SLEX method, that overcomes the limitations of the existing methods. The Auto-SLEX method is a statistical procedure that automatically segments the bivariate time series into approximately stationary blocks and automatically chooses the smoothing parameter for the estimation of time-varying spectra and coherence. The Auto-SLEX method is based on the smooth localized complex exponential (SLEX) transform, which forms a library of orthonormal transforms (collection of orthonormal SLEX bases). Orthonormal transforms are mathematically elegant and preserve the energy of the time series. Furthermore, they allow for the theory to be tractable (DMvS 1998). The SLEX transform uses SLEX basis vectors, which are closely related to the Fourier basis vectors and are obtained by applying two special windows to the Fourier vectors. The SLEX vectors are simultaneously localized in time and frequency, and, hence, are able to represent discrete random processes whose spectral properties change with time. In addition, the SLEX transform is complex valued. It has a phase that can model the time lag between the components of the bivariate time series. Moreover, the Auto-SLEX method can give estimates of coherence because it is designed for multivariate nonstationary time series. Although the Auto-SLEX procedure presented herein is for bivariate processes only, its extension to the multivariate setting is straightforward. Finally, the Auto-SLEX method yields results that are easy to interpret because they parallel existing spectral methods for analyzing stationary time series.

The rest of the article is organized as follows. In Section 2, we state the Dahlhaus model for locally stationary processes and then discuss apparent intuitive approximations by piecewise stationary and blended stationary processes. The SLEX transform is described in Section 3 and the Auto-SLEX method is discussed in Section 4. Theoretical results are presented in Section 5. Simulation results are given in Section 6. In Section 7 we present an application of the Auto-SLEX method to a speech signal. An application of the Auto-SLEX method to a bivariate EEG dataset recorded during an epileptic seizure is given in Section 8. In the conclusion section we discuss open questions and indicate how we intend to further develop our method to the direction of multichannel automatic spectral and coherence analysis. Proofs are deferred to the Appendix.

2. LOCALLY STATIONARY PROCESSES

In this section, we discuss the model of locally stationary processes given in Dahlhaus (1997). It is based on a time-dependent analog to the classical Cramér spectral representation of stationary processes. Our discussion is restricted to univariate zero-mean processes for ease of exposition. The multivariate extension (see also Dahlhaus, in press) is straightforward and parallels the extension of the classical spectral representation to the multivariate case (Brillinger 1981, section 4.6).

2.1 The Dahlhaus Model

Definition 1 (Dahlhaus). A sequence of zero-mean random variables \( \{X_{i,T}\}_{i=0}^{T-1} \), \( T \geq 1 \), is called locally stationary with transfer function \( A^0 \) if there exists a representation

\[
X_{i,T} = \int_{-1/2}^{1/2} A_{0,T}^0(\omega) \exp(i2\pi \omega t) \, dZ(\omega),
\]

where:

1. \( Z(\omega) \) is a zero-mean orthogonal increment process on \([-1/2, 1/2] \),
2. There exists a positive constant \( D \) and a smooth function \( A: [0, 1] \times [-1/2, 1/2] \to C \) with \( A(u, \omega) = A(u, -\omega) \), such that for all \( T \),

\[
\sup_{t, \omega} |A_{0,T}^0(\omega) - A(t/T, \omega)| \leq DT^{-1}.
\]

3. For all \( u, A(u, \omega) \) is continuous in \( u \).

Definition 2 (Dahlhaus). The time-varying or evolutionary spectrum of the Dahlhaus locally stationary process at time \( u \in [0, 1] \) and frequency \( \omega \in [-1/2, 1/2] \) is \( f_{XX}(u, \omega) = |A(u, \omega)|^2 \).

Remarks. Note that the first argument of \( A(t/T, \omega) \) is rescaled to live on the unit interval. Increasing the number of observations, \( T \), does not mean looking into the future. Rather, this asymptotic framework allows more data to be observed at a local structure and asymptotic inference starting from a single realization of \( \{X_{i,T}\} \). This is possible by Equation (3), which says that \( A_{0,T}^0(\omega) \approx A(t/T, \omega) \); that is, the smoothness of \( A \) in \( u \) controls the change of \( A_{0,T}^0 \) as a sequence in \( t \) such that it is allowed to change only slowly over time. The idea behind this is essentially that for each fixed \( T \), we implicitly assume some local interval of stationarity about each time point and a smooth change from one interval to the next.

In the literature (Dahlhaus 1996, 1997; Neumann and von Sachs 1997), a variety of possible smoothness conditions on \( A(u, \omega) \) exists. We state ours in the context of the theorems we intend to show. It is also interesting to note that the model in Adak (1998) uses a smoothness condition that is slightly different from Equation (3).

We have defined the time-varying spectrum \( f(u, \omega) \) of the model. This allows us to develop asymptotic estimation theory for \( f(u, \omega) \). The theory parallels nonparametric estimation of curves where questions of consistency, asymptotic normality, and related measures (such as \( L_2 \) risks) of the quality of an estimator of \( f(u, \omega) \) (e.g., a smoothed time-varying periodogram) are addressed.
The next two subsections discuss two classes of processes that can serve as intuitive simple approximations to locally stationary processes: the class of piecewise stationary processes and the class of blended stationary processes. Both, but particularly the second, processes are also useful to motivate the construction of estimators of the time-varying spectrum of a locally stationary process.

2.2 Piecewise Stationary Processes

A piecewise stationary processes can be defined (see, e.g., Adak 1998) to be

$$\tilde{X}_{t,T} = \sum_{b=1}^{B} X_b^T \mathcal{J}(t/T, I_b),$$  \hspace{1cm} (4)

where $X_b^T$ are stationary processes with spectra $f_b^T(\omega)$, and $I_b = [u_{b-1}, u_b)$, $\mathcal{J}(t/T, I_b) = 1$ if $t/T \in I_b$ and $\mathcal{J}(t/T, I_b) = 0$ if $t/T \notin I_b$. Time is rescaled to live on the interval $[0, 1]$ so that $0 = u_0 < \cdots < u_B = 1$ forms a partition of $[0, 1]$ that depends on $T$. It is not difficult to show that locally stationary processes can be well approximated by piecewise stationary processes. In fact this is just a variant of (the proof of) our more general Theorem 1 on the approximation of locally stationary processes by blended stationary processes. In general, the number of blocks $B$ needs to increase with the number of observations $T$ to make the approximation error tend to zero (see, again, Theorem 1). To assure that more observations fall within each stationary segment or block, we need to satisfy $B/T \to 0$ as $B \to \infty$ and $T \to \infty$. This is equivalent to saying that the upper bound, $M$, for the number of observations on the blocks fulfills $M \to \infty$, but $M/T \to 0$ as $T \to \infty$.

2.3 Time–Frequency plot

A useful tool for analyzing nonstationary time series is the time–frequency plot. A time–frequency plot is a representation of the time-varying spectrum. It gives the concentration of power at each frequency at each time point. An example of a time–frequency plot of a nonstationary process (defined later) is given in Figure 2. The time–frequency plot in Figure 2 is a representation of the time-varying spectrum of a piecewise stationary process defined as

$$X_t = \begin{cases} X_{1,t}, & \text{if } 0 \leq t < 511, \\ X_{2,t}, & \text{if } 512 \leq t < 1023, \end{cases}$$

where $X_{1,t}$ and $X_{2,t}$ are Gaussian AR(1) with parameters $\theta_1 = .9$ and $\theta_2 = -.9$, respectively (refer to Eq. (1)). In all time–frequency plots in this article, darker shades represent higher power (or stronger coherence). Power at the first half of the process is concentrated at lower frequencies. At the second half, power is shifted to higher frequencies.

An obvious limitation of piecewise stationary processes is that the spectral properties change abruptly at the change points and, hence, may not be reasonable models for processes that evolve slowly over time. Blended stationary processes, which are discussed in the next subsection, overcome these difficulties, because they allow some overlap between two adjoining stationary processes.

2.4 Blended Stationary Processes

A good model for biological processes that evolve slowly is one that has a spectrum with smooth transitions between adjoining stationary blocks. In this section, we define blended stationary processes as

$$\tilde{X}_{t,T}^B = \sum_{b=1}^{B} W_b ([t/T - u_{b-1}]/[u_b - u_{b-1}]) X_t^b,$$  \hspace{1cm} (5)

where $X_t^b$ is the stationary process for block $b$ with spectrum $f_b^T(\omega)$. The weights $W$ overlap and must satisfy

$$\sum_{b=1}^{B} W_b ([t/T - u_{b-1}]/[u_b - u_{b-1}]) = 1$$  \hspace{1cm} (6)

for any $B$ and for all $t$, where $B$ is the number of stationary blocks. Let $M$ be the upper bound on the number of observations on the blocks. A blended stationary process must satisfy the equivalent conditions

$$B \to \infty, \quad \text{but} \quad B/T \to 0,$$  \hspace{1cm} (7)

$$M \to \infty, \quad \text{but} \quad M/T \to 0$$  \hspace{1cm} (8)

as $T \to \infty$. We now state the result that the class of blended stationary processes (including piecewise stationary processes as a special case) can approximate locally stationary processes.

Theorem 1. Let $\{X_{t,T}\}$ be a locally stationary process with $A(u, \omega)$ being Lipschitz in $u$. Let the weights $W$ satisfy the condition in Equation (6). Moreover, let $B$ and $M$ satisfy (7) and (8), respectively. Then there exists a sequence of blended stationary processes $\{\tilde{X}_{t,T}^B\}$ and a positive constant $Q$ such that for all $T \geq 1$,

$$T^{-1} \sum_{t=0}^{T-1} \mathbb{E} (X_{t,T} - \tilde{X}_{t,T}^B)^2 \leq Q M^2 / T^2.$$  \hspace{1cm} (9)

The proof is given in the Appendix.

The preceding result justifies the standard approach of analyzing nonstationary time series, that is, segmenting the time series into approximately stationary blocks and then obtaining smoothed periodograms as estimates of the spectrum at each of the stationary blocks. In Theorem 1, the use of the weights allows a smooth transition of the spectral estimates from one stationary block to the next.
3. THE SLEX TRANSFORM

3.1 The SLEX Basis Vectors

The Fourier basis functions are perfectly localized in frequency and are periodic. Thus, they are ideal for representing a stationary time series, because the distributional properties of a stationary time series do not change with time. The Fourier basis functions, however, do not adequately represent processes that have spectral properties that evolve with time. To alleviate the time localization problem, smooth compactly supported windows can be applied to the Fourier functions (Daubechies 1992). A windowed Fourier function is of the form \( \varphi_{\omega}(u) = \Psi(u) \exp(i2\pi \omega u) \), where \( \Psi \) is a taper and \( \omega \in [-1/2, 1/2] \). The windowed Fourier functions, however, are generally no longer orthogonal. In fact, the Balian–Low theorem says that there exists no smooth window such that the windowed Fourier basis functions are simultaneously orthogonal, and localized in time and frequency (Wickerhauser 1994).

We next define SLEX basis functions and vectors, which are simultaneously orthogonal and localized in time and frequency. They evade the Balian–Low obstruction because they are constructed by applying a projection operator, rather than a window, on the Fourier functions. The details of the construction of a projection operator are given in Wickerhauser (1994). Ombao (1999) showed that the application of a projection operator is identical to applying two, rather than just a single, smooth and compactly supported windows (see Fig. 3) to the Fourier basis functions.

A SLEX basis function \( \phi_{\omega}(u) \) has the form

\[
\phi_{\omega}(u) = \Psi_+(u) \exp(i2\pi \omega u) + \Psi_-(u) \exp(-i2\pi \omega u),
\]

where \( \omega \in [-1/2, 1/2] \) and \( u \in [-\eta, 1+\eta] \), where \( 0 < \eta < .5 \). The windows \( \Psi_+ \) and \( \Psi_- \) come in pairs; that is, once \( \Psi_+ \) is specified, the other is known. Also note that the SLEX function [Eq. (10)] is a generalization of the windowed Fourier function.

SLEX basis functions generalize directly to orthogonal SLEX basis vectors for representing discrete time series. Define \( \alpha_1 > \alpha_0 \) to be two time points and let \( S = \{ \alpha_0, \alpha_0 + 1, \ldots, \alpha_1 - 1 \} \) be a block of time points. Define \( |S| = \alpha_1 - \alpha_0 \) and \( \varepsilon = \lceil \eta |S| \rceil \), where \( \lceil \cdot \rceil \) denotes the greatest integer less than or equal to its argument. A SLEX basis vector has elements \( \{ \phi_{S, \omega_j, t}, t = \alpha_0 - \varepsilon, \ldots, \alpha_0, \ldots, \alpha_1 - 1, \ldots, \alpha_1 - 1 + \varepsilon \} \), where

\[
\phi_{S, \omega_j, t} = \varphi_{\omega_j}(\frac{t - \alpha_0}{|S|})
\]

\[
= \Psi_+(\frac{t - \alpha_0}{|S|}) \exp\left\{ i2\pi \omega_j (t - \alpha_0) \right\}
\]

\[
+ \Psi_-(\frac{t - \alpha_0}{|S|}) \exp\left\{ -i2\pi \omega_j (t - \alpha_0) \right\},
\]

where \( \omega_j = k/|S| \), \( k = -|S|/2 + 1, \ldots, |S|/2 \). Different blocks are defined by different values of \( \alpha_0 \) and \( \alpha_1 \).

SLEX basis functions and vectors overlap. Thus they allow for smooth transitions between adjacent blocks. They are orthogonal, despite the overlap, because of the inclusion of the second smooth window \( \Psi_- \). They are localized in time because the windows \( \Psi_+ \) and \( \Psi_- \) have compact support. Finally, they are localized in frequency because the windows are smooth.

3.2 Computing the SLEX Transform

The first property of the SLEX transform is that it smoothly partitions the time axis by allowing the SLEX vectors to overlap. The SLEX transform divides the time series in a dyadic fashion. Suppose that each component of the time series \( (X_t, Y_t) (t = 0, \ldots, T - 1) \) has length \( T = 2^{T/2} \), where \( L = 11 \). (To simplify the discussion, we consider only the SLEX transform of one component of the bivariate time series.) At scale \( j = 0 \), the time series is kept undivided as one single block or segment. At scale \( j = 0, \ldots, J \), where \( J \) is some prespecified scale that satisfies \( J < L \), the time series is divided into \( 2^j \) overlapping segments. The segments correspond to the support of the SLEX basis vectors. We define \( S(j, b) \) to be the \( b \)th block on scale \( j \). The SLEX coefficients on block \( S(j, b) \) are defined as

\[
d_j^{(X)}(\omega_k) = \frac{1}{\sqrt{M_j}} \sum_t X_t \phi_{j, b, \omega_k, t},
\]

where \( M_j = |S(j, b)| = T/2^j \) and \( \phi_{j, b, \omega_k, t} = \phi_{S(j, b), \omega_k, t} \) is the SLEX basis vector on block \( S(j, b) \) oscillating at frequency \( \omega_k = k/M_j \), where \( k = -M_j/2 + 1, \ldots, M_j/2 \).

One important feature of the SLEX transform is that it is computationally efficient. The SLEX vectors are like tapered Fourier vectors. This property of the SLEX vectors allows the use of the fast Fourier transform (FFT) to compute the SLEX transform.

3.3 Best Segmentation

The SLEX transform forms a library of orthonormal transforms. After computing the SLEX transform, a well-defined cost [for example, Eq. (15)] is computed at each of the blocks. The cost for a particular segmentation of the time series is the sum of the costs at all the blocks that define that segmentation. The SLEX table (see Fig. 4) is an example of a SLEX library and the shaded blocks represent one orthonormal transform (or one segmentation or one basis). The cost for this particular transform is the sum of the costs at each of the shaded blocks. The BBA is then applied to the SLEX transform to obtain the unique orthonormal transform in the SLEX library.
that has the smallest cost. A detailed discussion of BBA and the cost functions is found in Wickerhauser (1994).

3.4 Remarks

The choice of the finest scale (or deepest level) of the transform should be governed by the following principles. As increases (while is fixed), the number of observations on the finest scale decreases. Although this is favorable from the point of view that bias due to nonstationarity can be controlled, fewer observations in a block mean reduced frequency resolution and increased variance of the spectral estimates.

When computing the first block at any given level , the block is padded with zeroes at the beginning to match the length of the SLEX vectors defined at that block. Similarly, the last blocks at every level are padded with zeroes at the end. Padding with zeroes does not affect the orthogonality of the SLEX vector and is a common practice in spectral analysis of time series (Shumway and Stoffer 2000).

The SLEX basis functions at different blocks overlap. Thus, the SLEX coefficients (and hence SLEX periodograms) in a particular block are computed using observations on that block and those at adjacent blocks. This implies that the SLEX transform already has a built-in time smoother to reduce the effect of blocky segmentation. As a consequence, the spectral values on either side of the boundary (or adjoining blocks) do not differ too drastically from each other.

The amount of overlap, , at the different scales need not be the same. To ensure orthogonality of the SLEX basis vectors, it is sufficient that the amount of overlap at scale satisfy .

DMvS motivates the use of BBA as follows: Desirable theoretical results for stationary random processes, such as consistency of the spectral estimators, are obtained in part because the Fourier transform approximately diagonalizes the covariance matrix of a stationary process. This idea can be transferred to locally stationary processes. It turns out that the transform, among the orthonormal transforms in a library, that minimizes the cost is the “best” diagonalizer of the covariance of a locally stationary process. Please refer to DMvS (sections 2, 10–12, and 14) for details.

3.5 The SLEX Transform and Other Localized Transforms

The recent literature on digital signal processing is permeated with the development of transforms that are localized in time and frequency. The wavelet transform is one popular transform, and now is widely discussed and used in nonparametric estimation. The wavelet transform is real valued and gives coefficients for a time series of length . The coefficients are obtained by computing the inner product between the time series and the wavelet basis vectors. Because they are localized in time and scale, wavelets are ideal for estimating functions that have sudden bursts and points of discontinuities, and those that have peaks at localized regions.

Another localized real-valued transform is the wavelet packet transform (WPT). The WPT divides the frequency axis of the two-dimensional time–frequency plane in a dyadic manner. The WPT consists of coefficients, where is the length of the time series and is the user-specified finest level of the transform. The WPT forms a library of orthonormal transforms. The best orthonormal transform (or best basis) can be selected using the BBA.

The CPT is yet another time and frequency localized transform. The CPT shares a common feature with the SLEX transform: both are based on trigonometric functions and both divide the time axis (rather than the frequency axis as in WPT) in a dyadic manner. Another similarity between the CPT and the SLEX transform is that both transforms are based on functions that are obtained by applying a projector. Cosine packets are obtained by applying a projector on the cosine function, whereas the SLEX functions are obtained by a projector on the complex exponential function.

The WPT and CPT are real valued. None of them has a phase and hence are not able to model the time lag between two components of a bivariate time series. The SLEX transform, on the other hand, is complex valued and hence is able to estimate the lag between components of a bivariate time series. Moreover, the SLEX transform uses the SLEX functions, which are a time- and frequency-localized generalization of the Fourier functions. Thus, our method totally parallels classical spectral analysis of stationary processes that is based on the Fourier functions. Our method is appealing to engineers and neurologists who are familiar with the Fourier-based spectral methods.

In addition, in Ombao, Raz, von Sachs, and Guo (2000) we proposed a model of a nonstationary random process that is based on the SLEX transform. Our model has as representation that is a time-dependent analog of the Cramér spectral representation. This model, again, is in the spirit of traditional
spectral analysis because it gives a spectrum whose physical meaning is a decomposition of power over frequency and time, and hence is consistent with the classical spectrum for stationary processes.

4. THE AUTO-SLEX METHOD

The Auto-SLEX method is an automatic statistical procedure that simultaneously segments the time series and gives the kernel-smoothed time-varying estimate of the spectra and coherence. Given a bivariate nonstationary time series \((X_t, Y_t)\), we aim to estimate the time-varying spectra \(f_{XX}\) and \(f_{YY}\) of the univariate components, the cross spectrum \(f_{XY} = f_{YY}\), the phase spectrum \(\text{arg}(f_{XY})\), and the coherence \(R_{XY} = |f_{XY}|/\sqrt{f_{XX}f_{YY}}\).

The method consists of two sequential steps: (1) automatic selection of the smoothing parameter and (2) automatic simultaneous segmentation and estimation. At each block of the SLEX transform, the best smoothing span is selected and then used to smooth (over frequency) the SLEX periodograms, and the real and imaginary parts of the SLEX cross-periodograms. The time series is then segmented into approximately stationary blocks by applying the BBA on the kernel-smoothed SLEX periodograms. The estimate of the time-varying spectra and coherence are the smoothed SLEX periodograms and SLEX coherence estimates at the blocks that define the best segmentation selected by BBA.

4.1 Automatic Selection of the Smoothing Parameter

The main outputs of this step are the kernel-smoothed SLEX periodograms and cross-periodograms whose bandwidths are automatically selected by generalized cross-validation (GCV). This step comes before finding the best segmentation of the nonstationary time series. The SLEX periodograms and cross-periodograms at each block of the SLEX transform are smoother prior to applying the BBA. The best smoothing span is selected at each block using a GCV-based procedure. We briefly describe the procedure herein. The details of the GCV method are reported in Ombao, Raz, Strawderman, and von Sachs (1999).

We first motivate the use of GCV in span selection. Note that the SLEX functions can be made arbitrarily close to the tapered Fourier functions [Eq. (10)] by adjusting the projection operator used in the construction of the tapers \(\Psi_x\) and \(\Psi_y\). Thus, the SLEX periodograms behave asymptotically like tapered Fourier periodograms. In Brillinger (1981), under mixing conditions and mild conditions on the taper, the tapered Fourier periodograms at frequency \(\omega = (0, 1/2]\) computed from a stationary block are approximately independently distributed gamma random variables with parameters \(\alpha = 1\) and \(\beta = f(\omega)\), where \(f(\omega)\) is the value of the spectrum at \(\omega\). We follow the parameterization of the gamma density in Lehmann and Casella (1998, p. 25). Thus, the problem of spectral estimation using SLEX periodograms can be considered in terms of a generalized additive model (GAM) with a gamma distribution that has a mean that depends on the unknown spectrum. Hastie and Tibshirani (1990) suggested choosing the smoothing parameter in GAM’s by minimizing a form of the GCV function. In Ombao et al. (1999), we proposed using the GCV function that is based on the deviance for gamma distributed random variables.

We now set the notation. Let \(d_{j,b}^{(X)}(\omega_k)\) and \(d_{j,b}^{(Y)}(\omega_k)\) \((k = -M_j/2 + 1, \ldots, M_j/2)\) be the SLEX coefficients on block \(S(j, b)\). The SLEX raw periodograms and cross-periodograms are defined as \(I_{j,b}^{(X)}(\omega_k) = |d_{j,b}^{(X)}(\omega_k)|^2, I_{j,b}^{(Y)}(\omega_k) = |d_{j,b}^{(Y)}(\omega_k)|^2\), and \(I_{j,b}^{(XY)}(\omega_k) = d_{j,b}^{(X)}(\omega_k)d_{j,b}^{(Y)}(\omega_k)\), respectively. Let \(\tilde{I}_{j,b}^{(X)}\) and \(\tilde{I}_{j,b}^{(Y)}\) be the SLEX periodograms that are smoothed using bandwidth \(v\). We denote the “error degrees of freedom” by \(df_{j,b,v}\) \(-1 - \text{trace}(H_v)/(M_j/2 + 1)\), where \(H_v\) is the smoother matrix.

The GCV functions for the components \(X_t\) and \(Y_t\) are, respectively,

\[
\text{GCV}^{(XX)}_{j,b}(v) = \frac{2}{(df_{j,b,v})^2} \sum_{k=0}^{M_j/2} \left(\frac{I_{j,b}^{(X)}(\omega_k)}{I_{j,b,v}^{(X)}(\omega_k)} - \log I_{j,b,v}^{(X)}(\omega_k) - 1\right). \tag{12}
\]

\[
\text{GCV}^{(YY)}_{j,b}(v) = \frac{2}{(df_{j,b,v})^2} \sum_{k=0}^{M_j/2} \left(\frac{I_{j,b}^{(Y)}(\omega_k)}{I_{j,b,v}^{(Y)}(\omega_k)} - \log I_{j,b,v}^{(Y)}(\omega_k) - 1\right). \tag{13}
\]

We define the best bandwidth for the block \(S(j, b)\) to be that which minimizes the function

\[
\mathcal{P}_{j,b}(v) = \text{GCV}^{(XX)}_{j,b}(v) + \text{GCV}^{(YY)}_{j,b}(v). \tag{14}
\]

The SLEX periodograms and the real and imaginary parts of the SLEX cross periodograms at each block in every scale are smoothed using the optimal bandwidth. We denote these by \(\tilde{I}_{j,b}^{(X)}\), \(\tilde{I}_{j,b}^{(Y)}\), and \(\tilde{I}_{j,b}^{(XY)}\), respectively. Remarks. To ensure that the estimate of the spectral density matrix is nonnegative, the kernel smoothers for the SLEX periodograms and crossperiodograms within each block must use the same bandwidth. Hence, we use a single minimizing function in Equation (14) that combines the GCV for the two components.

Another automatic span selection method is given in Lee (1997). The Lee span selector is based on unbiased estimation of the mean squared error between the smoothed periodogram estimate and the unknown spectrum. One problem with the Lee method is that it assumes that the periodograms at frequencies \(\omega_k, k = -M_j/2 + 1, \ldots, M_j/2\), are independent. This assumption is not correct (Brillinger 1981, theorem 5.2.6). Consequently, the span selector in Lee is now biased for the mean squared error. A bias-corrected version of the Lee span selector is given in Ombao et al. (1999). Moreover, we report that the GCV span selection method gives more stable and less variable spectral estimates than the Lee method.

The GCV method is also based on the full (asymptotic) likelihood because it uses the gamma deviance [Eq. (12)], that is, the deviance that is appropriate for the asymptotic distribution of the periodograms. We point out that the Lee span selector, on the other hand, uses the squared error deviance, which is more appropriate in an additive model with homoscedastic normal errors.
4.2 Automatic Segmentation and Estimation

The next step is to apply the BBA to the smoothed SLEX periodograms. In our simulations and applications, we used the log-energy cost plus a complexity penalty term as motivated in DMvS. The cost at each block $S(j, b)$ is defined as

$$\text{Cost}(j, b) = \left\{ \sum_{k=-M_j/2+1}^{M_j/2} \log \tilde{T}^{(XX)}_{j,b}(\omega_k) + \log \tilde{T}^{(YY)}_{j,b}(\omega_k) \right\} + \beta \sqrt{M_j},$$

where $\beta$ is the complexity penalty parameter. The cost for one segmentation is the sum of the costs for all blocks that define that segmentation. The BBA then searches for the segmentation with the smallest cost. The estimates of the time-varying spectra $\tilde{T}^{(XX)}_{j,b}$ and $\tilde{T}^{(YY)}_{j,b}$ and of the cross spectra $\tilde{T}^{(XY)}_{j,b}$ are extracted from blocks $S(j, b)$ that define the best segmentation selected by BBA. The estimate of the time-varying coherence is given by

$$\tilde{R}^{(XY)}_{j,b}(\omega_k) = \left| \tilde{T}^{(XY)}_{j,b}(\omega_k) / \sqrt{\tilde{T}^{(XX)}_{j,b}(\omega_k) \tilde{T}^{(YY)}_{j,b}(\omega_k)} \right| \tilde{R}.$$  

Remarks. To make meaningful coherence estimates, every component of the multivariate time series must have the same segmentation. In the procedure, we combine the cost of each component of the bivariate time series to obtain a single segmentation.

The cost function [Eq. (15)] includes the term $\beta \sqrt{M_j}$. This penalty term safeguards the procedure from obtaining a segmentation that has too many or too few blocks. A small value of $\beta$ leads to a procedure that tends to select a segmentation with too many small blocks. As previously discussed, this is favorable from the point of view that there is less bias due to nonstationarity. However, having less observations within each block leads to inflated variances of the estimates. A large value of $\beta$, on the other hand, leads to a procedure that tends to select a segmentation with very few blocks. Although variance of the estimates is reduced, having too few blocks may lead to bias due to nonstationarity (i.e., error due to not splitting a nonstationary block). The penalty parameter $\beta$ can be either approximated or computed via a data-driven procedure. In our work, we set $\beta = 1$, which is motivated by the thresholds set by Gao (1997) and supported by a theoretical argument in DMvS. Convincing empirical results were obtained using $\beta = 1$. We add that the sensitivity of the method to the choice of this parameter certainly will need further investigation.

The cost function [Eq. (15)] does not account for the cross-correlation structure between the multivariate time series. Our future work will involve fine tuning the Auto-SLEX method and this includes the cost function. The current form of the cost function, however, works well for cases where the rate at which the cross spectrum between any two time series changes over time is not faster than the spectra. Our neurologist believes that this is a reasonable assumption for epileptic seizure EEG’s, although further experiments are needed to confirm this assumption.

It is important to note that the cost function does not treat one bivariate process as two independent univariate series. Actually the two components are treated as one set of information and this one set of information yields one (“a common”) basis that can best represent the two components.

There are other cost functions that can be used. DMvS used the squared cost function, that is, the negative of the squared periodograms rather than the log periodograms. In numerical experiments, the squared cost function tends to divide stationary blocks unnecessarily.

4.3 Relationship of the Auto-SLEX Method to the Adak and DMvS Methods

A common thread that ties the three methods is the use of transforms that divide the time series in a dyadic manner. The transforms used, however, are different. Adak used the nonorthogonal windowed Fourier transform and DMvS used the real-valued cosine packet transform. The Auto-SLEX method uses the SLEX transform that is complex valued and forms a library of localized orthonormal transforms.

Adak (1998) suggested using two types of cost functions: (1) a distance measure between spectral estimates and (2) a distance measure plus a complexity penalty term. Adak’s method 2 is supposed to safeguard against choosing a segmentation with too many blocks. It incorporates a data-driven procedure for selecting the best penalty parameter. The Auto-SLEX method also uses a cost function that includes a complexity penalty for too many blocks. Simulation results in Section 6 show that the present implementation of the Auto-SLEX method, where the complexity penalty parameter is set to $\beta = 1$, works very well and performs better than the Adak method.

There is also a difference in the underlying principles for segmentation between the two methods. The Adak method searches for the best segmentation based on some well-defined criterion, that is, the cost computed for a parent block is a distance measure between the spectra of children blocks. The Adak method is not designed to select a basis. The Auto-SLEX method, on the other hand, uses the BBA, which searches for the basis that best represents the time series based on some well-defined cost function. The segmentation of the time series is implied by the best basis selected in the Auto-SLEX method.

DMvS showed that consistent estimators for the time-varying covariances can be obtained by applying BBA using the squared cost function (the cost for a block is the sum of the squared periodograms in that block). DMvS also suggested and provided theoretical justification for using the log-energy cost with complexity penalty. In addition to the theoretical motivation in DMvS, we implemented the log-energy cost because empirical results indicate that it is able to give a higher rate of correct segmentation than the squared cost.

Both the Adak method and the DMvS method (with log energy as a cost function) apply a segmentation procedure using raw periodograms. The Auto-SLEX method is more general because it allows the BBA to be applied to “smoothed” SLEX periodograms with span selected by an automatic procedure. This includes the raw SLEX periodograms as a special case that has a zero smoothing parameter. There is also empirical evidence that suggests that BBA is able to give the correct segmentation at a higher rate when applied to smoothed, rather than raw, SLEX periodograms.
Finally, DMvsS uses wavelet denoising, whereas Auto-SLEX uses kernel smoothing. It is well known that wavelet-denoised estimators are able to preserve important features of the data such as peaks and jumps. In our approach, we chose to use linear smoothers because when the same span is used in smoothing the periodograms and the cross periodograms, the estimate of the spectral density matrix is guaranteed to be nonnegative definite. This is justified by the Cauchy–Schwarz inequality.

5. ASYMPTOTIC RESULTS

In this section we prove consistency of the smoothed (univariate) SLEX periodogram as an estimator of the evolutionary spectrum \( f_{XX}(u, \omega) \) defined in the Dahlhaus model of local stationarity given in Section 2. We are careful to note here that the proof of consistency assumes that a segmentation of the nonstationary time series is already known. It does not address the overall consistency of the method, because it does not take into account the segmentation that is selected by BBA. This is a main focal point of our work in progress. Nevertheless, the theorem shows two main important results, that is, the effects of both the dyadic structure of the SLEX transform and the bell \( \Psi_\tau \) on the spectral estimator vanish asymptotically.

To prove consistency of our SLEX spectral estimators, we modify existing theoretical results that involve tapered Fourier periodograms, which are computed using just one smooth window (Dahlhaus 1997). Let us now set some necessary notation. Suppose that the univariate time series has length \( T = 2^L \) for some positive integer \( L \). Let \( J = J_T \) be the finest scale on which we consider the blocks \( S(J, b) \) as defined in Section 3.2. The blocks on scale \( J \) have length \( M = M_J = T/2^J = 2^{L-J} \). For our asymptotics, we assume that \( M/T \to 0 \) as \( T \to \infty \) and \( M \to \infty \).

Define a segment of observations centered at \( [uT] \), called \( S_{u,T} = [uT]-M/2+1, [uT]+M/2 \). In Dahlhaus (1996), the windowed Fourier periodograms were computed on \( S_{u,T} \). In our approach, we compute the SLEX periodograms \( I_k(u, \omega_k) \) for frequencies \( k \) using observations on the same segment \( S_{u,T} \).

As in the classical (Fourier) periodogram, we can obtain consistent estimators by smoothing the SLEX periodograms across frequency in each segment. We use a conventional kernel smoother,

\[
\hat{f}(u, \omega) = (Mb_j)^{-1} \sum_{k=-M/2+1}^{M/2} K_f((\omega - \lambda_k)/b_j)I_k(u, \lambda_k),
\]

(17)

where \( \lambda_k = k/M \), \( K_f(v) \) is the kernel function with compact support \([-1/2, 1/2], K_f(v) = K_f(-v), \), and \( \int K_f(v)dv = 1 \). Its bandwidth \( b_j \) is assumed to fulfill \( b_j \to 0 \) and \( Mb_j \to \infty \) as \( T \to \infty \).

It is interesting to observe that in the resulting kernel estimator in Equation (17) we automatically include some sort of “smoothing” in rescaled time. We observe that in rescaled time, \( u \), the shrinking length of the segment \( |S_{u,T}|/T \) is proportional to \( b_j = M_J/T \), which thus plays the role of a “bandwidth” in time.

**Theorem 2 (A consistency theorem).** Suppose \( \{X_{j,T}\} \) is a locally stationary process as defined in Theorem 1, with mean zero and transfer function \( A \) whose partial and mixed derivatives up to second order are continuous. Then, for \( k \neq 0, \omega \neq 0, \)

1. \( \mathbb{E} \{I_k(u, \omega_k)\} = f(u, \omega_k) + C_{\Psi,\lambda} b_j \theta + o(b_j^2) + O(\log(M_J)/M_J), \)
2. \( \mathbb{E} \{\hat{f}(u, \omega)\} = f(u, \omega) + C_{\Psi,\psi} b_j \theta^2 + C_{\lambda,\lambda} b_j \theta + o(b_j^2 + \theta^2 + \log(M_J)/M_J), \)
3. \( \text{var} \{\hat{f}(u, \omega)\} = (b_j b_j T)^{-1} f^2(u, \omega) \bar{C}_{\Psi,\lambda}^{-1} o((b_j b_j T)^{-1}), \)

where \( C_{\Psi,\lambda} \) denotes a constant that depends both on the window functions \( \Psi_\tau \) and \( \Psi_\lambda \) and on the second derivative \( f''_w \) of \( f(u, \omega_k) \) with respect to time \( u; \) \( C_{\lambda,\lambda} \) is a constant that depends on the kernel function \( K_f \) and \( \theta^2 \) is a constant that depends on the size of the windows \( \Psi_\lambda \) and \( \Psi_\tau \) and on the kernel function \( K_f \).

The proof is given in the Appendix.

**Remarks.** In Theorem 2 the smoothness conditions on the evolutionary spectrum [i.e., on the transfer function \( A(u, \omega) \)] can be substantially relaxed without losing the property of consistency (cf. Neumann and von Sachs 1997).

The rate of consistency is \( (b_j b_j T)^{-1} = (b_j M_J)^{-1} \), which is similar to the rate for smoothing a periodogram over \( M_J \) frequencies in a neighborhood of bandwidth \( b_j \).

As in Dahlhaus (1996), optimal rates for \( b_j \) (i.e., for \( M_J \)) and \( b_j \) can be calculated as rates that minimize the asymptotically leading term of the mean relative squared error \( \mathbb{E} \{ f(u, \omega)/\hat{f}(u, \omega) - 1 \}^2 \). Without going into details, following the usual technique of balancing the leading parts of squared bias and variance, we arrive at \( b_j \sim T^{-1/6} \) and \( b_j \sim T^{-1/6} \).

The kernel estimator [Eq. (17)] has the same structure as a segmented periodogram-based statistic in the central limit theorem of Dahlhaus (1997, theorem A.2). Hence with the same techniques applied on the higher cumulants of \( \hat{f}(u, \omega) \), asymptotic normality of our kernel estimator can be shown. In fact, by a consequence of a central limit theorem, even for non-Gaussian time series \( \{X_{j,T}\} \), the SLEX coefficients \( \{d(X)(o, j)\} \) are asymptotically normal.

The simulation studies reported in Sections 6.1 and 6.3 support the theoretical results that the effect of both the inclusion of the bell \( \Psi_\tau \) and the dyadic structure of the transform can be neglected for a sufficiently large sample size.

We have proven consistency given a fixed segmentation. A more general question arises concerning whether our method consistently estimates the spectrum when the segmentation is selected by BBA. Compare the approximation results of Section 2, in particular Theorem 1. We expect that our use of an orthogonal transform will be important in developing this more general theory. DMvsS developed a related theory based on the orthogonal local cosine transform. Although we have not yet developed more general theory, this argument provides additional support for our use of an orthogonal transform.

6. SIMULATIONS

To evaluate the proposed method, we performed simulation studies to compare it with the Adak method. We chose the
Adak method because it is closest to ours in terms of the goals of segmentation and estimation, use of a dyadic segmentation, and use of a complex-valued transform.

We performed simulation studies that (1) compared the kernel-smoothed SLEX periodogram spectral estimate with the kernel-smoothed tapered Fourier periodogram estimate of the spectrum of a stationary process, (2) compared the Auto-SLEX method with the Adak method for segmenting a piecewise stationary process, (3) compared the Auto-SLEX method with the Adak method for estimating the spectrum of a piecewise stationary process, (4) evaluated the performance of the Auto-SLEX method for estimating the time-varying spectrum of a piecewise stationary process when the change point is not at a dyadic break, and (5) evaluated the performance of the Auto-SLEX method for estimating the time-varying spectrum of a process that has a spectrum that changes very slowly over time.

6.1 SLEX Periodograms Versus Fourier Periodograms

It is already known that the Fourier complex exponentials best represent stationary processes. The goal of this simulation study is to show that stationary SLEX periodograms (periodograms computed using SLEX functions defined at level $J = 0$ only) can do reasonably well when used to estimate the spectrum of a stationary process. Moreover, we wanted to show that the difference between a tapered Fourier periodogram and the stationary SLEX periodogram is negligible when the length of the time series is sufficiently large.

We simulated $N = 100$ time series datasets, each of length $T = 1,024$, from the Gaussian AR(2) $(\theta_1 = 1.07, \theta_2 = -0.81)$ process [Eq. (1)]. For each simulated dataset, we computed the stationary SLEX transform and then smoothed the stationary SLEX periodograms using the Daniell window with a span that was automatically selected by the GCV method. We also smoothed the tapered Fourier periodogram using the Daniell window and with a span that was given by the GCV method. For each realization, the averaged squared error (ASE), with respect to the log spectrum, of each of the stationary SLEX estimates and the tapered Fourier estimates was computed,

$$
\text{ASE} = (T/2 + 1)^{-1} \sum_{k=0}^{T/2} \left\{ \log \hat{f}(\omega_k) - \log f(\omega_k) \right\}^2,
$$

(18)

where $\hat{f}(\omega_k)$ is the estimate (stationary SLEX or tapered Fourier) of the true spectrum $f(\omega_k)$. The mean and the quartiles of the ASE of the stationary SLEX and tapered Fourier estimate are given in Table 1.

A simulated dataset from the foregoing stationary process is given in Figure 5. The stationary SLEX and the tapered Fourier estimates based on this data set have ASE values (.0398 and .0380, respectively) that are both the median values of their corresponding ASE’s. The log of the spectral estimates using the stationary SLEX and tapered Fourier transforms, computed from this single realization is given in Figure 6.

We give the details of the implementation of the Auto-SLEX method. In the simulation study, we (1) used the log-energy plus complexity term as a cost function [Eq. (15)] in

![Figure 5. A Simulated Dataset of Length $T = 1,024$ From the Gaussian AR(2)(1.07, -0.81). The ASE values for the stationary SLEX periodogram estimate and the windowed Fourier periodogram estimate are both the median of their respective ASE values.](image)

![Figure 6. Log of the Smoothed Auto-SLEX Estimates and the Smoothed Tapered Fourier Estimates, Based on the Realization of Figure 5, of the Spectrum of a Gaussian AR(2)(1.07, -0.81).](image)

Table 1. Quartiles and Mean of the Respective ASE Values

<table>
<thead>
<tr>
<th></th>
<th>Quartile 1</th>
<th>Median</th>
<th>Mean</th>
<th>Quartile 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLEX</td>
<td>0.0315</td>
<td>0.0398</td>
<td>0.0397</td>
<td>0.0401</td>
</tr>
<tr>
<td>Fourier</td>
<td>0.0312</td>
<td>0.0380</td>
<td>0.0390</td>
<td>0.0432</td>
</tr>
</tbody>
</table>

NOTE: $N = 100$ datasets were simulated from the Gaussian AR(2) process $(\theta_1 = 1.07, \theta_2 = -0.81)$ in Section 6.1. For each dataset, an estimate using the stationary SLEX transform was computed as well as an estimate using the tapered Fourier transform. An ASE value is also computed for each of the stationary SLEX estimates and tapered Fourier estimates.
the search for the best segmentation; (2) computed the cost for all blocks down to scale \( J = 4 \); (3) used the Daniell window for smoothing; (4) used the GCV method to obtain bandwidths to be used for smoothing the SLEX periodograms.

We compared these simulations with the same setup for the Adak method where we (1) used the Kolmogorov–Smirnov distance measure without complexity penalty in the search for the best segmentation; (2) computed the cost for all blocks down to scale \( J = 4 \); (3) used the bell as the taper; (4) used the Daniell window for smoothing; (5) used the GCV methods to obtain bandwidths for smoothing. We followed the specification in Adak (1998) by selecting the best segmentation based on raw periodograms. In the simulation, we “enhanced” the Adak method by using the GCV span selector to give smoothed (and, hence, improved) estimates of the spectrum.

In preliminary numerical experiments using the piecewise processes in Adak (1998), we implemented the Adak procedure with the complexity penalty but noted that the segmentation results were not different from the procedure without complexity penalty. Furthermore, we were not able to replicate the results in Adak (1998) exactly from the information given in the article.

In evaluating the methods with respect to criterion 1, we performed two simulation studies that determined how well the Auto-SLEX method gave the correct segmentation of the time series and then compared the performance with the Adak method. We evaluated criterion 2 by computing the ASE of the log of the time-dependent spectrum over frequency in \([0, 1/2]\) as

\[
ASE = \left[ \frac{1}{T(M_j/2 + 1)} \sum_{t=0}^{T-M_j/2} \left( \log \hat{f}(t/T, \omega_k) \right)^2 \right]^{1/2},
\]

where \(\hat{f}(u, \omega_k)\) is an estimate of the true time-dependent spectrum \(f(u, \omega_k)\).

We obtained \(N = 200\) simulated datasets, each of length \(T = 1,024\), from a piecewise stationary process \(Y(t = 0, \ldots, 1,023)\), with change points at \(u = 2/4\) and \(3/4\), where

\[
Y_t = \begin{cases} 
X_{t,1}, & \text{if } 0 \leq t \leq 511, \\
X_{t,2}, & \text{if } 512 \leq t \leq 767, \\
X_{t,3}, & \text{if } 768 \leq t \leq 1,023,
\end{cases}
\]

where \(\{X_{t,1}\}\) is Gaussian AR(1) with parameters \(\theta_1 = 0.9\), \(\{X_{t,2}\}\) is Gaussian AR(2) with parameters \(\theta_1 = 1.69\) and \(\theta_2 = -0.81\), and \(\{X_{t,3}\}\) is Gaussian AR(2) with parameters \(\theta_1 = 1.32\) and \(\theta_2 = -0.81\). The same Gaussian AR(2) processes were used in the simulation in Adak (1998).

Table 2 compares the performance of the Auto-SLEX and Adak methods for the piecewise stationary case. It gives the percentage of simulated time series datasets (out of \(N = 200\)) that were segmented according to different segmentations. For example, the Auto-SLEX method correctly segmented 60.0% of the \(N = 200\) simulated time series datasets.

Furthermore, we plotted one simulated time series dataset in Figure 7. The ASE of the Auto-SLEX and Adak estimates based on this dataset are both close to their respective median ASE. In particular, the Adak ASE for this particular dataset is \(0.04336\) and corresponds to the 47th percentile of Adak ASE values. The Auto-SLEX ASE for this dataset is \(0.04236\) and corresponds to the median Auto-SLEX ASE values. Thus, the log-spectral estimates given in Figure 8, are valid representative estimates of the two methods.

**Discussion of Results.** The Auto-SLEX method was able to segment the time series correctly in 60.0% of the cases. In the other cases it happened to divide a block even if it was stationary. This error, however, is more tolerable than not splitting, because “undersplitting” can introduce nonstationary bias. With “oversplitting,” there is a chance to recover from the error if the spectral estimates at the splits of a stationary block are close. However, an undesirable effect of unnecessary oversplitting is that the variance of the spectral estimates becomes inflated due to fewer observations in a stationary block. In view of this, the Auto-SLEX method did remarkably well, because only in 6% of the cases did it decided for more than one additional accidental split compared to Adak’s 22%.
Figure 7. A Simulated Dataset From the Piecewise Stationary Process in Equation (20). The Auto-SLEX estimate on Figure 8, based on this simulated dataset, has ASE that is the median value of the $N = 200$ Auto-SLEX ASE’s. The Adak estimate has ASE that is the 47th percentile of the Adak ASE’s.

A comparison of the two methods shows that the Auto-SLEX method gave a higher rate of correct segmentation and a lower ASE for the correct segmentation than the Adak method. Moreover, the Adak method tended to oversplit more often than the Auto-SLEX method, in particular, with a much higher percentage for more than one additional false split.

6.3 Nondyadic segmentation

In practice it is somewhat unlikely that the segmentation of a piecewise stationary process follows a dyadic structure. An example of a process that does not have a dyadic structure is

$$X_t = \begin{cases} X_{1,t}, & \text{if } 0 \leq t \leq 196, \\ X_{2,t}, & \text{if } 197 \leq t \leq 1,023, \end{cases}$$

(21)

where $X_{1,t}$ and $X_{2,t}$ are AR(1) processes with parameters $\theta_1 = .9$ and $\theta_2 = -.9$, respectively.

An “ideal” segmentation of the time series would be that which gives very fine blocks around the change point. The exact dyadic change points that are used to approximate the nondyadic change point depend on the length of the time series and the prespecified level $J$. (Larger values of $J$ allow for finer segmentation.) In our setup ($T = 1,024, J = 4$), we can consider splits at $t = 2/16, 3/16, 4/16, 8/16$ to be the “ideal” segmentation.

We simulated $N = 200$ datasets from the preceding process, and obtained the Auto-SLEX and Adak estimates of the log spectrum. The ASE for each dataset also was computed [see Eq. (19)]. The plot of a simulated dataset is shown in Figure 9. The Auto-SLEX and the Adak estimates of the log spectrum computed from this dataset are shown Figure 10. The ASE’s of the Auto-SLEX and Adak estimates are .18833 and .27839, respectively, and correspond to the 47th and 43rd percentiles of their respective ASE values.

Discussion of Results and Dyadic Segmentations. We noted in Table 3 that the Auto-SLEX method was able to select the “ideal” segmentation more often than the Adak method (44% versus 18%). The Adak method suffers from being unable to detect the presence of a change. Note that the Adak method undersplit in 19% of the cases. This is a serious problem and can lead to biased estimates of the spectrum. The Auto-SLEX method, on the other hand was sensitive to actual changes in the spectrum.

In addition, the Adak method also suffers from another extreme case. The Adak method gave only 39% of the simulated datasets that either had the ideal segmentation or were only slightly oversplit. The Auto-SLEX method, on the other hand, performed much better with an 81% rate. As already discussed, oversplitting means that block sizes are reduced and hence variance of the estimates can be inflated.

Although we have seen that the drawback of using dyadic transforms to a nondyadic process is not too severe, it can be a limitation. We note that both the Adak and the DMvS methods suffer from this limitation as well. Future work may lead to the use of the cycle-spinning approach of Coifman and Donoho (1995).

Using a dyadic transform, however, is still very attractive from a computational point of view and can be justified theoretically as we have shown in Section 5. Dyadic transforms are part of the tradition of the enormous literature that uses binary trees (including DMvS and Adak) and the BBA to create useful and efficient methods.
6.4 Slowly Varying AR(2) Process

Some random processes have spectral properties that change slowly over time. Consider, for example, the slowly changing AR(2) process \( Y_t = a_t Y_{t-1} - .81 Y_{t-2} + \xi_t \), \( t = 0, \ldots, 1, 023 \), where \( a_t = .8[1 - .5 \cos(\pi t/1,024)] \); and \( \xi_t \) are iid standard normal.

The plot of a simulated time series is shown in Figure 11. The ASE's for the Auto-SLEX and Adak estimates are .24325 and .31063, respectively, and both are the 40th percentile of their respective ASE values. The Auto-SLEX and Adak estimates are shown in Figure 12. In this example, we see that both the Auto-SLEX and Adak methods perform reasonably well even when the true process is not piecewise and does not have a dyadic structure.

We also note in Table 4 that the Auto-SLEX method tends to give segmentations that have larger blocks, whereas the Adak method gave smaller blocks. The first parameter of the AR(2) process changes very slowly over time and the second parameter remains constant in time. Hence, a segmentation that does not give blocks that are too short is certainly reasonable from the point of view that the variance of spectral estimates is smaller for blocks that are larger. This is reflected in the ASE values reported in Table 4. The Auto-SLEX method gave ASE values that are generally smaller than those given by the Adak method.

7. ANALYSIS OF A SPEECH SIGNAL

We applied the Auto-SLEX method to a speech signal that is a recording of the word “greasy”; see Figure 13. The speech signal “G-R-E-A-S-Y” has \( T = 8,192 \) time points and also was analyzed in Adak (1998). The speech signal is clearly nonstationary, but it is reasonable to break up the signal into approximately stationary intervals that correspond to G, R, EA, S, and Y.
The Auto-SLEX estimate of the time-varying spectrum is given in Figure 14. Note how the Auto-SLEX method is able to detect the changes in the spectrum by giving a segmentation of the signal that roughly divides according to G, R, EA, S, and Y. The spectrum at the blocks that correspond to G have high power at the lowest frequencies (below .05 of the Nyquist frequency), whereas the blocks that correspond to S have high power at high frequencies (.3-.5 Nyquist frequency). It is also interesting to note that the blocks that correspond to EA show the evolution of power from lower frequencies to higher frequencies (0-.3 Nyquist frequency). Furthermore, the blocks that correspond to R show power at frequencies at .05-.1 Nyquist frequency)—slightly above that for G.

This information also is conveyed in the spectral estimate given in figure 11 of Adak (1998). The main difference between the Auto-SLEX estimate and the Adak estimate is that the Auto-SLEX method was able to detect the change between S and Y. The Adak method, based on figure 11 of Adak (1998), however, missed this very important distinction. Clearly, Figure 13 suggests that the waves oscillate faster at S than at Y.

8. ANALYSIS OF A BIVARIATE EEG DATASET

We present an application of the Auto-SLEX method to EEG’s recorded during an epileptic seizure at two channels, namely the left temporal (T3) and left parietal (P3), during an epileptic seizure. The EEG recordings are shown in Figure 1. Each EEG time series has length $T = 32,768$ and was observed at the sampling rate of 100 Hz.

The Auto-SLEX method gave a segmentation of the bivariate EEG time series where the first half was left undivided as a stationary block and the last half was divided into very fine small blocks; see Figure 15. Recall that seizure for this particular episode, according to our neurologist, started right about $t = 185$ s, which was approximately at “half-time.” Thus, the Auto-SLEX method was able to detect the change in the spectra at approximately the same time that the actual change occurred. After the onset of seizure, the Auto-SLEX method divided the second half of the EEG into smaller blocks. This is again consistent with the known fact that seizure is characterized by an evolution of waves from lower frequencies to higher frequencies. Over time, power is shifted from lower to higher frequencies. Seizure is not a static
Table 4. Slowly Varying AR(2) Process

<table>
<thead>
<tr>
<th>Number of blocks</th>
<th>Adak Ave. ASE (Std. dev)</th>
<th>Auto-SLEX Ave. ASE (Std. dev.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 4</td>
<td>0</td>
<td>14.0 (0.25271)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>27.0 (0.22790)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>35.0 (0.23191)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2.0</td>
<td>18.0 (0.24270)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7.0</td>
<td>15.0 (0.26920)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9.0</td>
<td>1.0 (0.30760)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>15.0</td>
<td>0 (0.31341)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 11</td>
<td>67.0</td>
<td>0 (0.36033)</td>
</tr>
</tbody>
</table>

NOTE: N = 100 simulated datasets were obtained from the slowly varying AR(2) process in Section 6.4. For each simulated dataset, we obtained the Adak and Auto-SLEX segmentation and estimates, and counted the number of blocks that define the segmentation chosen by each method.

process, but one that evolves over time. Thus, the Auto-SLEX segmentation of the bivariate EEG time series is not only biologically plausible, but also consistent with clinical facts about epileptic seizures.

Also note that prior to seizure, power was concentrated at the lower frequencies. During seizure, however, power spread to all frequencies. This is consistent with the known clinical fact that large populations of neurons with diverse firing frequencies are recruited during the seizure. As the seizure drew to an end, the concentration of power was slowly restored to lower frequencies. Another interesting observation is that dramatic changes in the estimate of the log spectrum at P3 came later than $t = 185$ s. This is also consistent with the fact that the patient was diagnosed with left temporal lobe epilepsy. Hence, seizure for this patient is expected to start at T3 and eventually spread to other regions.

The Auto-SLEX estimates of the time-varying coherence [cf. Eq. (16)], plotted in Figure 16, showed that the EEG's were strongly coherent at the alpha frequency band at about 9–12 Hz prior to the onset of seizure. Furthermore, there was strong coherence at a narrow band at about 40Hz. Another important observation is that the EEG's at T3 and P3 were strongly coherent at low frequencies soon after the onset of seizure toward the middle of the seizure period. Strong coherence, however, was shifted to higher frequencies at the end of the seizure. This is again consistent with the known fact that seizure is characterized by the evolution of waves with increasing frequencies. The two EEG's were in synchrony during seizure: both started with lower frequency waves that evolved into higher frequency waves at the end of the seizure.

9. CONCLUSION

The existing methods for analyzing nonstationary time series are built on the assumption that these nonstationary time series can be modeled by processes that are stationary within a small time interval. We have shown that the Auto-SLEX method has the necessary elements of a good methodology for analyzing bivariate nonstationary random processes.
It is a procedure that was developed for bivariate time series and, hence, it can give estimates of the time-varying spectra and coherence. It uses the SLEX transform, which is localized in both time and frequency, and, hence, is ideal for representing nonstationary processes. The method is also computationally efficient, because it takes advantage of existing powerful algorithms: the FFT and BBA. We also have given strong theoretical justifications for the Auto-SLEX method. We have rigorously shown that, under a known segmentation, the smoothed SLEX periodograms are consistent estimators for the time-varying spectrum of the Dahlhaus locally stationary process. This important theoretical result tells us that the effect of the dyadic structure and the inclusion of another window, \( \psi_k \), vanish asymptotically. Finally, the Auto-SLEX method yields results that are easy to interpret because they parallel existing methods for stationary processes.

The Auto-SLEX method also is being further developed in both practical and theoretical directions. In practice, EEG’s can be recorded at 128 channels. The method (both on the segmentation and the estimation) has a natural extension from the bivariate to the multivariate setting. In addition, we are working on implementing SLEX (frequency) domain principal components analysis to extract the important information from datasets with high dimensionality. It is also of primary interest to model the time lags and coherence between channels, which will help the neurologist to gain some understanding about seizure origin at a focal region and seizure spread to other regions. Furthermore, we have proposed a new model of a nonstationary random process, which we call the SLEX model (Ombao et al. 2000). The SLEX model gives a Cramér-like representation that is not in terms of the Fourier complex exponentials, as in Equation (2), but is in terms of the localized SLEX vectors. The SLEX model is in the spirit of traditional spectral analysis because it gives a time-varying SLEX spectrum that is a time-frequency decomposition of power. We have begun theoretical work on the open questions of consistency of the Auto-SLEX estimator when the segmentation is not known, but is selected from the data by the BBA. We expect that the SLEX model of a nonstationary processes will facilitate our theoretical development.

**APPENDIX: PROOFS**

Proof of Theorem 1

First, we denote \( A^0_b(t, \omega) \) to be the transfer function of stationary process \( X_b^r \) that is defined on block \( b \). By Parseval’s identity, we have

\[
\sum_{t=0}^{T-1} E(X_{t, \tau} - \hat{X}_{t, \tau})^2 = \sum_{t=0}^{T-1} \frac{1}{\tau} |A^0_{t, \tau}(\omega)|^2 - \frac{B}{b=1} \int W^B(t) A^B_b(t, \omega) \, dw. \tag{A.1}
\]

Because \( \sum_{b=1}^{B} W^B(t) = 1 \) for all \( t \), it follows that

\[
|A^0_{t, \tau}(\omega) - \sum_{b=1}^{B} W^B(t) A^B_b(t, \omega)|^2 \leq \left\{ \sum_{b=1}^{B} W^B(t)|A^0_{t, \tau}(\omega) - A^B_b(t, \omega)|^2 \right\} \tag{A.2}
\]

We now fix \( t = t^* \) and assume that \( t^*/T \in [u_b, \nu_{b+1}] \). Recall that \( M = M_f \) denotes the upper bound of the length of (any) segment that must fulfill \( M/T \to 0 \) as \( T \to \infty \). Let \( \epsilon \) denote the overlap between the \( b \)th and \( (b+1) \)th blocks and let \( \Delta = \epsilon/T \) be the rescaled overlap. Only one of the following cases holds:

Case 1. \( t^*/T \in [u_b, \nu_b + \Delta) \).
Case 2. \( t^*/T \in [u_b + \Delta, \nu_{b+1} - \Delta) \).
Case 3. \( t^*/T \in [\nu_{b+1} - \Delta, \nu_{b+1}) \).

We consider the second case. When \( t^*/T \in [u_b + \Delta, \nu_{b+1} - \Delta) \), \( W^B(t) = W^B(t) \delta(b = b^*) \). Moreover, \( W^B(t) \leq 1 \) for all \( b \) and \( t \). Hence,

\[
\left\{ \sum_{b=1}^{B} W^B(t)|A^0_{t, \tau}(\omega) - A^B_b(t, \omega)|^2 \right\} \leq |A^0_{t, \tau}(\omega) - A^B_b(t, \omega)|^2. \tag{A.3}
\]

Next, by the triangle inequality,

\[
|A^0_{t, \tau}(\omega) - A^B_b(t, \omega)| \leq |A^0_{t, \tau}(\omega) - A(t/T, \omega)| + |A(t/T, \omega) - A^B_b(t, \omega)|. \tag{A.4}
\]

We bound both terms on the right side of the inequality. From Equation (3), \( |A^0_{t, \tau}(\omega) - A(t/T, \omega)| \leq D/T \). Using the Lipschitz continuity of \( A(u, \omega) \) in \( u \), \( |A(t/T, \omega) - A^0_b(t, \omega)| \leq L|t/T - u_b| \), where \( L > 0 \). Using the fact that \( t^*/T = T^* \) has maximal distance \( M/T \) from \( u_b \), we have \( |A(t/T, \omega) - A^0_b(t, \omega)| \leq LM/T \). Hence, \( |A^0_{t, \tau}(\omega) - A^B_b(t, \omega)| \leq D/T + LM/T \). It follows that \( |A^0_{t, \tau}(\omega) - A^B_b(t, \omega)|^2 \leq QM^2/T^2 \) for some \( Q > 0 \). Thus,

\[
|A^0_{t, \tau}(\omega) - \sum_{b=1}^{B} W^B(t) A^B_b(t, \omega)|^2 \leq QM^2/T^2. \tag{A.5}
\]

The argument for the two other (overlapping) cases is similar. We can bound the error from above by the sum over the contribution from the two segments. The key point here is that an approximately stationary small interval can be formed around the change points. In fact, at a small neighborhood about \( \nu_{b+1} \): \( [\nu_{b+1} - \Delta, \nu_{b+1} + \Delta] \), the transfer function \( A^0_{t, \tau} \) is “close” to both \( A^b(t, \omega) \) and \( A(\nu_{b+1}, \omega) \) when the locally stationary process is slowly changing. Finally,

\[
T^{-1} \sum_{t=1}^{T-1} E(X_{t, \tau} - \hat{X}_{t, \tau})^2 \leq T^{-1} \sum_{t=1}^{T-1} \frac{1}{\tau} \frac{1}{\tau} S(t) = QM^2/T^2. \tag{A.6}
\]
Proof of Theorem 2

The proof works by analogy to Dahlhaus (1996, theorem 2.2), because the SLEX periodogram is analogous to a tapered (Fourier) periodogram where the windows $\Psi_{+,+}(t/T)$ and $\Psi_{-,+}(t/T)$ play the role of two taper functions. These fulfill the assumptions of classical smooth tapers. Furthermore, these two tapers are orthogonal: $\sum_{s \in \mathbb{Z}} \Psi_{+,+}(t/T) \Psi_{-,+}(t/T) = 0$. This is simply a consequence of the different parity of the two tapers: $\Psi_{+,+}$ is even, while $\Psi_{-,+}$ is odd on the segment $S$. Moreover, $\sum_{s \in \mathbb{Z}} (\Psi_{+,+}(t/T) + \Psi_{-,+}(t/T)) = M_T = T/2$. Note that the sum is independent of $\Psi_{+,+}$, $\Psi_{-,+}$, and $\epsilon$, and that our SLEX periodograms are correctly normalized by the sums of squares of the taper values, as is usual for tapered periodograms.

For clarity, we first compare how the SLEX periodograms differ from the tapered periodograms in the case of a stationary process. Under stationarity, Parseval’s equality gives

$$ E(\theta) = M_T \sum_{i,j} \gamma(i-s) \phi_{x,i,j} \phi_{x,i,j}^* (t) $$

where $\gamma(i-s) = \text{cov}(X_i, X_j)$, and $\phi_{x,i,j}$ denotes the finite Fourier transforms of the windows $\Psi_{+,+}$ and $\Psi_{-,+}$, respectively. As known from classical results (e.g., Brillinger 1981, theorem 4.2.2),

$$ M_T \int f(\eta) \Delta(\omega, \eta) d\eta = f(\omega) + O(M_T^{-1}), \quad (A.8) $$

because $|\hat{\Psi}_{+,+}(0)|^2 + |\hat{\Psi}_{-,+}(0)|^2 = M_T$. Note that the regularity assumptions on both the window functions and the spectrum are fulfilled.

For the cross term, we observe that for $\omega \neq 0$,

$$ M_T^{-1} \int f(\eta) \hat{\Psi}_{+,+}(\omega) \hat{\Psi}_{-,+}(\omega + \eta) d\eta = \sigma(M_T^{-1}), \quad (A.9) $$

because the Fourier transform of $\Psi_{+,+}$ asymptotically concentrates around $\omega$, whereas that of $\Psi_{-,+}$ concentrates around $-\omega$.

We continue with the variance of the SLEX-periodogram in the stationary case. For simplicity, consider the situation of a Gaussian process $\{X_i\}_{i=1}^N$, for which we can use the Isserlis theorem (Brillinger 1981, p. 21):

$$ \text{var}(E_d(\omega)) = \text{cov}(d_2(\omega), d_2(\omega)) \text{cov}(\hat{d}_2(\omega), \hat{d}_2(\omega)) + \text{cov}(d_2(\omega), \hat{d}_2(\omega)) \text{cov}(\hat{d}_2(\omega), d_2(\omega)). \quad (A.10) $$

Using the preceding result for the expectation $E_d(\omega)$, we observe that the first term in the sum tends to $f^2(\omega)$ with a rate on the order of $M_T^{-1}$. The second term tends to zero with this rate, because $\text{cov}(d_2(\omega), \hat{d}_2(\omega)) = E_{\hat{d}_2}^2(\omega) = M_T^{-1} \int f(\eta) \hat{\Delta}(\omega, \eta) d\eta$, with

$$ \hat{\Delta}(\omega, \eta) = \hat{\Psi}_{+,+}(\omega + \eta) \hat{\Psi}_{-,+}(\omega - \eta) + \hat{\Psi}_{-,+}(\omega + \eta) \hat{\Psi}_{+,+}(\omega - \eta) $$

$$ + 2 \hat{\Psi}_{+,+}(\omega - \eta) \hat{\Psi}_{-,+}(\omega - \eta). \quad (A.11) $$

Here, the first two terms behave exactly like the cross term in Equation (A.9) whereas for the final term, we observe (e.g., once again by the proof of theorem 4.3.2 of Brillinger 1981, pp. 402 ff) that

$$ M_T^{-1} \int f(\eta) \hat{\Psi}_{+,+}(\omega - \eta) \hat{\Psi}_{-,+}(\omega - \eta) d\eta $$

$$ = M_T^{-1} \sum_t \Psi_{+,+}(t/T) \Psi_{-,+}(t/T) f(\omega) + O(M_T^{-1}). \quad (A.12) $$

Whereas $\sum_t \Psi_{+,+}(t/T) \Psi_{-,+}(t/T) = 0$, it follows that $M_T^{-1} \int f(\eta) \hat{\Psi}_{+,+}(\omega - \eta) \hat{\Psi}_{-,+}(\omega - \eta) d\eta = O(M_T^{-1})$. In total, var$(I_d(\omega)) = f^2(\omega) + O(M_T^{-1})$.

Turning to our situation of local stationarity, we adapt the proof of theorem 2.2 part (i) of Dahlhaus (1996) to the situation of our particular taper functions (Theorem 2, parts 1-3) to arrive at the same leading terms as in the stationary case. Note again that the assumptions for the tapers of Dahlhaus (1997, lemmas A.5 and A.9) are fulfilled. Actually, a taper function of bounded variation would be sufficient. We give only a brief sketch.

**Proof of Part 1.** To treat $E_d(\omega)$ we need to replace the term $\gamma(t-s)$ in equation (A5) with

$$ \gamma(t,s) = \text{cov}(X_i, X_j) $$

Further, we replace $A_{\omega,\epsilon}^{(i)}(t,\epsilon)$ with $A_{\omega,\epsilon}^{(i)}(t,\epsilon)$ and use the same techniques as in Dahlhaus (1996) to bound the replacement error by a term of $O(\log(M_T)/M_T)$. We do the same for the factor with $s$ instead of $t$ and integrate over $\omega$. Taylor expansion of $f(u-(M/2+t)/T, \omega)$ around $u$ allows us to use the same ideas as in the stationary case and to finally arrive at the result of part (1) with

$$ C_{\Psi_d}(u) = 1/2 \int_0^1 (x-1/2)^2 \left( \Psi_d^2(x) + \Psi_d^2(x) \right) dx, \quad (A.13) $$

**Proof of Part 2.** Part 2 follows from part 1, again by Taylor expansion of the spectrum, but now around $\omega$. Here,

$$ C_{K_d}(u) = 1/2 \int \eta^2 K(u, \omega) d\omega, \quad (A.14) $$

**Proof of Part 3.** Part 3 is a consequence of the proof of lemma A.9 in Dahlhaus (1997) for the case where the length of overlap on each side $\epsilon = M/2$ (or the case $S = N$ following the notation of the author), again with the necessary modification due to the presence of two terms, $\Psi_{+,+}$ and $\Psi_{-,+}$, in the taper function. It can be observed that again the principle applies as shown for the variance of $I_d(\omega)$ in the stationary case. Hence, for the factor that inflates the asymptotic variance of the kernel estimator $f(u, \omega)$, we simply get

$$ \lim_{T \to \infty} M_T^{-1} \int_0^1 \left[ \Psi_d^2(t/T) + \Psi_d^2(t/T) \right]^2 dx = \int_0^1 \left( \Psi_d^2(x) + \Psi_d^2(x) \right)^2 dx, \quad (A.15) $$

where we denote $C_{\Psi_d}$ for $C_{\Psi_d}(u, \omega) d\omega$. This ends the proof of Theorem 2.

[Received July 1999. Revised March 2000.]

**REFERENCES**


