ELSEVIER

Journal of Statistical Planning and Inference 57 (1997) 195–214

journal of statistical planning and inference

Optimal transformations and the spectral envelope for real-valued time series

A.J. McDougall[a], D.S. Stoffer[b], D.E. Tyler[a,*]

[a] *Department of Statistics, Rutgers University, New Brunswick, NJ 08903, USA*

[b] *Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260, USA*

Received 1 February 1995; revised 1 May 1995

Abstract

The concept of a *spectral envelope* for exploring the periodic nature of real-valued time series is introduced. This concept follows naturally from the data-dependent approach proposed by Stoffer et al. (1993) for spectral analysis and scaling of categorical processes. Here, the notion of the spectral envelope is applied in the context of transformations of a time series, and a data-dependent approach for selecting optimal transformations is proposed. These transformations help emphasize periodicities that may exist in the real-valued process. The definition of the spectral envelope is also extended to include multivariate time series. Several examples are used to illustrate the application of this methodology and asymptotic properties of the procedure are established.

AMS classification: 62M15

Keywords: Spectral envelope; Transformations; Time series

1. Introduction

In recent work, Stoffer et al. (1993) discuss the frequency domain analysis of categorical time series and introduce a concept called the *spectral envelope* of a categorical time series. They note that the obvious approach to exploring the periodic nature of a categorical process is to assign numerical values to each of the categories followed by a spectral analysis of the resulting discrete-valued time series. The resulting spectrum however, will depend upon the particular values chosen for the categories. Thus, rather than choose arbitrary values, Stoffer et al. (1993) propose a data-dependent approach based on the spectral envelope for selecting values which help emphasize any periodic features that may exist in the categorical process. An application of this methodology to a problem involving DNA data is given in Stoffer et al. (1993).

* Corresponding author. Tel.: +1 908 445 2691; fax: +1 908 445 3428; E-mail: dtyler@stat.rutgers.edu.

PII S0378-3758(96)00044-4

Similar arguments can be made for real-valued series. In exploring the periodic nature of a real-valued process, one can consider the spectral analysis of not only the original time series, but also of transformations of the time series. The spectrum of a transformed time series will depend on the particular transformation. Bloomfield (1976) notes that simple power transformations can sometimes be used to make a time series more closely sinusoidal and hence the resulting spectrum easier to interpret than that of the original process. Again, rather than arbitrarily choosing a transformation, we propose in the next section a data-dependent approach for selecting transformations which help emphasize any periodic features that may exist in the real-valued process. This data-dependent approach for real-valued series is based on an extension of the spectral envelope methodology developed in Stoffer et al. (1993). The definition of the spectral envelope can be extended to include multivariate time series as we also show in the next section.

The paper is organized as follows. In the next section, we define the spectral envelope and discuss its use in finding interesting real-valued transformations for univariate or multivariate time series. We also discuss the construction of the spectral envelope and note a relationship between the spectral envelope and Brillinger's (1981) principal components analysis of a multivariate time series. In Section 3, we discuss estimation of the spectral envelope and present some asymptotic results. Most of the asymptotic results follow readily from the results given in Stoffer et al. (1993). Applications of the spectral envelope and data-dependent transformations are presented in Section 4. Three of the examples involve real data, including one in which a multivariate time series is examined. We conclude with some final comments in Section 5.

2. The spectral envelope

Let X_t, $t = 0, \pm 1, \pm 2, \ldots$, be a stationary time series and consider some real-valued transformation of the time series $g(X_t)$. Our goal here is to find transformations from some class $\mathscr{G}$ so that the resulting spectral density is in some sense most informative. The approach we take is to first consider a given frequency $-\pi < \omega \leqslant \pi$, and observe that the transformation of the time series which best emphasizes the frequency ω is that in which the resulting normalized spectral density $f_g(\omega)/\sigma_g^2$ is largest at ω. Here, $f_g(\omega)$ represents the spectral density of $g(X_t)$ and $\sigma_g^2 = \int f_g(\omega)\,\mathrm{d}\omega$ is the total power of $g(X_t)$. We then propose looking at the frequencies and the corresponding transformations for which $f_g(\omega)/\sigma_g^2$ is relatively large. Note that if the sequence X_t is iid with finite variance, also referred to as (pure or strict) white noise, then $f_g(\omega)/\sigma_g^2 = 1/2\pi$ since $g(X_t)$ will also be white noise, and so no frequency or transformation 'stands out'.

The above idea leads naturally to the notion of the "spectral envelope" given below. This approach is applicable to multivariate time series as well as univariate, and so we formally introduce a general definition of the spectral envelope.

Definition 2.1. Let X_t, $t = 0, \pm 1, \pm 2, \ldots$, be a time series on $\mathbb{R}^p$ and let $\mathscr{G}$ be a class of transformations from $\mathbb{R}^p$ to $\mathbb{R}$ such that for any $g \in \mathscr{G}$, the spectral density $f_g(\omega)$

of $g(X_t)$ exists. The *spectral envelope* of X_t with respect to $\mathscr{G}$, is defined to be

$$\lambda(\omega) = \sup_{g \in \mathscr{G}} \{ f_g(\omega) / \sigma_g^2 \}, \quad -\pi < \omega \leqslant \pi.$$

The quantity $\lambda(\omega)\,\mathrm{d}\omega$ represents the largest proportion of total power that can be attributed to the frequencies $\omega\,\mathrm{d}\omega$ for *any* transformation within the class $\mathscr{G}$. We refer to the transformation $g \in \mathscr{G}$ at which the supremum is obtained as the *optimal* transformation at ω with respect to the class $\mathscr{G}$. The name 'spectral envelope' is appropriate since the graph of $\lambda(\omega)$ defines an upper bound or 'envelope' for *all* the normalized spectra associated with the class $\mathscr{G}$. That is, $f_g(\omega)/\sigma_g^2 \leqslant \lambda(\omega)$ with equality being obtained at the optimal transformation. The optimal transformation is not the same for all ω. As will be seen in the examples in Section 4, different transformations may help emphasize different frequencies. If X_t is white noise then $\lambda(\omega) = \frac{1}{2}\pi$ and any transformation would by default satisfy the definition of an optimal transformation.

In Definition 2.1 we do not specifically require that X_t be a weakly stationary time series since the existence of $\lambda(\omega)$ only depends on the transformed time series possessing spectral densities. In practice, simple non-instantaneous transformations are often used to induce stationarity in a time series; the growth rate $\nabla \log X_t$ for example, is a standard type of transformation employed in economics. To be able to include such cases, Definition 2.1 can be applied to the vector process $Y_t = (X_{1t}, X_{2t}, \ldots, X_{pt})^\tau$ where the components X_{jt} are judiciously choosen. Thus, the implicit instantaneous form of Definition 2.1 is not as restrictive as it might first seem.

In general, using Definition 2.1 directly to construct a spectral envelope is complicated since we need to take the supremum over a class of functions for each frequency. However, if $\mathscr{G}$ corresponds to a finite-dimensional vector space, then we show below that finding the supremum reduces to a finite-dimensional eigenvalue problem for each frequency. In particular, we note a special case of this situation is where X_t is a categorical series with k states. Here, assigning all possible numerical values to the states corresponds to the class of transformations $\mathscr{G}$ represented by all real-valued functions on the set $\{1, 2, \ldots, k\}$. Alternatively, an orthogonal basis for $\mathscr{G}$ can be employed by setting the $g_j(X_t)$ that appear in (2.1) below, to be the indicator functions for $\{X_t \in \text{state } j\}$. Using the definition of the spectral envelope with respect to $\mathscr{G}$ is then seen to be consistent with the definition of the spectral envelope given in Stoffer et al. (1993) for categorical time series. While there exists a natural choice of $\mathscr{G}$ for categorical series, this is not the case with real-valued series and we shall return to this issue following the examples presented in Section 4.

Henceforth, we assume that $\mathscr{G}$ is a k-dimensional vector space. As such, there exists a set of basis functions $\{g_1, \ldots, g_k\}$ such that for any $g \in \mathscr{G}$

$$g(X_t) = \beta_1 g_1(X_t) + \cdots + \beta_k g_k(X_t). \tag{2.1}$$

Since $\mathscr{G}$ is fixed, (2.1) may be represented as

$$X_t(\beta) = \beta^\tau Y_t, \tag{2.2}$$

where Y_t is the vector process

$$Y_t = (g_1(X_t), \dots, g_k(X_t))^\tau \tag{2.3}$$

and $\beta = (\beta_1, \dots, \beta_k)^\tau \in \mathbb{R}^k$. If the *basis* vector Y_t is assumed to possess a continuous spectral density, $f(\omega)$ say, then $X_t(\beta)$ will have a continuous spectral density $f(\omega;\beta)$ for all $\beta \in \mathbb{R}^k$. Moreover, (2.2) implies that

$$f_g(\omega) \equiv f(\omega;\beta) = \beta^\tau f(\omega)\beta$$

and

$$\sigma_g^2 \equiv \mathrm{Var}[X_t(\beta)] = \beta^\tau V\beta,$$

where

$$V = \mathrm{Var}(Y_t) = \int_{-\pi}^{\pi} f(\omega)\,\mathrm{d}\omega$$

is the total power of Y_t. Thus, the spectral envelope over $\mathscr{G}$ is well-defined and can be expressed as

$$\lambda(\omega) = \sup_{\beta \in \mathscr{B}} \{\beta^\tau f(\omega)\beta / \beta^\tau V\beta\}, \quad -\pi < \omega \leqslant \pi, \tag{2.4}$$

where $\mathscr{B} = \{\beta \mid V\beta \neq 0\}$.

The problem of obtaining $\lambda(\omega)$ is simplified by the fact that only real-valued scalings are considered. Let $f^{\mathrm{re}}(\omega)$ denote the real part of $f(\omega)$. Since $f(\omega)$ is Hermitian, $\beta^\tau f(\omega)\beta = \beta^\tau f^{\mathrm{re}}(\omega)\beta$ for all $\beta \in \mathbb{R}^k$, and so (2.4) may be replaced by

$$\lambda(\omega) = \sup_{\beta \in \mathscr{B}} \{\beta^\tau f^{\mathrm{re}}(\omega)\beta / \beta^\tau V\beta\}, \quad -\pi < \omega \leqslant \pi.$$

In addition, $f^{\mathrm{re}}(-\omega) = f^{\mathrm{re}}(\omega)$ implies that $\lambda(\omega)$ is symmetric so it suffices to consider $\lambda(\omega)$ on $0 \leqslant \omega \leqslant \pi$. The relationship between $\lambda(\omega)$ and the optimal linear coefficients $\beta(\omega)$ can also be defined as the largest scalar $\lambda(\omega)$ such that for some $\beta(\omega)$, with $V\beta(\omega) \neq 0$,

$$f^{\mathrm{re}}(\omega)\beta(\omega) = \lambda(\omega)V\beta(\omega). \tag{2.5}$$

When V has full rank, the supremum is over all $\beta \neq 0$ and it follows that $\lambda(\omega)$ is the largest eigenvalue associated with the determinantal equation

$$|f^{\mathrm{re}}(\omega) - \lambda V| = 0. \tag{2.6}$$

The assumption that V has full rank is essentially a costless one since in order for V to be singular, we must have $\alpha^\tau Y_t \equiv 0$ a.s. for some $\alpha \neq 0$. This would imply that one (or more) of the components of Y_t is identically equal to zero, or there exists an *exact* linear dependency between the components of Y_t. In either case, the dimension of $\mathscr{G}$ can be reduced without any loss of information.

When $X_t = (X_{1t}, X_{2t}, \dots, X_{pt})^\tau$ is a multivariate time series, one class of transformations of interest are linear combinations of X_t. In Tiao et al. (1993) for example, linear transformations are used in a time domain approach to investigate dynamic and contemporaneous relationships among the components of a multivariate time series. In the context of a spectral envelope analysis, this class forms a k-dimensional vector space of transformations and a natural choice of basis functions are the k components themselves; that is, $Y_t = X_t$ in (2.2). In this case, the spectral envelope has an interesting relationship to the principal components analysis of multivariate time series which is discussed in Brillinger (1981). If we take a multivariate time series and standardize it so the total power is the identity matrix, then Brillinger's first principal component spectrum would correspond to $\lambda(\omega)$ in (2.5), but with $f^{\mathrm{re}}(\omega)$ replaced by $f(\omega)$. Alternatively, the same result would be obtained if we allow for complex linear combinations in our definition of the spectral envelope. In this case, the complex linear combination corresponds to the instantaneous transfer function associated with $\lambda(\omega)$. Thus, a spectral envelope interpretation can be given to Brillinger's first principal component spectrum which also corresponds to the best rank one filtering of the multivariate time series.

3. Estimation of the spectral envelope

To apply the spectral envelope procedure in practice, $\lambda(\omega)$ needs to be estimated from the observed data and this can be effected by replacing $f(\omega)$ in (2.4) by a spectral estimate, $f_T(\omega)$, computed from Y_t, $t = 0, 1, \dots, T-1$. The periodogram

$$I_T(\omega) = \frac{1}{2\pi T} \left| \sum_{t=0}^{T-1} Y_t \exp\{-\mathrm{i}\omega t\} \right|^2, \quad -\pi < \omega \leqslant \pi \tag{3.1}$$

provides a simple estimate of $f(\omega)$ and in this case we refer to the resulting $\lambda_T(\omega)$ as the *sample spectral envelope* estimate, and $\beta_T(\omega)$ as the *sample scalings*. Note that $I_T(\omega)$ may also be expressed as

$$I_T(\omega) = (2\pi T)^{-1} d_T(\omega) d_T^*(\omega), \qquad d_T(\omega) = \sum_{t=0}^{T-1} Y_t \exp\{-\mathrm{i}\omega t\},$$

where $d_T(\omega)$ is the finite Fourier transform of Y_t with * denoting conjugation with transposition. Alternatively, a smoothed periodogram estimate or a consistent spectral window estimate for $f(\omega)$ can be used to obtain $\lambda_T(\omega)$ and $\beta_T(\omega)$.

The theory for estimating the spectral density of a vector process is well established (Brillinger, 1981; Hannan, 1970; Rosenblatt, 1959), and can be applied to $f(\omega)$ corresponding to the basis vector (2.3). However, in order that certain assumptions be satisfied by the Y_t process, we shall first need to place some restrictions on the transformations used to construct Y_t. To motivate these restrictions, we reconsider the transformation (2.1) associated with $\mathscr{G} = \{g_1, \dots, g_k\}$ which is repeated here, for convenience,

$$g(X_t) = \beta_1 g_1(X_t) + \cdots + \beta_k g_k(X_t).$$

A sufficient condition for the spectrum of $g(X_t)$ to be bounded and uniformly continuous is that $\sum |\gamma_g(r)| < \infty$ where $\gamma_g(r) = \mathrm{Cov}[g(X_t), g(X_{t-r})]$. This implies that the individual autocovariances associated with each $g_j(X_t)$ and the cross-covariances between any pair of basis functions in $\mathcal{G}$, must also satisfy the summability condition. That is, $\sum \|\Gamma(r)\| < \infty$ where

$$\Gamma(r) = [\mathrm{Cov}[g_i(X_t), g_j(X_{t-r})]], \quad i,j = 1,\dots,k, \; r = 0, \pm 1, \dots \tag{3.2}$$

is the $k \times k$ autocovariance matrix associated with $\mathcal{G}$. As mentioned previously we may take $E[g_i(X_t)] = 0$ without loss of generality since the spectral density of $g(X_t)$ is identical to that of $g(X_t) - \beta_0$ except at the zero frequency which is not used in computing $\lambda_T(\omega)$. This implies that none of the basis transformations can be a constant function of X_t and we shall also assume that $\mathrm{Var}[Y_t]$ has full rank.

We should point out that the lack of linear independence does not preclude the use of the spectral envelope procedure since $\lambda(\omega)$ can be defined as in (2.5). Computationally however, it is more convenient to introduce a constraint matrix A such that $A^{\tau} V A$ is positive definite of rank r and $A^{\tau} A = I_r$. Then $\lambda(\omega)$ can be defined as the largest eigenvalue of the determinantal equation

$$|A^{\tau} f^{\mathrm{re}}(\omega) A - \lambda A^{\tau} V A| = 0,$$

where $\lambda(\omega)$ does not depend on the choice of A. Although the corresponding eigenvector, $b(\omega)$ say, does depend on A, the equivalence class of scalings associated with $\beta(\omega) = Ab(\omega)$ does not depend on A. In the categorical case considered by Stoffer et al. (1993), the need for a constraint matrix arose naturally from the multinomial form of Y_t. In summary then, the above conditions on $\mathcal{G}$ may be formally written as follows.

Condition A. For a stationary time series X_t, $t = 0, \pm 1, \dots$, let $\mathcal{G} = \{g_1, \dots, g_k\}$ denote a vector space of a.e. continuous real-valued zero-mean transformations of X_t, satisfying

(a) $\sum \|\Gamma(r)\| < \infty$ where $\Gamma(r)$ is given by (3.2),
(b) $\Gamma(0)$ is strictly positive definite.

To allow for the application of a general theory in obtaining asymptotic distributions for the estimates of the spectral density $f(\omega)$, it will be necessary to impose further assumptions on the Y_t process. These conditions could certainly be relaxed but we do not pursue that issue here. Rather than introduce excessive notation we simply state the assumptions as follows.

Condition B. In addition to Condition A, X_t is assumed to be strictly stationary and the fourth-order cumulant spectrum exists for all $g \in \mathscr{G}$.

Under Condition B it follows from Doob (1953) that Y_t will be a strictly stationary process and possesses a fourth-order cumulant spectrum. The second part of these conditions may be contrasted with the categorical case treated in Stoffer et al. (1993) where it was costless to assume all cumulant spectra exist. This distinction for real-valued series is necessary since we intend to accommodate nonlinear processes, such as the bilinear model presented in the next section, within the framework of the spectral envelope methodology. In the case of a bilinear process for example, it can be shown that $E[X_t^m]$ does not exist for m sufficiently large (Tong, 1990; Granger and Anderson, 1978). However, the above assumption is sufficient for the conclusions we shall require about $f_T(\omega)$ which are comprehensively treated in Hannan (1970). The results obtained by Stoffer et al. (1993) for the categorical case will therefore be applicable to the real-valued case, subject to the above conditions being satisfied. Before considering these results it is worth noting that the approach we have adopted here avoids the need to deal specifically with the X_t process. The problem with nonlinear transformations of a random process is that there does not exist a systematic theory comparable to that associated with linear systems which would make an analytical treatment exceedingly complex. An illuminating discussion for the case where X_t is Gaussian is given in Hannan (1970, II, Section 7).

Following Stoffer et al. (1993) we may now define $\lambda_T(\omega)$, without loss of generality, to be the largest eigenvalue of $h_T^{\mathrm{re}}(\omega)$ where

$$h_T(\omega) = V_T^{-1/2} f_T(\omega) V_T^{-1/2}$$

and V_T is the sample covariance matrix of Y_t, $t = 0, 1, \dots, T-1$ constructed from $\{g_1, \dots, g_k\}$ which satisfy Condition A. The scaling $\beta_T(\omega)$ is then defined by $V_T^{1/2}\beta_T(\omega)$ which is the eigenvector of $h_T^{\mathrm{re}}(\omega)$ associated with $\lambda_T(\omega)$. We shall take $\beta_T(\omega)$ to be the normalized scaling satisfying $\beta_T(\omega)^{\tau} V_T \beta_T(\omega) = 1$, with the first non-zero entry of $V_T^{1/2}\beta_T(\omega)$ being positive. In the results presented below, all limiting statements are taken as $T \to \infty$ and the distinct frequencies ω_j, $j = 1, \dots, J$ associated with $f_T(\omega)$ are assumed to be in the interval $(0, \pi]$. Let $W(p, v, \Sigma)$ denote the Wishart distribution of dimension p with v degrees of freedom and $p \times p$ covariance parameter Σ. Similarly, the notation $W_c(p, v, \Sigma)$ refers to the complex Wishart distribution.

We first consider the case where the periodogram $I_T(\omega)$ defined in (3.1) is used as the spectral estimate for $f(\omega)$ and Y_t satisfies Condition B. Then from Hannan (1970) we have

Lemma 3.1. *Under the established conditions on Y_t and $\mathscr{G}$, $I_T(\omega_j)$, $j = 1, \dots, J$ converge in distribution to independent $W_c[k, 1, f(\omega_j)]$, $j = 1, \dots, J$.*

This lemma enables the asymptotic distributions of *sample* estimates $\lambda_T(\omega)$ and $\beta_T(\omega)$ to be obtained by noting that eigenvalues and eigenvectors are continuous functions of a matrix argument a.e. with respect to Lebesgue measure.

Theorem 3.1. *Under the established conditions on Y_t and $\mathscr{G}$ the collection*

$$\{\lambda_T(\omega_j), \beta_T(\omega_j)\} \xrightarrow{\mathrm{d}} \{\lambda_j, u_j\}, \quad j = 1, \dots, J,$$

where $v_j = V^{-1/2}u_j$ and $\{\lambda_j, u_j\}$, $j = 1, \dots, J$ are the largest eigenvalue and eigenvector of independent $W_c^{\mathrm{re}}[k, 1, h(\omega_j)]$ matrices $j = 1, \dots, J$, with u_j normalized so that $u_j', u_j = 1$ and the first non-zero entry of u_j is positive.

Of particular interest is the case where X_t (and hence Y_t) is white noise.

Theorem 3.2. *Under the established conditions on Y_t and $\mathscr{G}$, and if in addition X_t is white noise, then the collection*

$$\lambda_T(\omega_j) \xrightarrow{\mathrm{d}} \lambda_j, \quad j = 1, \dots, J,$$

where the λ_j are iid with

$$P[2\pi\lambda_1 < x] = P[\chi^2_{2k} < 4x] - \pi^{1/2}x^{(k-1)/2}\exp(-x)P[\chi^2_k < 2x]/\Gamma(k/2) \tag{3.3}$$

for $0 < x < \infty$.

The proof of this result follows immediately from that given in Stoffer et al. (1993, Theorem 3.2) and provides us with an asymptotic test which can be used to assess the significance of peaks in the sample spectral envelope plot.

The asymptotic distributions of $\lambda_T(\omega)$ and $\beta_T(\omega)$ can also be derived when alternative estimators for $f(\omega)$ are used, such as the smoothed periodogram estimate

$$f_T(\omega) = (2m+1)^{-1} \sum_{s=-m}^{m} I_T\left(\omega + \frac{2\pi s}{T}\right) \tag{3.4}$$

or a consistent spectral window estimate

$$f_T(\omega) = \frac{2\pi}{T} \sum_{s=1}^{T-1} W_T\left(\omega - \frac{2\pi s}{T}\right) I_T\left(\frac{2\pi s}{T}\right), \tag{3.5}$$

where $W_T(\alpha) = B_T^{-1}\sum_{j=-\infty}^{\infty} W(B_T^{-1}(\alpha + 2\pi j))$ and $W(\alpha)$ is a spectral window function satisfying $\int_{-\infty}^{\infty} W(\alpha)\,\mathrm{d}\alpha = 1$, $\int_{-\infty}^{\infty} |W(\alpha)|\,\mathrm{d}\alpha < \infty$. The bandwidth parameter B_T is assumed to satisfy the conditions $B_T \to 0$, $B_T T \to \infty$ as $T \to \infty$. It will also be convenient to define the normalization constant

$$v_T = (B_T T)^{1/2}\left\{2\pi \int_{-\infty}^{\infty} W(\alpha)^2\,\mathrm{d}\alpha\right\}^{1/2}. \tag{3.6}$$

For the smoothed periodogram estimate given in (3.4), all the previous results hold with minor modification. In Lemma 3.1, $f_T(\omega)$ and $W_c[k, 2m+1, f(\omega_j)]/(2m+1)$ replace $I_T(\omega)$ and $W_c[k, 1, f(\omega_j)]$ with the analogous adjustment in Theorem 3.1. Similarly, Theorem 3.2 also holds with $f_T(\omega)$, but the distribution of $2\pi\lambda_1$ is that of the largest root of a $W(k, 4m+2, I_k)/(2m+1)$ matrix which differs from the relationship given by (3.3). For a discussion of the largest root of a $W(p, v, I_p)$ matrix we refer the reader to Muirhead (1982). The situation for (3.5) however, is more complicated and depends on whether the largest root of $h^{\mathrm{re}}(\omega_j)$ is distinct or not. For the case where the largest root of $h^{\mathrm{re}}(\omega_j)$ is distinct we have from Stoffer et al. (1993) and Hannan (1970):

Theorem 3.3. *Under the established conditions on Y_t and $\mathscr{G}$ with $f_T(\omega)$ defined by* (3.5), *then provided the largest root of $h^{\mathrm{re}}(\omega_j)$ is distinct for each $j = 1, \ldots, J$,*

$$\left\{ v_T \left(\frac{\lambda_T(\omega_j) - \lambda(\omega_j)}{\lambda(\omega_j)} \right), v_T(\beta_T(\omega_j) - \beta(\omega_j)) \right\} \xrightarrow{\mathrm{d}} \{z_j, y_j\}, \quad j = 1, \ldots, J,$$

where v_T is defined as in (3.6) *and z_j, y_j being independent for $j = 1, \ldots, J$. Furthermore, for each $j = 1, \ldots, J$, $z_j \sim \mathrm{N}(0, 1)$ and is independent from y_j which is multivariate normal with mean 0. The variance of $V^{-1/2} y_j$ is given by*

$$\{\lambda(\omega_j) H(\omega_j)^+ h^{\mathrm{re}}(\omega_j) H(\omega_j)^+ - a(\omega_j) a(\omega_j)^{\tau}\}/2, \tag{3.7}$$

where $H(\omega_j) = h^{\mathrm{re}}(\omega_j) - \lambda(\omega_j) I_{k-1}$, $a(\omega_j) = H(\omega_j)^+ h^{\mathrm{im}}(\omega_j) V^{1/2} \beta(\omega_j)$ and $H(\omega_j)^+$ refers to the Moore–Penrose inverse of $H(\omega_j)$.

The proof of this result follows that given in Stoffer et al. (1993, Appendix) and enables asymptotic normal confidence intervals and tests for $\lambda(\omega)$ to be constructed. Similarly, asymptotic confidence ellipsoids and chi-square tests can also be constructed for $\beta_T(\omega)$ using Theorem 3.3. As shown in the above reference, a simpler but more conservative asymptotic test can be constructed by omitting the $a(\omega)$ term in (3.7). Specifically, under the conditions of Theorem 3.3,

$$2v_T \left\{ \lambda_T(\omega)^{-1} \beta(\omega_j)^{\tau} f_T^{\mathrm{re}}(\omega) \beta(\omega) + \lambda_T(\omega) \beta(\omega_j)^{\tau} V_T f_T^{\mathrm{re}}(\omega)^{-1} V_T \beta(\omega) - 2 \right\} \tag{3.8}$$

converges in distribution to a distribution which is stochastically bounded between a χ^2_{k-2} and χ^2_{k-1} distribution.

Lastly, all the above results may be extended to the case where X_t is a vector process, with minor modification. In particular, we require that no two components of X_t are stochastically identical for all t, and that the product space $\mathscr{M} = \mathscr{G}_1 \times \cdots \times \mathscr{G}_k$ satisfy Condition A.

4. Applications

Example 1. Let X_t satisfy the bilinear model

$$X_t = \alpha \varepsilon_{t-1} X_{t-2} + \varepsilon_t,$$

where ε_t is an unobserved iid sequence with $E[\varepsilon_t] = 0$ and $E[\varepsilon_t^2] = \sigma^2$. This particular example is treated in detail in Granger and Andersen (1978) and permits a stationary solution when $\alpha^2 \sigma^2 < 1$. As is easily verified, X_t is an uncorrelated sequence and hence has all the appearance of a second-order white noise process. However, it can be shown that X_t^2 has the same autocorrelation structure as that of an identifiable ARMA(2,1) model which clearly distinguishes X_t from pure white noise.

It is instructive then to evaluate the spectral envelope procedure with respect to $\mathscr{G} = \{x, x^2\}$. From Granger and Andersen (1978) we note that X_t and X_t^2 are uncorrelated and thus,

$$V = \mathrm{Var}[(X_t, X_t^2)^\tau] = \mathrm{diag}(\sigma_x^2, \sigma_{xx}^2),$$

$$f(\omega) = f^{\mathrm{re}}(\omega) = (2\pi)^{-1}\mathrm{diag}(\sigma_x^2, \sigma_{xx}^2|k(\omega)|^2)$$

which exist provided $\alpha^4 E[\varepsilon_t^4] < 1$. Here $k(\omega)$ denotes the transfer function of X_t^2 which has the form of an identifiable ARMA(2,1) process as indicated above. From (2.6) it follows that the spectral envelope is given by

$$2\pi\lambda(\omega) = \max(1, |k(\omega)|^2)$$

which is a non-constant function of ω since $k(\omega) \not\equiv 1$ for all ω. This implies $|k(\omega)|^2 > 1$ over some non-empty subset of $(-\pi, \pi]$ so that a graph of $\lambda(\omega)$ would also distinguish X_t from pure white noise. In addition, the optimal scaling associated with $|k(\omega)|^2 > 1$ is $\beta(\omega) = (0,1)^\tau$ which identifies X_t^2 as the source of the correlation.

The above example provides a useful insight into the situation where the X_t are 'residuals' obtained from some modeling procedure. If the fitted model is appropriate then the residuals should exhibit properties similar to a white noise process. Although departures of the data from the fitted model may suggest model mis-specification, non-Gaussian data or the existence of a non-linear structure, the spectral envelope provides a simple diagnostic technique that would be useful for this type of residual analysis. This approach is pursued with a real data set in the next example.

Example 2. Here, we consider quarterly US GNP from 1947(1)–1991(1). The series was obtained from the Citibase database and is seasonally adjusted. After fitting an MA(2) model to the growth rate series, the residual ACF/PACF correlograms, Portmanteau test and spectrum estimate all indicated white noise residuals which we shall denote as X_t. Following the approach of Example 1, the absolute value and square transformations were applied to X_t and a spectral envelope estimate based on (3.4) with $m = 5$, was computed with respect to $\mathscr{G} = \{x, |x|, x^2\}$.

As can be seen from Fig. 1, there is a noticeable concentration of spectral power at the low-frequency end of the spectral envelope graph. Theorem 3.3 was applied to obtain an approximate 95% confidence interval about the first fundamental ($\omega_1 = 0.006$). The lower threshold of this interval was 0.027 which is about twice the baseline level. A possible explanation for the presence of this significant low-frequency power follows from Tiao and Tsay (1994), who investigated this growth rate series by fitting a nonlinear Threshold AR(2) model. Their results are interesting and indicate that the growth rate behaves differently during periods of contraction and expansion. Using a

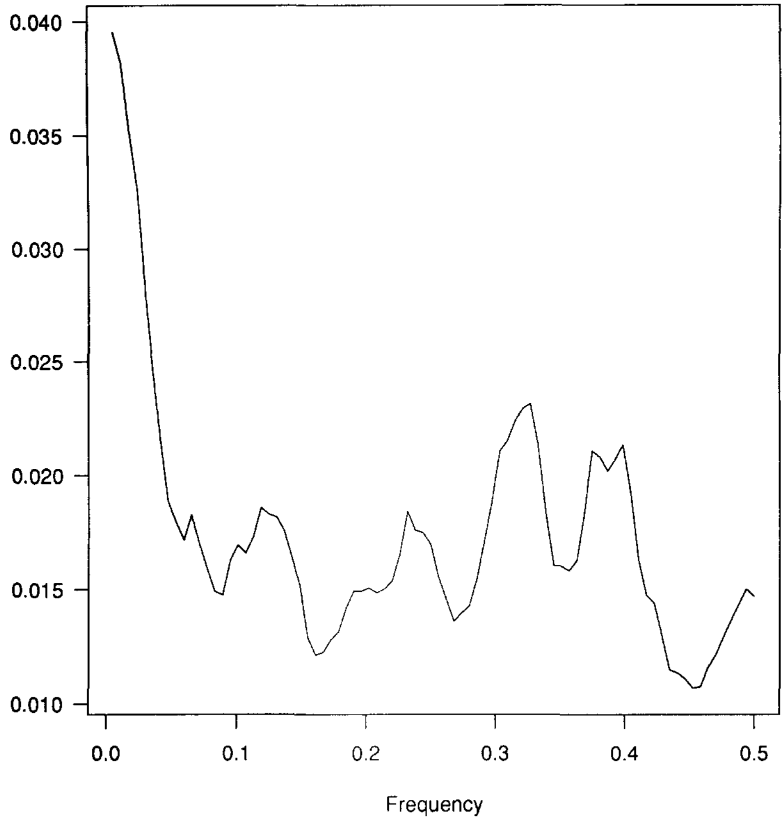

Fig. 1. Spectral envelope for GNP residuals $\mathscr{G} = (x, |x|, x^2)$, $T = 176$. Smoothed periodogram ($m = 5$) based estimate.

finer division of these regimes, the model was able to identify the 6 periods since 1947 that have been classified as economic recessions.

The presence of spectral power at very low frequency in detrended economic series is also interesting in terms of long-range dependence and our choice of transformations above, was partly influenced by the recent work of Ding et al. (1993), who applied transformations of the form $|X_t|^d$, $d \in (0,3]$, to the S&P 500 stock market series. In Fig. 2, the optimal transformation corresponding to ω_1 is shown. Here, it can be seen that while the low-frequency power is primarily due to the absolute value transformation, it has been modified by the square transformation. The square transformation also provides a dampening effect of two extreme residuals. These correspond to the values at $t = 12$ and $t = 133$ which were noted as possible outliers in Tiao and Tsay (1994) although their magnitudes are not substantial.

We conclude this example with an interesting footnote which concerns the deterministic drift component that was included in the MA(2) fitted model. Given that the original data was differenced, a zero mean might have been expected. But if the drift term is ignored, the ACFs for *all* the individual transformations in $\mathscr{G}$ indicate white

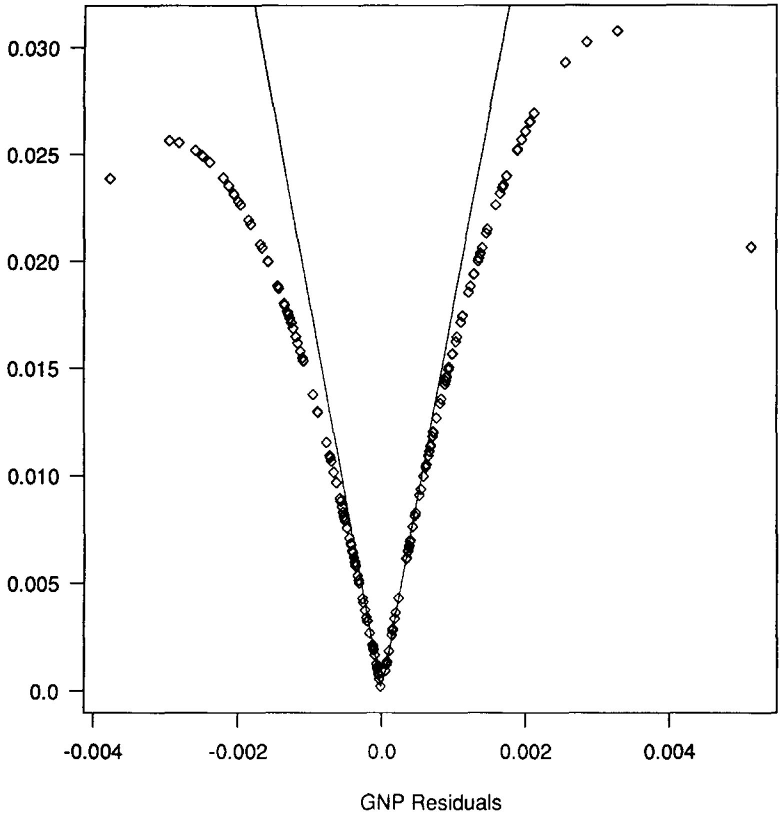

Fig. 2. Optimal transformation for GNP residuals $\omega = 0.006$. $\beta(0.006) = (0.8, 18.0, -2872.5)^{\tau}$. (Solid line) $18 * |x|$ (Spectral envelope estimate as in Fig. 1.).

noise and only the spectrum for the absolute transformation provides some indication of a departure from white noise. In this case, evidence of low-frequency power still exists in the spectral envelope, but a larger proportion of spectral power is also allocated over the frequency range 0.3–0.4.

Example 3. Let X_t be generated by the non-linear model

$$X_t = \exp\{\alpha \sin(2\pi\omega_0 t + \phi) + \varepsilon_t\},$$

where ϕ is a random phase, $\alpha > 0$ and ω_0 are constants, and ε_t is Gaussian white noise with variance σ^2. This example is adapted from Breiman and Friedman (1985) where the ACE procedure is discussed and, hence, motivates us to consider the result of applying the spectral envelope when the basis for $\mathscr{G}$ does *not* include the log transformation.

In the results presented below, realizations of this model were generated using the parameter values, $\alpha = 3$, $\sigma = 4$ and $\omega_0 = 50/T$, with $T = 512$ as the sample length of the simulation. As might be expected, these parameter values can produce realizations

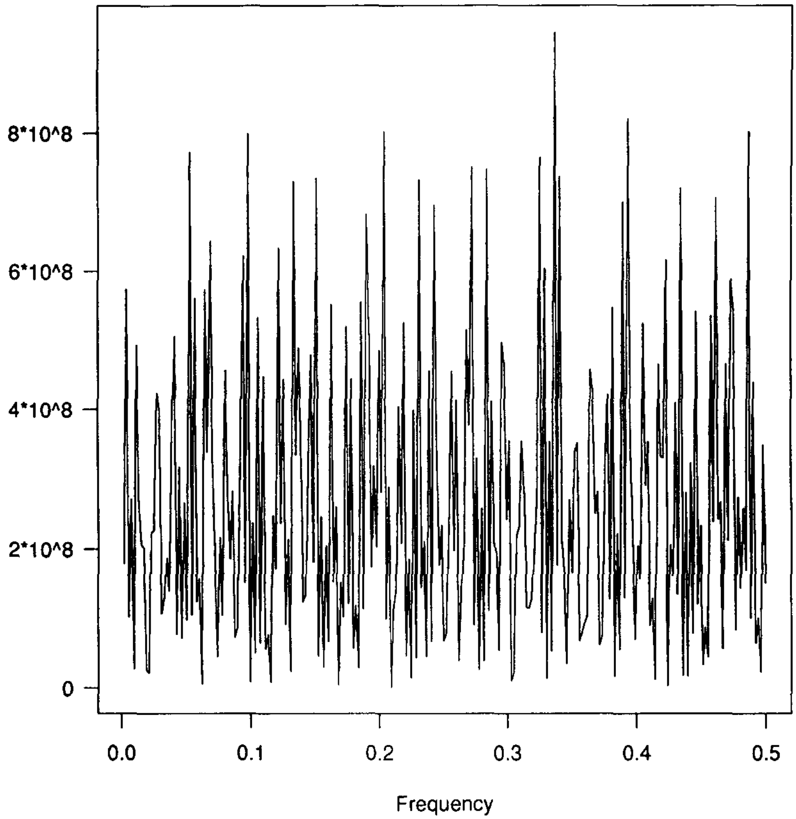

Fig. 3. Periodogram for simulated model (Example 3) $X_t = \exp(\alpha \sin(2\pi\omega_0 t + \phi) + \varepsilon_t)$. $T = 512$, $\alpha = 3$, $\omega_0 = 50/T$, $\sigma = 4$.

which sometimes contain extremely large (600,000+) observations. The periodogram for a typical realization is given in Fig. 3 and provides no evidence of any dominant frequency, including $\omega = \omega_0$ which is close to 0.1.

In contrast, the sample spectral envelope computed with respect to $\mathscr{G} = \{x, \sqrt{x}, \sqrt[3]{x}\}$ has little difficulty in isolating the frequency ω_0 as evident by the large peak in Fig. 4. Under the assumption that this process is white noise, the significance of a single a priori specified frequency can be assessed using Theorem 3.2 to approximate $P[2\pi\lambda_T(\omega) < Tx/2]$ for $0 \leqslant x \leqslant 1$. Note that $4\pi T^{-1}\lambda_T(\omega)$ can be interpreted as the largest proportion of total power at frequency ω that can be obtained for any transformation in the class $\mathscr{G}$. In Fig. 4, the approximate 0.0001 ($x = 4.84\%$) null significance threshold is plotted.

The scaling corresponding to $\lambda_T(\omega_0)$ is $\beta_T(\omega_0) = (1.52e-7, -3.11e-04, 1.9e-3)^{\tau}$. Although not shown here, the shape of the optimal transformation generated by $\beta_T(\omega_0)$ is very close to a $\log(x)$ curve over the domain $x < 4000$ (which excludes 13 extreme values). A final note also worth mentioning is the result we obtained when the rather inappropriate basis $\{x, x^2, x^3\}$ was employed. Surprisingly, the spectral envelope pro-

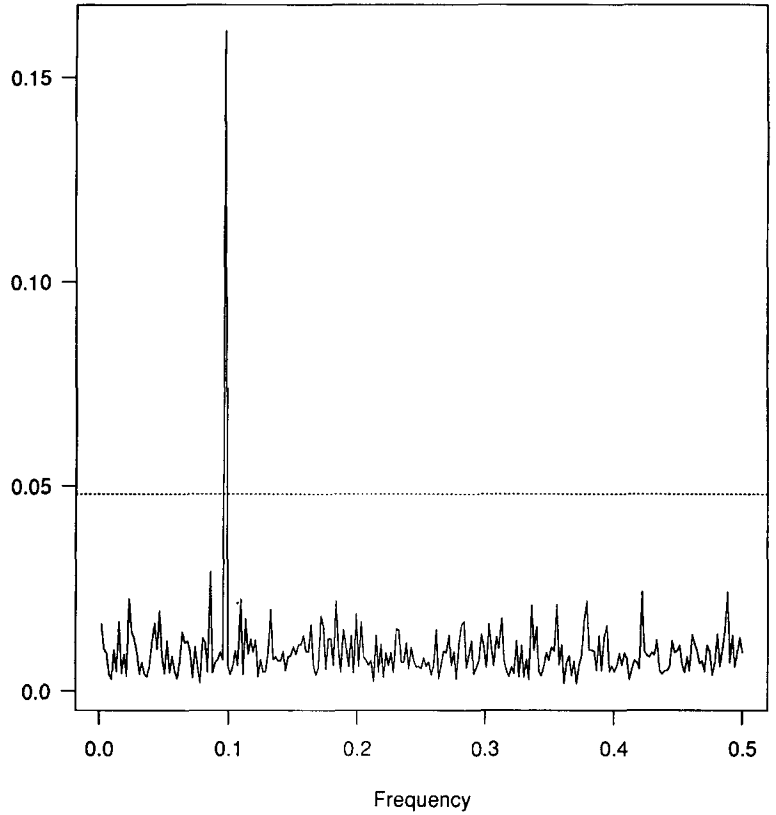

Fig. 4. Spectral envelope for simulated model (Example 3). $X_t = \exp(\alpha \sin(2\pi\omega_0 t + \phi) + \varepsilon_t)$. $T = 512$, $\alpha = 3$, $\omega_0 = 50/T$, $\sigma = 4$. $\mathscr{G} = (x, \sqrt{x}, \sqrt[3]{x})$. Periodogram based estimate. (— — ——) approximate 0.0001 null threshold.

cedure often managed to assign a noticeably greater proportion of total power to the frequencies surrounding ω_0 due to the x^3 component.

Example 4. In Bloomfield's (1976) exploratory analysis of the Sunspot series, the square root transformation was applied and compared to the untransformed spectrum. Bloomfield's results suggest some benefit is gained by using a transformation, although the evidence is not overwhelming. (This can be mainly attributed to the asymmetry of the sunspot cycles; they typically rise more rapidly than they fall.) With the spectral envelope procedure however, an objective assessment can be made with respect to an entire class of transformations.

Fig. 5 shows the result obtained from applying the spectral envelope procedure to the semi-annual aggregated series with respect to $\mathscr{G} = \{x, \sqrt{x}, \log(x+1)\}$. To facilitate examination of Fig. 5, we have plotted the sample spectral envelope on a log scale and included the normalized log periodogram of X_t which is represented by the broken line. Of interest is the presence of a peak near frequency 0.08 which corresponds to an

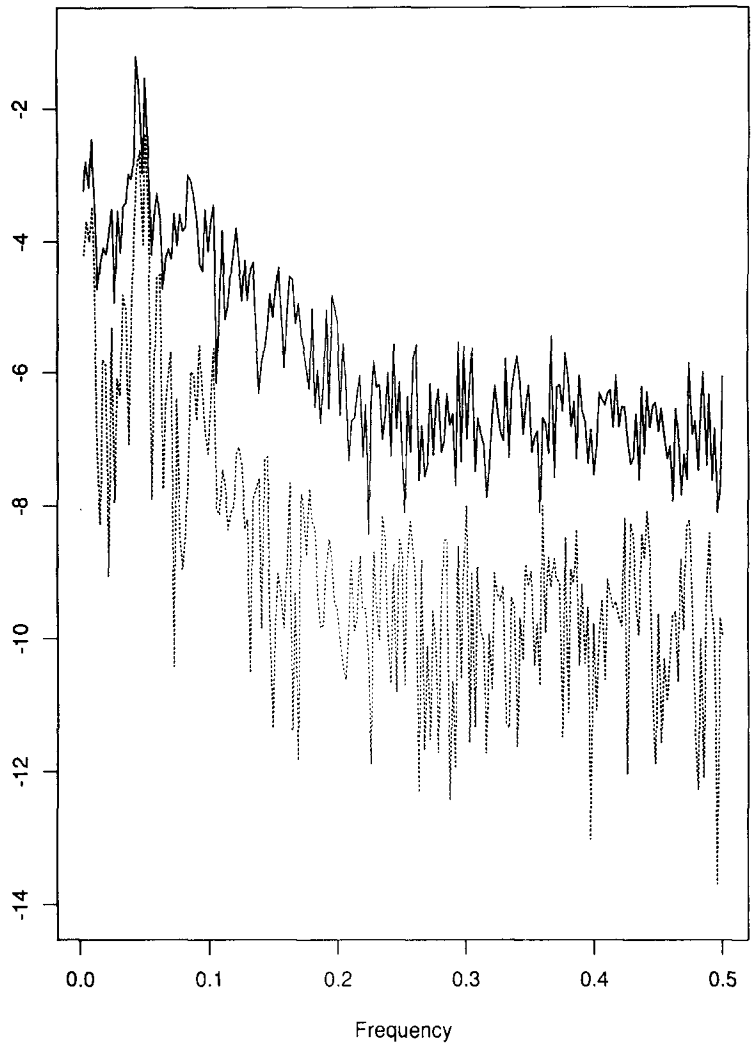

Fig. 5. Log spectral envelope for sunspot series $T = 456$, semi-annual aggregates. $\mathscr{G} = (x, \sqrt{x}, \log(x+1))$. Periodogram based estimate. (– – – ––) log periodogram of sunspot series.

approximate $5\frac{1}{2}$ year cycle. This particular harmonic is discussed in Shumway (1988) and is shown to be statistically significant using standard spectral diagnostics.

In Fig. 6 we compare the effect of the transformations corresponding to the 11 and $5\frac{1}{2}$-year cycles. (To assist the visual comparison, the series has been rescaled and shifted so the y-axis graduations are not relevant here.) From a careful examination of these two series, it can be seen that the effect of the $5\frac{1}{2}$ year transformation is to essentially "fold" the 11 year harmonic series (solid line) upon itself horizontally. This, of course, would produce a second harmonic with double the frequency which is evidently the case here. Shumway (1988) also discusses the presence of higher-order

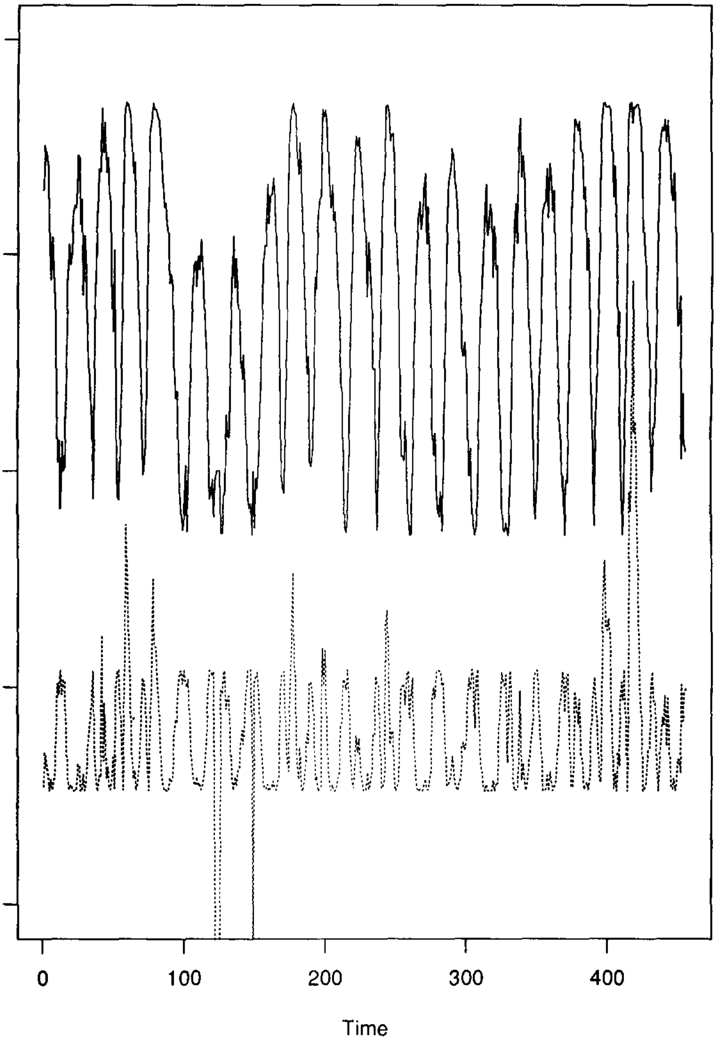

Fig. 6. Transformed series associated with sunspot series $T = 456$, $\mathscr{G} = (x, \sqrt{x}, \log(x+1))$. Periodogram based estimate. (Solid line) 11 year transformed series, $\beta = (0, -0.59, 1)$; (Broken line) $5\frac{1}{2}$ year transformed series, $\beta = (0, -0.42, 1)$ (Values rounded to 2 dp's).

harmonics in this data which could be investigated in similar fashion using the spectral envelope.

Example 5. In our final example we apply the spectral envelope procedure to a multivariate time series. For illustrative purposes we have used a data set that would be readily available to the interested reader. Thus, we take the S data set *bonds.yield* which consists of Daily Yields from six AT&T Bonds during the period April 1975 to December 1975 (Becker et al., 1988). Each of these series appear to exhibit a slow cycle rather than any obvious trend, and it is of interest whether *all* of the component

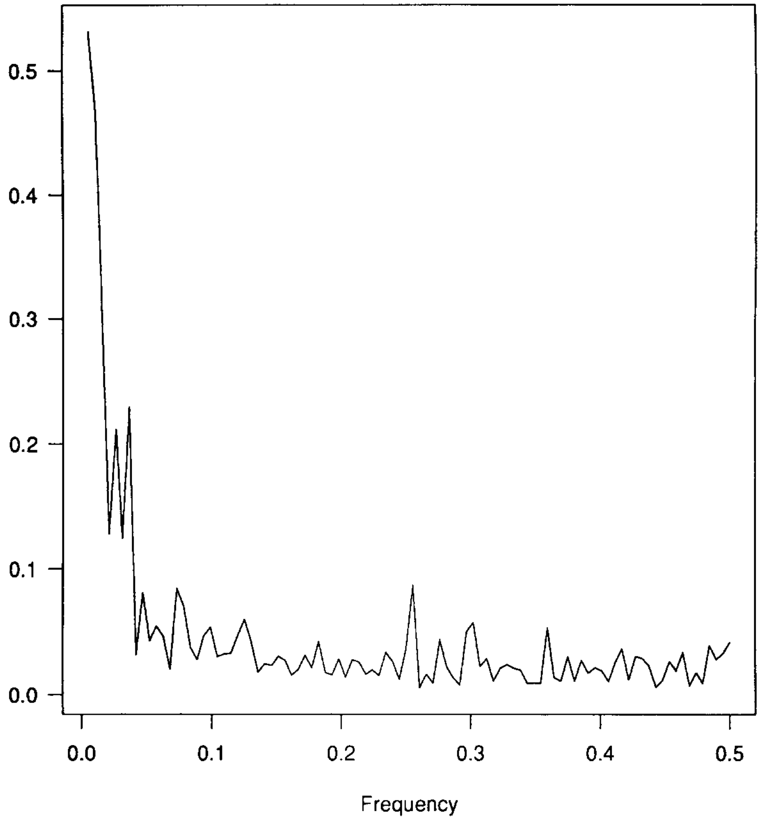

Fig. 7. Spectral envelope for S dataset: *bonds.yield* $T = 192$, Daily bond yields for 6 AT&T bonds. Multivariate periodogram based estimate. $\mathscr{G} = \{X_{1t}, \ldots, X_{6t}\}$.

series contribute to this periodicity. If not, then we should expect to see some of the $\beta(\omega)$ coefficients close to zero, and similarly at other frequencies of interest. Note that asymptotic tests for the coefficient vector can be obtained via (3.8).

From the sample spectral envelope shown in Fig. 7, we can identify two particular frequencies that seem to be of interest; $\omega_1 = 1/192$ or $2/192$, and possibly $\omega_2 = 0.25$. The presence of a large first fundamental is worth investing further since this can be caused by a simple trend rather than a sinusoid. Closer inspection of the individual series and their spectra (which all have a peak at frequency $2/192$) would seem to suggest a possible slight trend (all the series drop rapidly towards the end), in conjunction with a strong cycle that has a frequency somewhere between the first two fundamentals.

The sample scalings for the first two fundamental frequencies are, $\beta_T(1/192) = (-0.77, 0.03, -0.41, -0.07, 0.09, 1.00)^{\tau}$ and $\beta_T(2/192) = (0.03, 1.00, -0.63, -0.22, 0.74, 0.35)^{\tau}$. Since the variances associated with individual time series can be regarded as equal, we have scaled these vectors relative to their largest (absolute) component for clarity. Despite the apparent differences in the relative weightings of the components of X_t, both scalings produce transformed series with very similiar spectral properties. This inconsistency is resolved by employing a smoothed spectrum estimate to compute the β_T scalings. Using (3.4) with $m = 5$, we obtain $\beta_T(1/192) = (-0.96, -0.57, -0.22, -0.02, 0.16, 1.00)^\tau$ corresponding to $\lambda_T(1/192) = 0.23$, with similar weightings at the next fundamental. The low-frequency behavior is now seen to be dominated by the difference between X_{1t} and X_{6t}, with X_{2t} contributing as well. Applying (3.8) to get an asymptotic test for $\beta(1/192) = (-2, -1, 0, 0, 0, 2)^\tau$ we obtained an asymptotic p-value between 0.28 and 0.49. Omitting the X_{2t} component to see if a simple difference would suffice however, resulted in p-values that were less than 0.001. The scaling corresponding to the ω_2 frequency, suggests this four-day cycle is governed by the difference between X_{5t} and X_{6t}. However, its associated λ_T value (which was rather marginal to begin with) becomes negligible when smoothing is applied.

5. Discussion

We have presented a general method for analyzing real-valued time series in the spectral domain which provides a useful complement to existing spectral diagnostic methods. The examples presented in Section 4 have shown that the spectral envelope procedure can be readily applied in variety of situations and simultaneously incorporates a range of results that can be used for interpretation. This article is not intended to be an exhaustive treatment of the topic of course, and we conclude with a brief discussion of some problems for further research.

The transformations we have considered so far have been restricted to univariate functions of X_t, or in the vector case, X_{jt}. In Section 2 we indicated that the components of a vector could also be constructed from a univariate series in order to accommodate non-instantaneous transformations. More generally, multivariate functions such as $g_j(X_t, X_{t-1})$ could just as easily be employed in the spectral envelope procedure. If X_t is strictly stationary then, under suitable regularity conditions on the class $\mathscr{G}$, the results of Section 3 remain valid except for Theorem 3.2. This follows from Doob (1953) since any measurable function defined on the σ-field of X_t that is unitary with respect to the shift operator, will also be strictly stationary. An obvious problem is that the interpretation of the spectral envelope relative to X_t then becomes more complicated since we no longer have Theorem 3.2 as a 'point of reference'. A simple example which can be interpreted is the case where $\lambda(\omega)$ optimizes the transformation $g(X_t, X_{t-1}) = X_t + \alpha X_{t-1}$ over α, where X_t is white noise. This is equivalent to setting $Y_t = (X_t, X_{t-1})^\tau$ and rescaling the first component of $\beta(\omega)$ to 1. By evaluating $f^{\mathrm{re}}(\omega)$ we obtain $\lambda(\omega) = 1 \pm \cos\omega$ and $\beta(\omega) = (1, \pm 1)^\tau$, with the minus case applying over $\omega > \pi/2$. Thus, in this simple case of a MA(1) model, we see that the spectral envelope provides the correct result of unit-root filters that maximize the spectral power of g at $\omega = 0$ or π.

The examples presented in this article have demonstrated some *exploratory* applications of the spectral envelope methodology where the choice of $\mathscr{G}$ was primarily motivated by practical experience. As Example 3 suggests however, a semiparametric approach could also be employed to generate $\mathscr{G}$ from some suitable parent class, $\mathscr{H}$ say. Although further investigation is needed to more fully assess the finite sample properties of this type of approximation procedure, asymptotic considerations lead to the following research problem which is also of theoretical interest. By defining the Hilbert space

$$\mathscr{H} = \{g : E[g(X_t)] = 0,\ E[g(X_t)^2] < \infty\}$$

with inner product $\langle g_1, g_2 \rangle = E[g_1(X_t)g_2(X_t)]$, the spectral envelope can be defined in terms of a linear operator on $\mathscr{G} \subset \mathscr{H}$. Thus, under suitable regularity conditions, a meaningful interpretation of the limit can be given in the situation where $\dim(\mathscr{G}) \to \infty$ as $T \to \infty$, which describes the limiting behavior associated with an approximation procedure.

Finally, it is also possible to extend the notion of the spectral envelope to *multidimensional* processes by defining it in terms of the *wave number spectrum* $f(\omega)$ where $\omega = (\omega_1, \ldots, \omega_d)^\tau$. In geostatical applications for example, a number of attributes, X_{js}, $j = 1, \ldots, k$ say, are measured at each location $s \in \mathscr{D} \subset \mathbb{R}^d$. Here, the spatial correlation of the process is of interest, and when $k = 1$, the *Covariogram* provides an effective diagnostic tool. However, when $k > 1$, cross-covariograms need to be considered and since these are more difficult to model, linear combinations of the attributes are often employed to reduce the dimensionality or allow a covariogram analysis to be performed. (For details on this subject we refer the reader to Cressie (1993).) If $\mathscr{D}$ is a regular grid then the spectral envelope clearly enables an *optimal* linear combination to be obtained at each ω which maximizes $f(\omega)$ with respect to the total power. When $d = 2$, this information can be presented in terms of a contour plot with high density regions indicating strong wave number association. Whether such a representation would be useful in geostatistics applications remains to be seen.

Acknowledgements

The work of Andrew McDougall was supported in part by a grant from the National Science Foundation. The work of David Stoffer was supported in part by a grant from the National Science Foundation.

References

Becker, R., J. Chambers and A. Wilks (1988). *The New S Language – A Programming Environment for Data Analysis and Graphics.* Wadsworth and Brooks/Cole, Pacific Grove, CA.

Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction.* Wiley, New York.

Breiman, L. and J.H. Friedman (1985). Estimating optimal transformations for multiple regression and correlation. *JASA* **80**, 580–619.

Brillinger, D.R. (1981). *Time Series: Data Analysis and Theory*, 2nd ed. Holden-Day, San Francisco.

Cressie, N. (1993). *Statistics for Spatial Data.* Wiley, New York.

Ding, Z., C.W.J. Granger and R.F. Engle (1993). A long memory property of stock market returns and a new model. *J. Empirical Finance* **1**, 83–106.

Doob, J.L. (1953). *Stochastic Processes.* Wiley, New York.

Granger, C.W.J. and A.P. Andersen (1978). *An Introduction to Bilinear Time Series Models.* Vandenhock and Ruprecht, Göttingen.

Hannan, E.J. (1970). *Multiple Time Series.* Wiley, New York.

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

Rosenblatt, M. (1959). Statistical analysis of stochastic processes with stationary residuals. In: U. Grenander, Ed., *Probability and Statistics*, Wiley, New York, pp. 246–275.

Shumway, R.H. (1988). *Applied Time Series Analysis.* Prentice-Hall, Englewood Cliffs, NJ.

Stoffer, D.S., D.E. Tyler and A.J. McDougall (1993). Spectral analysis for categorical time series: scaling and the spectral envelope. *Biometrika* **80**, 611–622.

Stoffer, D.S., D.E. Tyler, A.J. McDougall and G. Schachtel (1993). Spectral analysis of DNA sequences. *Bull. Internat. Statist. Inst.* (to appear).

Tiao, G.C. and R.S. Tsay (1994). Some advances in nonlinear and adaptive modeling in time series analysis. *J. Forecasting* **13**, 109–131.

Tiao, G.C., R.S. Tsay and T. Wang (1993). Usefulness of linear transformations in multivariate time series analysis. *Tech. Report.*, Statistics Research Center, Graduate School of Business, University of Chicago.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford Univ. Press, Oxford.