The Exponential Model for the Spectrum of a Time Series: Extensions and Applications

Tommaso Proietti
University of Rome Tor Vergata

Alessandra Luati
University of Bologna

Abstract. The exponential model for the spectrum of a time series and its fractional extensions are based on the Fourier series expansion of the logarithm of the spectral density. The coefficients of the expansion form the cepstrum of the time series. After deriving the cepstrum of important classes of time series processes, also featuring long memory, we discuss likelihood inferences based on the periodogram, for which the estimation of the cepstrum yields a generalized linear model for exponential data with logarithmic link, focusing on the issue of separating the contribution of the long memory component to the log-spectrum. We then propose two extensions. The first deals with replacing the logarithmic link with a more general Box-Cox link, which encompasses also the identity and the inverse links: this enables nesting alternative spectral estimation methods (autoregressive, exponential, etc.) under the same likelihood-based framework. Secondly, we propose a gradient boosting algorithm for the estimation of the log-spectrum and illustrate its potential for distilling the long memory component of the log-spectrum.

Key words and phrases: Frequency Domain Methods; Generalized linear models; Long Memory; Boosting.

1 Introduction

The analysis of stationary processes in the frequency domain has a long tradition in time series analysis; the spectral density provides the decomposition of the total variation of the process into the contribution of periodic components with different frequency as well as a complete characterization of the serial correlation structure of the process, so that it contains all the information needed for prediction and interpolation. Inferences on the spectrum are based on the periodogram, which possesses a well established large sample distribution theory that leads to convenient likelihood based estimation and testing methods.

This paper is concerned with a class of generalized linear models formulated for the logarithm of the spectral density of a time series, known as the exponential (EXP) model, which emerges by truncating the Fourier series expansion of the log-spectrum. Modelling the log-spectrum (rather than the spectrum) is most amenable for processes whose spectral densities can be factorized as products of spectra of different components, as it is the case for processes with long range dependence, for which the long memory and the short memory components enter multiplicatively. The coefficients of the expansion are known as the cepstral coefficients and are in turn obtained from the discrete Fourier transform of the log-spectrum; their collection form the cepstrum. This terminology was introduced by Bogert, Healey and Tuckey (1963), cepstral and cepstrum being anagrams of spectral and spectrum, respectively.

The Fourier transform of the logarithm of the spectral density function plays an important role in the analysis of stationary stochastic processes. It is the key element of the spectral factorization at the basis of prediction theory, leading to the Kolmogorov-Szegö formula for the prediction error variance (see Doob, 1953, theorem 6.2, Grenander and Rosenblatt, 1957, section 2.2, and Pourahmadi, 2001, Theorem VII). Bogert, Healey and Tuckey (1963) advocated its use for the analysis of series that are contaminated by echoes, namely seismological data, whose spectral densities typically factorize as the product of two components, one of which is the the contribution of the echo. We refer the reader to Oppenheimen and Schafer (2010 ch. 13), Brillinger (2001) and Childers, Skinner and Kemerait (1977) for historical reviews on the cepstrum and its applications in signal processing.

Bloomfield (1973) introduced the exponential (EXP) model and discussed its maximum likelihood estimation, relying on the distributional properties of the periodogram, based on Whittle (1953) and Walker (1964). As illustrated by Cameron and Turner (1987), maximum likelihood estimation is computationally very attractive, being carried out by iteratively reweighted least squares.

Local likelihood methods with logarithmic link for spectral estimation have been considered by Fan and Kreutzberg (1998). Also, the exponential model has played an important role in the Bayesian estimation of the spectrum (Wahba, 1980; Carter and Kohn, 1997), for regularized spectral estimation, where smoothness priors are enforced by shrinking higher order cepstral coefficients toward zero, and has been recently considered in the estimation of time-varying spectra (Rosen, Stoffer and Wood, 2009, and Rosen, Wood

and Stoffer, 2012). Among other uses of the EXP model we mention discrimination and clustering of time series, as in Fokianos and Savvides (2008).

The model was then generalized to processes featuring long range dependence by Robinson (1991) and Beran (1993), originating the fractional EXP model (FEXP), whereas Janacek (1982) proposed a method of moments estimator of the long memory parameter based on the sample cepstral coefficients estimated nonparametrically using the log-periodogram. Maximum likelihood estimation of the FEXP model has been dealt with recently by Narukawa and Matsuda (2011). Hurvich (2002) addresses the issue of predicting with it. The sampling distribution of the periodogram is also at the basis of log-periodogram regression, that is widely applied for long memory estimation (Geweke and Porter-Hudak, 1983; Robinson, 1995; Moulines and Soulier, 1999; Andrews and Guggenberger, 2003; Hsu and Tsai, 2009).

Against this background, the paper contributes to the current literature in the following way.

- After deriving the cepstrum of important time series models, we illustrate the potential and the limitations of cepstral analysis for long memory time series. In particular, we focus on the issue of separating the long memory component from the short one; the sampling distribution of the maximum likelihood estimators of the parameters points at an inherent difficulty in disentangling the contribution of the long memory component from the log-spectrum. This is the topic of sections 2-4.

- We introduce the class of generalized linear cepstral models with Box-Cox link, according to which a linear model is formulated for the Box-Cox transformation of the spectral density. The link function depends on a power transformation parameter, and encompasses the exponential model, which corresponds to the case when the transformation parameter is equal to zero. Other important special cases are the inverse link (which leads to modelling the inverse spectrum), and the identity link. The coefficients of the model are related to the generalized autocovariances, see Proietti and Luati (2012), and are termed generalized cepstral coefficients. To enforce the constraints needed to guarantee the positivity of the spectral density, we offer a reparameterization of the generalized cepstral coefficients and we show that our framework is able to nest alternative spectral estimation methods, in addition to the exponential approach, namely autoregressive spectral estimation (inverse link) and moving average estimation (identity link), so that the appropriate method can be selected in a likelihood based framework. We also discuss testing for white noise in this framework. This is the content of section 5.

- We introduce a boosting algorithm for variable selection and spectral estimation with a fractional exponential model and illustrate its great potential for separating long memory from short (section 6). The procedure offers clear advantages over the selection of the order of the truncation of the FEXP model, due to the regularization properties of the boosting algorithm.

3

These points are illustrated in section 7 by three case studies and a Monte Carlo simulation experiment. Finally, in section 8 we offer some conclusions.

2   The Exponential Model and Cepstral Analysis

Let $\{y_t\}_{t \in T}$ be a stationary zero-mean stochastic process indexed by a discrete time set T, with covariance function $\gamma_k = \int_{-\pi}^{\pi} e^{\imath \omega k} dF(\omega)$, where $F(\omega)$ is the spectral distribution function of the process and $\imath$ is the imaginary unit. We assume that the spectral density function of the process exists, $F(\omega) = \int_{-\pi}^{\omega} f(\lambda) d\lambda$, and that the process is regular (Doob, 1953, p. 564), i.e. $\int_{-\pi}^{\pi} \ln f(\omega) d\omega > -\infty$.

As $f(\omega)$ is a positive, smooth, even and periodic function of the frequency, its logarithm can be expanded in a Fourier series as follows,

$$\ln[2\pi f(\omega)] = c_0 + 2 \sum_{k=1}^{\infty} c_k \cos k\omega, \tag{1}$$

where the coefficients c_k, $k = 0, 1, \ldots$, are obtained by the (inverse) Fourier transform of $\ln 2\pi f(\omega)$,

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln[2\pi f(\omega)] \exp(\imath \omega k) d\omega.$$

The coefficients c_k are known as the *cepstral coefficients* and the sequence $\{c_k\}_{k=0,1,\ldots}$ is known as the *cepstrum* (Bogert, Healy and Tukey, 1963). The interpretation of the cepstral coefficients as pseudo-autocovariances is also discussed in Bogert, Healy and Tukey (1963) and essentially follows from the analogy with the Fourier pair $2\pi f(\omega) = \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(k\omega)$ and $\gamma_k = \int_{-\pi}^{\pi} f(\omega) \exp(\imath \omega k) d\omega$.

Important characteristics of the underlying process can be obtained from the cepstral coefficients. The intercept is related to the the one-step ahead prediction error variance (p.e.v.), $\sigma^2 = \text{Var}(y_t | \mathcal{F}_{t-1})$, where $\mathcal{F}_t$ is the information up to time t: by the Szegö-Kolmogorov formula,

$$\sigma^2 = \exp\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln[2\pi f(\omega)] d\omega\right]$$

we get immediately that $c_0 = \ln \sigma^2$. Moreover, the long run variance is obtained as

$$2\pi f(0) = \exp\left(c_0 + 2 \sum_{k=1}^{\infty} c_k\right).$$

Also, if we let $y_t = \varphi(B)\xi_t$ denote the Wold representation of the process, with $\varphi(B) = 1 + \varphi_1 B + \varphi_2 B^2 + \ldots, \sum_j |\varphi_j| < \infty, \xi_t \sim \text{WN}(0, \sigma^2)$, where B is the lag operator, $B^j y_t = y_{t-j}$, then the moving average coefficients of the Wold representation are obtained recursively from the formula

$$\varphi_j = j^{-1} \sum_{r=1}^{j} r c_r \varphi_{j-r}, \quad j = 1, 2, \ldots, \tag{2}$$

with $\varphi_0 = 1$. The derivation, see Janacek (1982), Pourahmadi (1983) and Hurvich (2002), is based on the spectral factorization $2\pi f(\omega) = \sigma^2 \varphi(e^{-\iota\omega}) \varphi(e^{\iota\omega})$; setting $\varphi(z) = \exp\left(\sum_{k=1}^{\infty} c_k z^k\right)$, and equating the derivatives of both sides with respect to z at the origin, enables to express the Wold coefficients in terms of the cepstral coefficients, giving (2). The autoregressive representation $\pi(B)y_t = \xi_t$, where $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j = \varphi(B)^{-1}$, is easily determined from the relationship $\ln\pi(z) = -\ln\varphi(z)$, and it is such that $\pi_0 = 1$ and $\pi_j = -j^{-1}\sum_{r=1}^{j} rc_r\pi_{j-r}, \quad j = 1, 2, \ldots$.

The mutual information between the past and the future of a Gaussian time series is defined in terms of the cepstral coefficients by Li (2005), $I_{p-f} = \frac{1}{2}\sum_{k=1}^{\infty} kc_k^2$, provided that $\sum_{k=-\infty}^{\infty} kc_k^2 < \infty$, and the following relation hold between cepstral coefficients and the partial autocorrelation coefficients, $\{\phi_{kk}\}_{k=1,2,\ldots}$, the so called reflectrum identity, $\sum_{k=1}^{\infty} kc_k^2 = -\sum_{k=1}^{\infty} k\ln(1 - \phi_{kk}^2)$ and $c_0 = \ln\gamma_0 + \sum_{k=1}^{\infty} \ln(1 - \phi_{kk}^2)$, the latter being a consequence of the Kolmogorov-Szegö formula.

We also note that the Fourier expansion (1) is equivalent to express the logarithm of the spectral density function as $\ln[2\pi f(\omega)] = c_0 + s(\omega)$ where $s(\omega)$ is a linear spline function, $s(\omega) = \int_0^{\omega} B(z)\mathrm{d}z$, and $B(z)$ is a Wiener process, when the canonical representation for the spline basis functions is chosen, i.e. via the Demmler-Reinsch basis functions (Demmler and Reinsch, 1975, see also Eubank, 1999). This representation is applied in a Bayesian setting by Rose, Wood and Stoffer (2012) and Rosen, Stoffer and Wood (2009).

2.1 Cepstral Analysis of ARMA Processes

If y_t is a white noise (WN) process, $c_k = 0, k > 0$. Figure 1 displays the cepstrum of the AR(1) process $y_t = \phi y_{t-1} + \xi_t, \xi_t \sim \text{WN}(0, \sigma^2)$ with $\phi = 0.9$ and coefficients $c_k = \phi^k/k$ (top left plot). The behaviour of the cepstrum is analogous to that of the autocovariance function, although it will dampen more quickly due to the presence of the factor k^{-1}. The upper right plot is the cepstrum of the MA(1) process $y_t = \xi_t + \theta\xi_{t-1}$, with $\theta = 0.9$; the general expression is $c_k = -(-\theta)^k/k$. Notice that if $c_k, k > 1$, are the cepstral coefficients of an AR model, that of an MA model of the same order with parameters $\theta_j = -\phi_j$ are $-c_k$. Hence, for instance, the cepstral coefficients of an MA(1) process with coefficient $\theta = -.9$ are obtained by reversing the first plot. The bottom plots concern the cepstra of two pseudo-cyclical processes: the AR(2) process $y_t = 1.25y_{t-1} - 0.95y_{t-2} + \xi_t$ with complex roots, and the ARMA(2,1) process $y_t = 1.75y_{t-1} - 0.95y_{t-2} + \xi_t + 0.5\xi_{t-1}$. The cepstra behave like a damped sinusoidal, and again the damping is more pronounced than it shows in the autocovariance function. Notice also that even for finite p and q we need infinite coefficients c_k to represent an ARMA model.

For a general ARMA process $y_t \sim \text{ARMA}(p, q)$, $\phi(B)y_t = \theta(B)\xi_t$, with AR and MA polynomials

Figure 1: Cepstral coefficients $c_k, k = 1, \ldots, 20$ for selected ARMA models
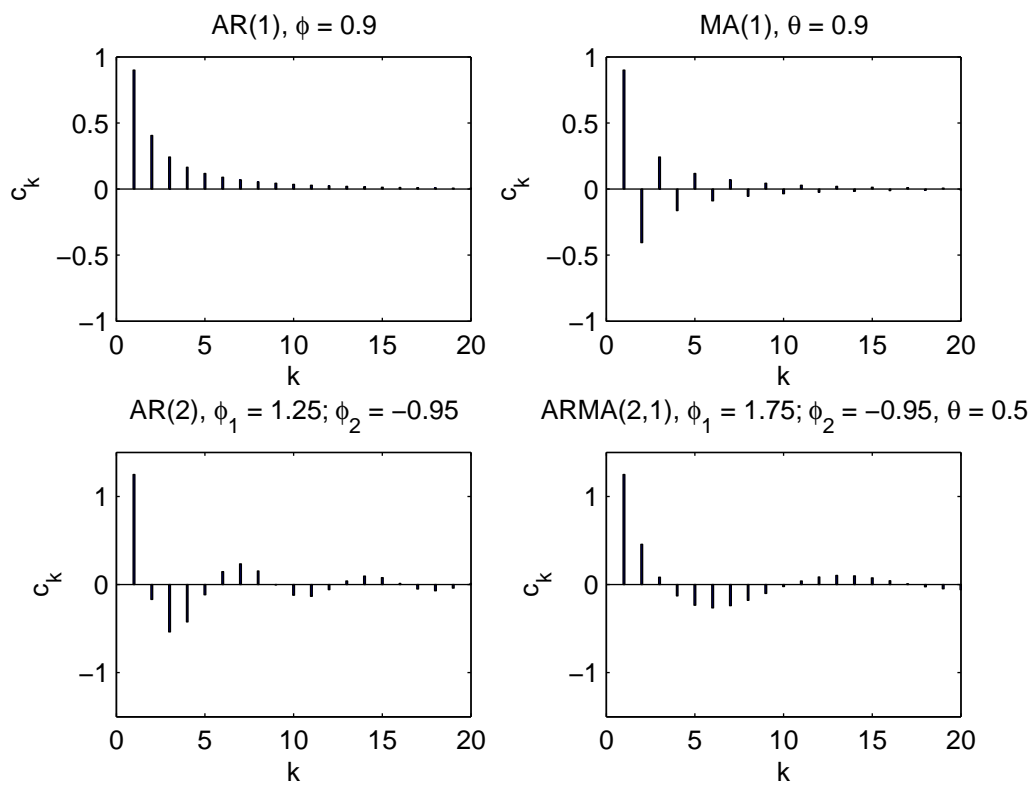
factorized in terms of their roots, as in Brockwell and Davis, 1991, section 4.4,

$$\phi(B) = \prod_{j=1}^{p}(1 - a_j^{-1}B), \quad \theta(B) = \prod_{j=1}^{q}(1 - b_j^{-1}B), |a_j| > 1, |b_j| > 1,$$

we have the following general result

$$\ln[2\pi f(\omega)] = c_0 + 2\sum_{k=1}^{\infty}\left(\sum_{j=1}^{q}c_{jk}^{(b)} - \sum_{j=1}^{p}c_{jk}^{(a)}\right)\cos(k\omega),$$

where
$$c_{jk}^{(a)} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\ln|1 - a_j^{-1}e^{-\imath\omega}|^2\cos(\omega k)\mathrm{d}\omega, \quad c_{jk}^{(b)} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\ln|1 - b_j^{-1}e^{-\imath\omega}|^2\cos(\omega k)\mathrm{d}\omega.$$

This is the sum of elementary cepstral processes corresponding to polynomial factors. When a_j and b_j are real $c_{jk}^{(a)} = -a_j^{-k}/k$ and $c_{jk}^{(b)} = -b_j^{-k}/k$ (see Gradshteyn and Ryzhik , 1994, 4.397.6).

When there are two complex conjugate roots, $a_j = r^{-1}e^{\imath\varpi}, \bar{a}_j = r^{-1}e^{-\imath\varpi}$, with modulus $1/r$ and phase ϖ, their contribution to the cepstrum is via the coefficients $r^k\cos(\varpi k)/k$. Hence, the cepstral coefficients of the stationary cyclical process $(1 - 2r\cos\varpi B + r^2 B^2)y_t = \xi_t$ are $c_k = r^k\cos(\varpi k)/k, k = 1, 2, \ldots$; see the bottom left plot of figure 1, for which $r = 0.97$ and $\varpi = 0.87$: the cepstral coefficient have a period of about 7 units.

2.2   Truncated Cepstral Processes

The class of stochastic processes characterised by an exponential spectrum was proposed by Bloomfield (1973), who suggested truncating the Fourier representation of $\ln 2\pi f(\omega)$ to the K term (EXP(K) process), so as to obtain:

$$\ln[2\pi f(\omega)] = c_0 + 2\sum_{k=1}^{K}c_k\cos(\omega k). \tag{3}$$

The EXP(1) process, characterized by the spectral density $f(\omega) = (2\pi)^{-1}\exp(c_0 + 2c_1\cos\omega)$, has autocovariance function

$$\gamma_k = \sigma^2 I_k(2c_1) = \sigma^2\sum_{j=0}^{\infty}\frac{c_1^{2j+k}}{(k+j)!j!},$$

where $I_k(2c_1)$ is the modified Bessel function of the first kind of order k, see Abramowitz and Stegun, (1972), 9.6.10 and 9.6.19, and $\sigma^2 = \exp(c_0)$. Notice that there is an interesting analogy with the Von Mises distribution on the circle, $f(\omega) = \sigma^2 I_0(2c_1)g(\omega; 0, 2c_1)$, where $g(.)$ is the density of the Von Mises distribution, see Mardia and Jupp (2000). For integer k, γ_k is real and symmetric. Moreover,

$$\sum_{k=-\infty}^{\infty}\gamma_k = \sigma^2\exp\{2c_1\}$$

which follows from $\sum_{k=0}^{\infty} \gamma_k = \sigma^2 \frac{1}{2}(I_0(2c_1) + \exp(2c_1))$, see Abramowitz and Stegun (1972, 9.6.37). The coefficients of the Wold representation are obtained from (2): $\varphi_j = (j!)^{-1}c_1^j, j > 0$, which highlights the differences with the impulse response of autoregressive (AR) process of order 1 (it converges to zero at a faster rate than geometric).

The truncated cepstral process of order K, with $f(\omega) = \frac{1}{2\pi}\exp(c_0 + 2\sum_{k=1}^{K} c_k \cos(\omega k))$, is such that the spectral density can be factorized as

$$f(\omega) = \frac{\sigma^2}{2\pi} \prod_{k=1}^{K} \left[I_0(2c_k) + 2\sum_{j=1}^{\infty} I_j(2c_k)\cos(\omega k j) \right].$$

This result comes from the Fourier expansion of the factors $\exp(2c_k \cos(\omega k))$.

2.3 Fractional Exponential (FEXP) processes

Let us consider the process y_t, generated according to the equation $y_t = (1-B)^{-d}\xi_t$, where ξ_t is a short memory process and d is the long memory parameter, $0 < d < 0.5$. The spectral density can be written as $|2\sin(\omega/2)|^{-2d} f_\xi(\omega)$, where the first factor is the power transfer function of the filter $(1-B)^{-d}$, i.e. $|1 - e^{-\iota\omega}|^{-2d} = |2\sin\frac{\omega}{2}|^{-2d}$, and $f_\xi(\omega)$ is the spectral density function of a short memory process, whose logarithm admits a Fourier series expansion.

The logarithm of the spectral generating function of y_t is thus linear in d and in the cepstral coefficients of the Fourier expansion of $\ln 2\pi f_\xi(\omega)$, denoted $c_k, k = 1, 2, \ldots$:

$$\ln[2\pi f(\omega)] = -2d\ln\left|2\sin\frac{\omega}{2}\right| + c_0 + 2\sum_{k=1}^{\infty} c_k \cos(\omega k). \tag{4}$$

Here, c_0 retains its link to the p.e.v., $\sigma^2 = \exp(c_0)$, as $\int_{-\pi}^{\pi} \ln\left|2\sin\frac{\omega}{2}\right| d\omega = 0$. In view of the result

$$-\ln\left|2\sin\frac{\omega}{2}\right| = \sum_{k=1}^{\infty} \frac{\cos(\omega k)}{k}$$

(see also Gradshteyn and Ryzhik, 1994, formula 1.441.2), which tends to infinity when $\omega \to 0$, we rewrite (4) as

$$\ln[2\pi f(\omega)] = c_0 + 2\sum_{k=1}^{\infty} (c_k^* + c_k)\cos(\omega k),$$

with

$$c_k^* = -\frac{1}{2\pi}\int_{-\pi}^{\pi} 2d\ln|2\sin(\omega/2)|\cos(k\omega)\mathrm{d}\omega = \frac{d}{k}, k > 0.$$

Hence, for a fractional noise (FN) process, for which $y_t \sim \text{WN}(0, \sigma^2)$, the cepstral coefficients show an hyperbolic decline ($c_k = d/k, k > 0$).

When $\ln 2\pi f_\xi(\omega)$ is approximated by an EXP(K) process, i.e. the last summand of (4) is truncated at K, a fractional exponential process of order K, FEXP(K), arises. The fractional noise case corresponds to the FEXP(0) process.

Finally, consider the long memory Gegenbauer processes $(1 - 2\cos\varpi B + B^2)^d y_t = \xi_t$ (see Hosking, 1981, Gray, Zhang and Woodward, 1989, and McElroy and Holan, 2012), where $\xi_t \sim \text{WN}(0, \sigma^2)$ and $\varpi \in [0, \pi]$. The log-spectrum is linear in the memory parameter (although it depends nonlinearly on λ) and, in particular,

$$\ln[2\pi f(\omega)] = \ln\sigma^2 - 2d\ln\left|4\sin\left(\frac{\omega + \varpi}{2}\right)\sin\left(\frac{\omega - \varpi}{2}\right)\right|.$$

By straightforward algebra,

$$\ln[2\pi f(\omega)] = \ln\sigma^2 + 2\sum_{k=1}^{\infty}\frac{2d\cos(\varpi k)}{k}\cos(\omega k)$$

and by application of Gradshteyn and Ryzhik (1994), formula 1.448.2, the cepstral coefficients for $k = 1, \ldots, \infty$ of the above Gegenbauer process are

$$c_k = \frac{2d}{k}\cos(\varpi k).$$

It is perhaps interesting to remark the difference with the short memory cyclical process, $(1 - 2r\cos\varpi B + r^2 B^2)y_t = \xi_t$, considered in section 2.1: if r is close to one it will be difficult to discriminate this process from a long memory Gegenbauer process.

3 The Periodogram and its Sampling Properties

The main tool for estimating the spectral density function and its functionals is the periodogram. Due to its sampling properties, a generalized linear model for exponential random variables with logarithmic link can be formulated for the spectral analysis of a time series. The strength of the approach lies in the linearity of the log-spectrum in the cepstral coefficients and the long memory parameter.

Let $\{y_t, t = 1, 2, \ldots, n\}$ denote a time series, which is a sample realisation from a stationary Gaussian process, and let $\omega_j = \frac{2\pi j}{n}$, $j = 1, \ldots, [n/2]$, denote the Fourier frequencies, where $[\cdot]$ denotes the integer part of the argument. The periodogram, or sample spectrum, is defined as

$$I(\omega_j) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n}(y_t - \bar{y})e^{-\imath\omega_j t}\right|^2,$$

where $\bar{y} = n^{-1}\sum_{t=1}^{n}y_t$. In large samples (Koopmans, 1974, ch. 8)

$$\frac{I(\omega_j)}{f(\omega_j)} \sim \text{IID}\frac{1}{2}\chi_2^2, \qquad \omega_j = \frac{2\pi j}{n}, \ j = 1, \ldots, [(n-1)/2], \tag{5}$$

9

whereas $\frac{I(\omega_j)}{f(\omega_j)} \sim \chi_1^2, \omega_j = 0, \pi$, where χ_m^2 denotes a chi-square random variable with m degrees of freedom, or, equivalently, a Gamma($m/2$,1) random variable. As a particular case, $\frac{1}{2}\chi_2^2$ is an exponential random variable with unit mean.

The above distributional result is the basis for approximate maximum likelihood inferences for the EXP(K) or FEXP(K) as models for the spectrum a time series. In the sequel we shall exclude the frequencies $\omega = 0, \pi$ from the analysis - the latter may be included with little effort, but their effect on the inferences is negligible in large samples. This mode of inference is often referred to as Whittle likelihood estimation and will constitute the topic of the next section.

The same theory has motivated the estimation of the cepstral coefficients via log-periodogram regression, which, for instance for the EXP(K) model, yields unbiased, but inefficient, estimates of the parameters c_k by applying ordinary least squares to the log-periodogram regression:

$$\ln\left[2\pi I(\omega_j)\right] - \psi(1) = c_0 + 2\sum_{k=1}^{K} c_k \cos(\omega k) + \epsilon_j, \qquad \omega_j = \frac{2\pi j}{n}, j = 1, \dots, \left[\frac{n-1}{2}\right], \qquad (6)$$

where $\psi(\cdot)$ is the digamma function, $\psi(1) = -\gamma$, where $\gamma = 0.57722$ is Euler's constant, and ϵ_j is a centered log-chisquare variable with $\text{Var}(\epsilon_j) = \psi'(1) = \pi^2/6$. Log-periodogram regression for seasonal long memory processes is considered in Hsu and Tsai (2009). The Geweke and Porter-Hudak (1983) estimator of d is based on the OLS regression of the log-periodogram ordinates on a constant and $-2\ln\left|2\sin\frac{\omega}{2}\right|$. An improved bias-reduced estimator based on log-periodogram regression is analysed in Andrews and Guggenberger (2003).

4 Approximate (Whittle) Likelihood Inference

Estimation by maximum likelihood (ML) of the truncated cepstral, or EXP(K), model was proposed by Bloomfield (1971); later Cameron and Turner (1987) showed that ML estimation is carried out by iteratively reweighted least squares (IRLS). Whittle estimation of the FEXP(K) model was proposed by Beran (1993) and enjoys wide popularity; see, among others, Hurvich (2002) and, more recently, Narukawa and Matsuda (2011). We now review this methodology with reference to the FEXP(K) model (which nests the EXP(K) model) and discuss the sampling properties of the ML estimator of the long memory parameter and the cepstral coefficients.

Denote by z_j the vector of explanatory variables at the Fourier frequency ω_j and by θ the unknown coefficients, e.g. in the long memory case $z_j' = [1, 2\cos\omega_j, 2\cos(2\omega_j), \dots, 2\cos(K\omega_j), -2\ln|2\sin(\omega_j/2)|]$, $\theta' = (c_0, c_1, c_2, \dots, c_K, d)$, so that the exponential model is parameterised as

$$\ln 2\pi f(\omega_j) = z_j'\theta.$$

Exploiting the result that, for a Gaussian time series, $I(\omega_j)$ at the Fourier frequencies $0 < \omega_j < \pi$ is exponential with mean $f(\omega_j)$, the log-likelihood of $\{I(\omega_j), j = 1, \ldots, N = [(n-1)/2]\}$, is:

$$\ell(\theta) = -\sum_{j=1}^{N}\left[\ln f(\omega_j) + \frac{I(\omega_j)}{f(\omega_j)}\right] = N \ln 2\pi - \sum_{j=1}^{N}\ell_j(\theta), \qquad (7)$$

where

$$\ell_j(\theta) = z_j'\theta + \frac{2\pi I(\omega_j)}{\exp(z_j'\theta)}.$$

The score vector and and the Hessian matrix are

$$s(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = -\sum_j z_j u_j, \quad u_j = 1 - \frac{2\pi I(\omega_j)}{\exp(z_j'\theta)}$$

$$H(\theta) = \frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta'} = -\sum_j W_j z_j z_j', \quad W_j = \frac{2\pi I(\omega_j)}{\exp(z_j'\theta)}$$

so that the expected Fisher information is $\mathcal{I}(\theta) = -\mathrm{E}[H(\theta)] = \sum_j z_j z_j'$.

Estimation is carried out by the Newton-Raphson algorithm, i.e. iterating until convergence

$$\tilde{\theta}_{i+1} = \tilde{\theta}_i - [H(\tilde{\theta}_i)]^{-1}s(\tilde{\theta}_i)$$

or by the method of scoring:

$$\tilde{\theta}_{i+1} = \tilde{\theta}_i + [\mathcal{I}(\tilde{\theta}_i)]^{-1}s(\tilde{\theta}_i)$$

with starting value obtained by log-periodogram regression,

$$\tilde{\theta}_0 = \left(\sum_j z_j z_j'\right)^{-1}\sum_j z_j(\ln[2\pi I(\omega_j)] + \gamma).$$

In the former case, $\tilde{\theta}_{i+1}$ is obtained equivalently by iteratively reweighted least squares (IRLS), i.e. by the regression of $z_j'\tilde{\theta}_i - u_j/W_j$ on z_j with weights $W_j, j = 1, \ldots, N$, where both u_j and W_j are evaluated at $\tilde{\theta}_i$, see Cameron and Turner (1987). On the contrary, the Fisher scoring update is carried out by the unweighted regression of $z_j'\tilde{\theta}_i - u_j$ on z_j.

The estimator is asymptotically normal (Dahlhaus, 1989), with $\sqrt{n}(\theta - \tilde{\theta}) \to_d \mathrm{N}(0, V)$,

$$V^{-1} = \frac{1}{4\pi}\int_{-\pi}^{\pi} z(\omega)z(\omega)'\mathrm{d}\omega = \begin{bmatrix} \frac{1}{2} & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 1 \\ 0 & 0 & 1 & \ddots & 0 & \frac{1}{2} \\ \vdots & \vdots & 0 & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \ldots & 1 & \frac{1}{K} \\ 0 & 1 & \frac{1}{2} & \ldots & \frac{1}{K} & \frac{\pi^2}{6} \end{bmatrix}.$$

11

where $z(\omega)' = (1, 2\cos\omega, 2\cos(2\omega), \ldots, 2\cos(K\omega), -2\ln|2\sin(\omega/2)|)$.

Partitioning V as

$$V = \begin{bmatrix} V_c & V_{cd} \\ V_{dc} & V_d \end{bmatrix},$$

where V_c is the block corresponding to the cepstral parameters (the first $K+1$ elements of θ), V_d is the asymptotic variance of the long memory estimator, V_{cd} is the $K \times 1$ vector containing the asymptotic covariances between short and long memory parameter estimators and $V_{dc} = V_{cd}'$, it can be shown that

$$V_d = \left(\frac{\pi^2}{6} - \sum_{k=1}^{K} \frac{1}{k^2} \right)^{-1}, V_c = \begin{bmatrix} 2 & 0' \\ 0 & I + V_d rr' \end{bmatrix}, V_{cd} = \begin{bmatrix} 0 \\ -V_d r \end{bmatrix}, \quad (8)$$

where we have set $r = [1, 1/2, \ldots, 1/K]'$.

Notice that, in view of $\pi^2/6 = \sum_{k=1}^{\infty} k^{-2}$, $V_d \to \infty$ as $K \to \infty$.

This implies that the asymptotic correlation between the MLEs of $c_k, k = 1, \ldots, K$, and d is

$$\rho_{c_k d} = -\sqrt{\frac{V_d}{V_d + k^2}},$$

which is negative and tends to be large when K is large and larger when k is small compared to K.

The selection of the order K is the main specification issue: information criteria like AIC and BIC can be used for that purpose. Diagnostic checking can be based on the Pearson's residuals $\frac{2\pi I(\omega_j)}{\exp(z_j'\hat\theta)} - 1 = -u_j$.

A great deal of attention has been attracted by the issue of estimating the long memory parameter. However, the sampling properties of the FEXP parameter estimators are such that separating long memory from short will prove problematic, to say the least. First, it should be noticed that while the MLE of d is asymptotically independent of the MLE of the intercept, it is correlated with that of the cepstral coefficients $c_k, k > 1$: in large samples, $\mathrm{corr}(2\cos(\omega k), -2\ln|2\sin(\omega/2)|) = \frac{\sqrt{6}}{\pi}\frac{1}{k} = \frac{0.78}{k}$), so that, as we have seen before, the estimator of d is negatively correlated with that of c_k. As K increases, the correlation tends to -1 (see the formula for $\rho_{c_k d}$).

The inherent difficulty of separating long memory from short memory can be explained by the nature of the explanatory variables in the FEXP model and the fact that an high order EXP model can accommodate long memory effects, when K is sufficiently large.

5 Generalized Linear Cepstral Models with Power Link

The EXP model is a generalized linear model (McCullagh and Nelder, 1989) for exponentially distributed observations with mean function given by the spectral density a logarithmic link function. The generalization that we propose in this section is based on the observation that any continuous monotonic transform

of the spectral density function can be expanded as a Fourier series. We focus, in particular, on a parametric class of link functions, the Box-Cox link (Box and Cox, 1964), depending on a power transformation parameter, that encompasses the EXP model (logarithmic link), as well as the the identity and the inverse links; the latter is also the canonical link for exponentially distributed observations.

Let us thus consider the Box-Cox transform of the spectral generating function $2\pi f(\omega)$,

$$
g_\lambda(\omega) = \begin{cases} \frac{[2\pi f(\omega)]^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \ln[2\pi f(\omega)], & \lambda = 0. \end{cases}
$$

Its Fourier series expansion, truncated at K, is

$$
g_\lambda(\omega) = c_{\lambda,0} + 2 \sum_{k=1}^{K} c_{\lambda,k} \cos(\omega k), \tag{9}
$$

and the coefficients

$$
c_{\lambda k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} g_\lambda(\omega) \cos(\omega k) \mathrm{d}\omega
$$

will be named *generalised cepstral coefficients* (GCC).

Hence, a linear model is formulated for $g_\lambda(\omega)$. The spectral model with Box-Cox link and mean function

$$
f(\omega) = \begin{cases} \frac{1}{2\pi} \left[1 + \lambda g_\lambda(\omega)\right]^{\frac{1}{\lambda}}, & \lambda \neq 0, \\ \frac{1}{2\pi} \exp[g_\lambda(\omega)], & \lambda = 0 \end{cases}
$$

will be referred to as a generalized cepstral model (GCM) with parameter λ and order K, GCM(λ, K), in short. The EXP model thus corresponds the the case when the power parameter λ is equal to zero, and $c_{0k} = c_k$, are the usual cepstral coefficients.

For $\lambda \neq 0$, the GCC's are related to the generalised autocovariance function, introduced by Proietti and Luati (2012),

$$
\gamma_{\lambda k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} [2\pi f(\omega)]^\lambda \cos(\omega k) \mathrm{d}\omega \tag{10}
$$

by the following relationships:

$$
c_{\lambda 0} = \frac{1}{\lambda}(\gamma_{\lambda 0} - 1), \quad c_{\lambda k} = \frac{1}{\lambda} \gamma_{\lambda k}, k \neq 0. \tag{11}
$$

In turn, the generalised autocovariances are interpreted as the traditional autocovariance function of the power-transformed process:

$$
u_{\lambda t} = \left[\sigma\varphi\left(B^{\mathrm{s}(\lambda)}\right)\right]^\lambda \xi_t^*, \tag{12}
$$

where $\xi_t^* = \sigma^{-1}\xi_t$, $\mathrm{s}(\lambda)$ is the sign of λ, and $\left[\sigma\varphi\left(B^{\mathrm{s}(\lambda)}\right)\right]$ is a series in the lag operator whose coefficients can be derived in a recursive manner based on the Wold coefficients by Gould (1974). For $\lambda = 1$, $c_{1k} = \gamma_k$, the autocovariance function of the process is obtained. In the case $\lambda = -1$ and $k \neq 0$, $c_{-1,k} = -\gamma i_k$, where

γi_k is the inverse autocovariance of y_t (see Cleveland, 1972). The intercept $c_{\lambda 0}$ for $\lambda = -1, 0, 1$, is related to important characteristics of the stochastic process, as $1/(1 - c_{-1,0})$ is the interpolation error variance, $\exp(c_{0,0}) = \sigma^2$, the prediction error variance, and $c_{1,0} + 1 = \gamma_0$ is the unconditional variance of y_t. Also, for $\lambda \to 0$, $c_{\lambda k} \to c_k$, i.e. the cepstrum is the limit of the GCC as λ goes to zero.

For a fractional noise process the GCCs are zero for $\lambda = -d^{-1}$ and $k > 1$. This is so since $[2\pi f(\omega)]^{\lambda} = \sigma^2 |2\sin(\omega/2)|^{-2d\lambda} = \sigma^2 |1 - e^{-\iota\omega}|^2$ for $\lambda = -d^{-1}$, which is the spectrum of a non-invertible moving average process of order 1, whose autocovariances are $\gamma_{\lambda k} = 0$ for $k > 1$.

Let $g_{\lambda}(\omega_j) = z_j'\theta_{\lambda}$ where now $z_j = [1, 2\cos\omega_j, 2\cos(2\omega_j), \dots, 2\cos(K\omega_j)]'$, and $\theta_{\lambda}' = [c_{\lambda 0}, c_{\lambda 1}, \dots, c_{\lambda K}]'$. Then, the Whittle likelihood is

$$\ell(\theta_{\lambda}) = N \ln 2\pi - \sum_{j=1}^{N} \ell_j(\theta_{\lambda})$$

where,

$$\ell_j(\theta_{\lambda}) = \begin{cases} \frac{1}{\lambda} \ln(1 + \lambda z_j'\theta_{\lambda}) + \frac{2\pi I(\omega_j)}{(1 + \lambda z_j'\theta_{\lambda})^{\frac{1}{\lambda}}}, & \lambda \neq 0, \\[2mm] z_j'\theta_0 + \frac{2\pi I(\omega_j)}{\exp(z_j'\theta_0)}, & \lambda = 0. \end{cases}$$

The score vector and and the Hessian matrix, when $\lambda \neq 0$ (for the case $\lambda = 0$ see section 4), are respectively

$$s(\theta_{\lambda}) = \frac{\partial \ell(\theta_{\lambda})}{\partial \theta_{\lambda}} = -\sum_{j} z_j^* u_j, \quad u_j = 1 - \frac{2\pi I(\omega_j)}{(1 + \lambda z_j'\theta_{\lambda})^{\frac{1}{\lambda}}}, \quad z_j^* = \frac{z_j}{1 + \lambda z_j'\theta_{\lambda}},$$

$$H(\theta_{\lambda}) = \frac{\partial^2 \ell(\theta_{\lambda})}{\partial \theta_{\lambda} \partial \theta_{\lambda}'} = -\sum_{j} W_j^* z_j^* z_j^{*\prime}, \quad W_j^* = \frac{2\pi I(\omega_j)}{(1 + \lambda z_j'\theta_{\lambda})^{\frac{1}{\lambda}}} - \lambda u_j.$$

It follows from theorem 2.1 in Dahlhaus (1989) that

$$\sqrt{n}(\tilde{\theta}_{\lambda} - \theta_{\lambda}) \to_d N(0, V_{\lambda}), \quad V_{\lambda}^{-1} = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{1}{[2\pi f(\omega)]^{2\lambda}} z(\omega) z(\omega)' d\omega.$$

Dahlhaus (1989) also proves the efficiency of the Whittle estimator.

5.1 White Noise and Goodness of Fit Test

We consider the problem of testing the white noise hypothesis $H_0 : c_{\lambda 1} = c_{\lambda 2} = \dots = c_{\lambda K} = 0$ in the GCM framework, with λ and K given. Interestingly, the score test statistic is invariant with λ and is asymptotically equivalent to the Box-Pierce (1970) test statistic. This is immediate to see for the traditional EXP(K) case, that is when $\lambda = 0$. Under the assumption $y_t \sim WN(0, \sigma^2)$, $2\pi f(\omega) = \exp(c_0)$ and the Whittle estimator of c_0 is the logarithm of the sample periodogram mean, $2\pi \bar{I} = \frac{1}{N} \sum_j 2\pi I(\omega_j)$ (which is also an estimate of the variance - for a WN process the p.e.v. equals the variance): $\tilde{c}_0 = \ln(2\pi \bar{I})$.

The score test of the null $H_0 : c_1 = c_2 = \cdots = c_K = 0$ in an EXP(K) model is

$$\mathcal{S}_{WN}(K) = \frac{1}{n} \left(\sum_j z_j \tilde{u}_j \right)' \left(\sum_j z_j \tilde{u}_j \right) \approx n \sum_{k=1}^{K} \hat{\rho}_k^2,$$

where $z_j' = [1, 2\cos\omega_j, 2\cos(2\omega_j), \ldots, 2\cos(K\omega_j)]$, $\tilde{u}_j = 1 - I(\omega_j)/\bar{I}$, and we rely on the large sample approximations: $2N \approx n$, $\frac{1}{N} \sum_{j=1}^{N} \frac{I(\omega_j)}{\bar{I}} \cos(\omega_j k) \approx \hat{\rho}_k$, the lag k sample autocorrelation (see Brockwell and Davis, prop. 10.1.2). Hence, $\mathcal{S}_{WN}(K)$ is the same as the Box-Pierce (1970) portmanteau test statistic. The same holds in the case when $\lambda \neq 0$, where $g_\lambda(\omega) = c_0$, $\tilde{c}_0 = \frac{(2\pi\bar{I})^\lambda - 1}{\lambda}$ and $1 + \lambda\tilde{c}_0 = (2\pi\bar{I})^\lambda$.

On the contrary, the likelihood ratio test can be shown to be equal to

$$LR = 2N \left(\ln \bar{I} - \frac{1}{N} \sum_{j=1}^{N} \ln \tilde{f}(\omega_j) \right) = 2N \ln \frac{V}{p.e.v.},$$

where $V = 2\pi\bar{I}$ is the unconditional variance of the series, estimated by the averaged periodogram, and the prediction error variance in the denominator is estimated by the geometric average of the estimated spectrum under the alternative. The former is also the p.e.v. implied by the null model; the latter depends on λ. Interestingly, the LR test can be considered a parametric version of the test proposed by Davis and Jones (1968), based on the comparison of the unconditional and the prediction error variance.

5.2 Reparameterization

The main difficulty with maximum likelihood estimation of the GCM(λ, K) model for $\lambda \neq 0$ is enforcing the condition $1 + \lambda z_j' \theta_\lambda > 0$. This problem is well known in the literature concerning generalized linear models with the inverse link for gamma distributed observations, for which is the canonical link is the inverse link (McCullagh and Nelder, 1989). Several strategies may help overcoming this problem, such as periodogram pooling (Bloomfield, 1973, Moulines and Soulier, 1999, Faÿ, Moulines and Soulier, 2002), which reduces the influence of the periodogram ordinates close to zero, and weighting the periodogram, so as to exclude in the estimation those frequencies for which the positivity constraint is violated.

The most appropriate solution is to reparameterize the generalized autocovariances and the cepstral coefficients as follows:

$$[2\pi f(\omega)]^\lambda = \sigma_\lambda^2 b_\lambda(e^{-\imath\omega}) b_\lambda(e^{\imath\omega}), \quad b_\lambda(e^{-\imath\omega}) = 1 + b_1 e^{-\imath\omega} + \cdots + b_K e^{-\imath\omega K}, \tag{13}$$

where the b_k coefficients are such that the roots of the polynomial $1 + b_1 z + \cdots + b_K z^K$ lie outside the unit circle, so that, for $\lambda \neq 0$, the GCC's are obtained as

$$c_{\lambda 0} = \frac{1}{\lambda} \left[\sigma_\lambda^2 (1 + b_1^2 + \cdots + b_K^2) - 1 \right], \quad c_{\lambda k} = \frac{1}{\lambda} \sigma_\lambda^2 \sum_{j=k}^{K} b_j b_{j-k}.$$

15

To ensure the positive definiteness and the regularity of the spectral density we adopt a reparameterization due to Barndorff-Nielsen and Schou (1973) and Monahan (1984): given K coefficients $\varsigma_{\lambda k}, k = 1, \ldots, K$, such that $|\varsigma_{\lambda k}| < 1$, the coefficients of the polynomial $b_\lambda(z)$ are obtained from the last iteration of the Durbin-Levinson recursion

$$b_{\lambda j}^{(k)} = b_{\lambda j}^{(k-1)} + \varsigma_{\lambda k} b_{\lambda, k-j}^{(k-1)}, \quad b_{\lambda k}^{(k)} = \varsigma_{\lambda k},$$

for $k = 1, \ldots, K$ and $j = 1, \ldots, k-1$. The coefficients $\varsigma_{\lambda j}$ are in turn obtained as the Fisher inverse transformations of unconstrained real parameters $\vartheta_j, j = 1, \ldots, K$, i.e. $\varsigma_{\lambda j} = \frac{\exp(2\vartheta_j)-1}{\exp(2\vartheta_j)+1}$ for $j = 1, \ldots, K$, which are estimated unrestrictedly. Also, we set $\vartheta_0 = \ln(\sigma_\lambda^2)$.

By this reparameterisation, alternative spectral estimation methods are nested within the GCM(λ, K) model. In particular, along with the EXP model ($\lambda = 0$), autoregressive estimation of the spectrum arises in the case $\lambda = -1$, whereas $\lambda = 1$ (identity link) amounts to fitting the spectrum of a moving average model of order K to the series. The profile likelihood of the GCM(λ, K) as λ varies can be used to select the spectral model for y_t. A similar idea has been used by Koenker and Yoon (2009) for the selection of the appropriate link function for binomial data; another possibility is to test for the adequacy of a maintained link (e.g. the logarithmic one) using the goodness of link test proposed by Pregibon (1980).

In conclusion, the GCM framework enables the selection of a spectral estimation method in a likelihood based framework. Another possible application of the GCM(λ, K) is the direct estimation of the inverse spectrum and inverse autocorrelations up to the lag K, which arises for $\lambda = -1$ (this corresponds to the inverse link) and of the optimal interpolator (Battaglia, 1983), which is obtained in our case from the corresponding b_k coefficients as $\sum_{k=1}^{K} \rho_{-1,k} y_{t\pm k}$ with

$$\rho_{-1,k} = \frac{\sum_{j=k}^{K} b_j b_{j-k}}{\sum_{j=0}^{K} b_j^2}.$$

which represents the inverse autocorrelation at lag k of y_t.

6 Gradient Boosting and Regularization for Cepstral Estimation

A crucial element in fitting an exponential model is the selection of the truncation parameter K. Rather than selecting K, another approach is regularizing the cepstral coefficients. This approach was considered by Wahba (1980), Chow and Grenander (1985) and Pawitan and O'Sullivan (1994), among others. Furthermore, modelling the log spectrum via periodic splines is a key ingredient in modelling locally stationary processes, see Rosen, Wood and Stoffer (2009, 2012).

Bayesian estimation of the spectrum by log-periodogram regression has a long tradition. Wahba (1980) estimated the model

$$\ln[2\pi I(\omega_j)] - \psi(1) = C(\omega_j) + \epsilon_j, \tag{14}$$

where $C(\omega) = \ln[2\pi f(\omega)]$ is a natural cubic spline, assuming a normal distribution for the error term, $\epsilon_j \sim \text{NID}(0, \psi(1)')$. The resulting estimate of the log-spectrum is

$$\hat{C}(\omega) = \hat{c}_0 + \sum_{k=1}^{N} \frac{1}{1 + \rho k^2} \hat{c}_k \cos(\omega k),$$

where ρ is a smoothness parameter and $\hat{c}_k$ are the sample Fourier coefficients of the centered log-periodogram. This amounts to applying a Butterworth filter to the empirical cepstral coefficients. Carter and Kohn (1997) modelled $C(\omega)$ in (14) as a periodic cubic spline, and approximated the distribution of ϵ_j in by a Gaussian mixture with 5 components.

Pawitan and O'Sullivan (1994) proposed a nonparametric estimator of $C(\omega) = \ln[2\pi f(\omega)]$ based on the maximisation of the penalised least squares criterion

$$L(\theta) = \int_{-\pi}^{\pi} \left[C(\omega) + \frac{2\pi I(\omega)}{\exp C(\omega)} \right] d\omega + \frac{\rho}{(2\pi)^{2r}} \int_{-\pi}^{\pi} |C^{(r)}(\omega)|^2 d\omega, \tag{15}$$

where ρ is a penalty parameter and $C^{(r)}(\omega)$ is the r-th derivative of $C(\omega)$. For the EXP model, $C(\omega) = c_0 + 2 \sum_{k=1}^{K} \cos(\omega k)$, the penalised least squares estimator maximises

$$\sum_{j=1}^{N} \left[\ln f(\omega_j) + \frac{I(\omega_j)}{f(\omega_j)} \right] + \rho \sum_{k=1}^{N} k^{2r} c_k^2$$

Interestingly, this is equivalent to a generalised ridge regression embodying a smoothness prior $c_k \sim \text{NID}(0, \rho k^{-2r}), k > 0$ on the cepstral coefficients.

It is not clear whether the above smoothness priors are reasonable. They surely are for cepstra that originate from ARMA processes. We will present an illustration, dealing with monthly sunspots, for which they are not, due to the presence of cycles with small periodicity, that require high order cepstral coefficients to be represented. Moreover, they do not extend straightforwardly to the FEXP model, for which a prior for the long memory parameter must also be entertained.

In this section we propose a gradient boosting algorithm for the estimation of the FEXP model, which exploits the generalized additive nature of the model. The proposed algorithm performs incremental forward stagewise fitting (Friedman, Hastie and Tibshirani, 2000) of the log-spectrum, by a sequence of Newton-Raphson iterations that iteratively improve the fit. At each step, a single candidate regressor is selected among the explanatory variables of the FEXP model and the current solution is moved towards the direction that increases maximimally the Whittle likelihood by a small fraction governed by the learning rate. This prevents the algorithm from becoming too greedy and yields coefficient patterns that have a close connection with the lasso (see Bühlmann, 2006, and Hastie, Tibshirani and Friedman, 2009). The stopping iteration can be selected by an information criterion. Excellent reviews of boosting are Bühlmann and Hothorn (2007) and Hastie, Tibshirani and Friedman (2009, chapter 10).

Suppose that we start from the estimate of log-spectrum $C(\omega)$ and we look for an improved estimate $C(\omega) + \nu c(\omega)$, where ν is a learning parameter, that we set equal to 0.1 to prevent overfitting, and $c(\omega)$ takes the form $\alpha z^+(\omega)$, where $z^+(\omega)$ is an element of the vector of the $K + 2$ explanatory variables in the FEXP model, $z(\omega) = (1, 2\cos\omega, 2\cos(2\omega), \ldots, 2\cos(K\omega), -2\ln|2\sin(\omega/2)|)'$, chosen so as to optimise the improvement in the fit, in the sense specified below. Here K is taken as possibly large number, e.g. N^a, $a \in [0.3, 0.6]$. When long memory is due to the presence of a Gegenbauer process at frequency ϖ, the last element in the vector $z(\omega)$ is obviously replaced by $2\ln\left|4\sin\left(\frac{\omega+\varpi}{2}\right)\sin\left(\frac{\omega-\varpi}{2}\right)\right|$.

For $z^+(\omega)$ given, the coefficient α can be estimated consistently by Whittle likelihood, performing a single Newton-Raphson iteration, with starting value $\hat{\alpha}_0 = 0$, which yields (see also section 4):

$$\hat{\alpha} = \frac{\sum_{j=1}^{N} z^+(\omega_j)u_j}{\sum_{j=1}^{N} W_j[z^+(\omega_j)]^2}, \quad u_j = 1 - \frac{2\pi I(\omega_j)}{\exp\{C(\omega_j)\}}, \quad W_j = \frac{2\pi I(\omega_j)}{\exp\{C(\omega_j)\}}.$$

Equivalently, $\hat{\alpha}$ results from the weighted least squares regression of u_j/W_j on $z^+(\omega_j)$, with weights W_j.

The selection of the regressor $z^+(\omega)$ is based on the maximum deviance reduction, were the deviance associated with the log-spectral fit $C(\omega)$ is

$$D(C) = 2\sum_{j=1}^{N} \left[C(\omega_j) + 2\pi I(\omega_j)/\exp\{C(\omega_j)\}\right]. \tag{16}$$

Hence, at every boosting iteration, $m = 1, \ldots, M$, and starting from the initial constant configuration $C_0(\omega) = \frac{1}{N}\sum_{j=1}^{N}\left(\ln[2\pi I(\omega_j)] - \psi(1)\right)$, we perform $K + 2$ weighted regressions and select the one which yields the largest reduction in the deviance (equivalently, the maximum increase in the Whittle likelihood), and we update the current estimate according to the learning rate ν as follows:

$$C_m(\omega) = C_{m-1} + \nu\alpha z_m^+(\omega), m = 1, 2, \ldots, M. \tag{17}$$

The stopping iteration is determined by an information criterion which imposes a penalty for the complexity of the fit, measured by the number of degrees of freedom. For the m-th update

$$df_m = df_{m-1} + \nu\frac{\sum_{j=1}^{N}[z_m^+(\omega_j)]^2}{\sum_{j=1}^{N} W_j[z_m^+(\omega_j)]^2}.$$

The initial value, corresponding to the constant spectral fit, is $df_0 = 1$. Alternatively, out of sample validation can be used, in which case we select the spectral fit that generalizes best to data that have not been used for fitting.

Figure 2: Box and Jenkins (1970) Series A. Chemical process concentration readings.
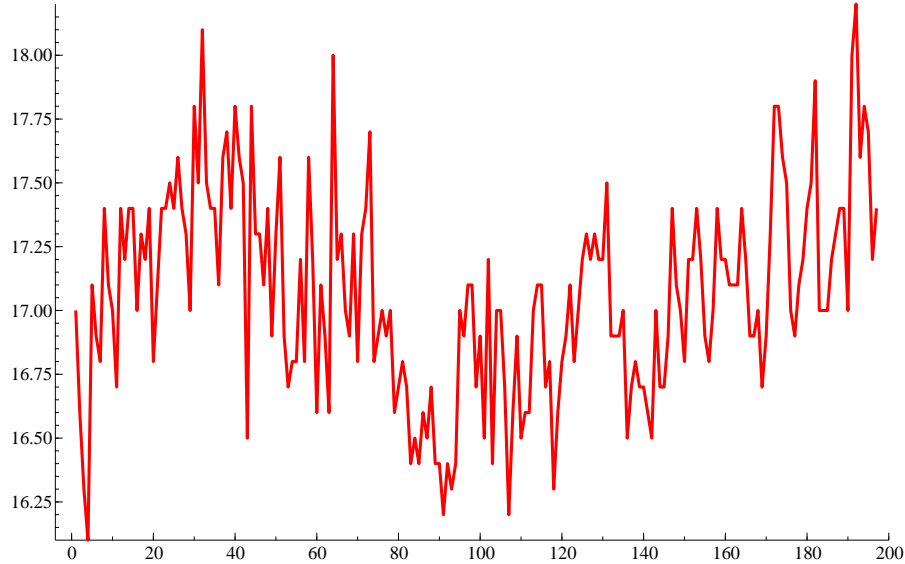
7 Illustrations

7.1 Box and Jenkins Series A

Our first empirical illustration deals with a time series popularised by Box and Jenkins (1970), concerning a sequence of $n = 200$ readings of a chemical process concentration, known as *Series A*. The series, plotted in figure 2, was investigated in the original paper by Bloomfield (1973), with the intent of comparing the EXP model with ARMA models. Bloomfield fitted a model with K chosen so as to match the number of ARMA parameters. Box and Jenkins (1970) had fitted an ARMA(1,1) model to the levels and an AR(1) to the differences. The estimated p.e.v. resulted 0.097 and 0.101, respectively. Thus, Bloomfield fitted the EXP(2) model to the levels and an EXP(1) to the 1st differences by maximum likelihood, using a modification which entails concentrating σ^2 out of the likelihood function. The estimated p.e.v. resulted 0.146 and 0.164, respectively. He found this rather disappointing and concluded that ARMA models are more flexible.

Actually, there is no reason for constraining K to the number of parameters of the ARMA model. If model selection is carried out and estimation by MLE is performed by IRLS, AIC selects an EXP(7) for the levels and an EXP(5) for the 1st differences. The estimated p.e.v. is 0.099 and 0.097, respectively. BIC selects an EXP(3) for both series. The p.e.v. estimates are 0.103 and 0.103.

Also, the FEXP(0) provides an excellent fit: the d parameter is estimated equal to 0.437 (with standard error 0.058), and the p.e.v. is 0.100.

Figure 3: BJ Series A. Spectrum and log-spectrum estimation by an exponential model with K selected by AIC and a fractional exponential model.
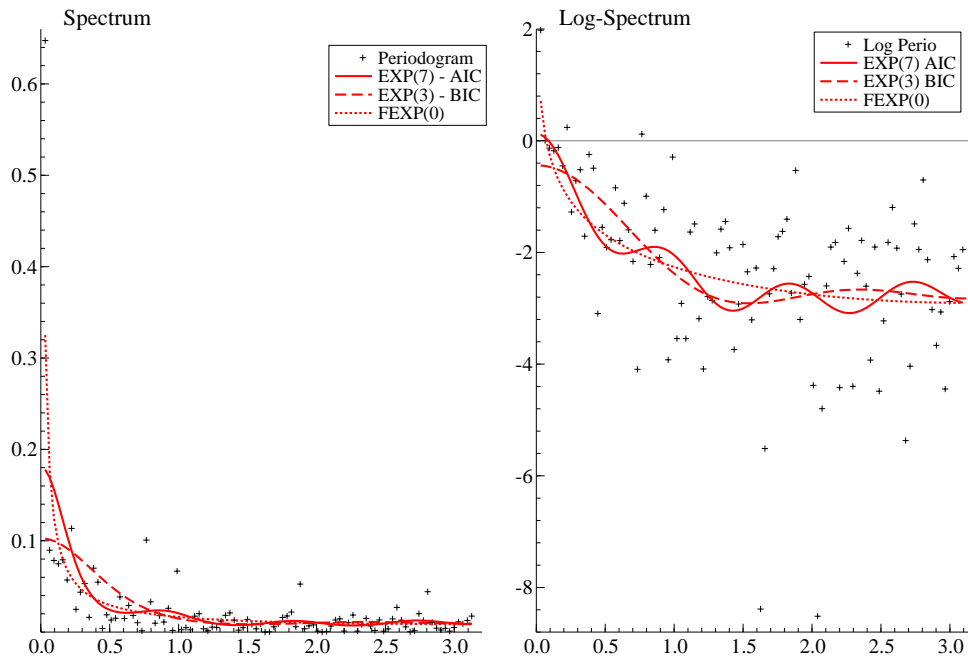

Figure 3 presents the periodogram and the fitted spectra for the two EXP specifications and the FEXP model (left plot). The right plot displays the centered log-periodogram $\ln\left[2\pi I(\omega_j)\right] - \psi(1)$ and compares the fitted log-spectra. It could be argued that EXP(7) is prone to overfitting and that the FEXP(0) model provides a very good fit, the first periodogram ordinate $I(\omega_1)$ being very influential in determining the fit.

The FEXP(0) model estimated on the the first differences yields an estimate of the memory parameter d equal to -0.564 (s.e. 0.056), and the p.e.v. is 0.098. These results are consistent with the FEXP model applied to the levels, as a negative d is estimated.

This example illustrates that the exponential model provides a fit that is comparable to that of an ARMA model, in terms of the prediction error variance. There is a possibility that the series has long memory, which has not been explored in the literature.

When we move to fitting the more general class of GMC(λ, K) models, both AIC and BIC select the model GCM(-2.29, 1); see table 1, which refers to the AIC. Notice that the EXP(5) and EXP(7) are characterised by a much smaller AIC (see the row corresponding to $\lambda = 0$). Table 2 displays the values of the estimated b_1 coefficient and the corresponding generalised cepstral coefficient $c_{\lambda 1}$, as well as the value of the maximised likelihood and prediction error variance, for the first order model GCM$(\lambda, 1)$, as the transforma-

Table 1: BJ Series A. Values of the Akaike Information Criterion for GCM(λ, K) models. The selected model is GCM(-2.29, 1).

				Values of K			
λ	1	2	3	4	5	6	7
-2.50	-3.131	-3.121	-3.118	-3.108	-3.099	-3.089	-3.079
-2.29	-3.133	-3.124	-3.118	-3.108	-3.100	-3.090	-3.080
-2.00	-3.128	-3.126	-3.117	-3.108	-3.100	-3.093	-3.084
-1.50	-3.095	-3.119	-3.109	-3.100	-3.094	-3.091	-3.091
-1.00	-3.049	-3.104	-3.097	-3.090	-3.083	-3.082	-3.091
-0.50	-3.000	-3.083	-3.085	-3.079	-3.071	-3.070	-3.087
0.00	-2.955	-3.055	-3.072	-3.069	-3.061	-3.056	-3.077
0.50	-2.918	-3.024	-3.057	-3.062	-3.053	-3.045	-3.063
1.00	-2.888	-2.994	-3.037	-3.053	-3.048	-3.038	-3.049

tion parameter varies. For the specification selected according to information criteria, the implied spectrum is $2\pi \tilde{f}(\omega) = \sigma_\lambda^{2/\tilde{\lambda}} |1 + \tilde{b}_1 e^{-\iota\omega}|^{2/\tilde{\lambda}} = \sigma_\lambda^{2/\tilde{\lambda}} |2\sin(\omega/2)|^{-2\times 0.44}$, which results from replacing $\tilde{b}_1 = -1$ and $\tilde{\lambda} = -2.29$ and by application of the two-angle trigonometric formula. This is the spectral density of a fractional noise process with memory parameter $d = 0.44$. It is indeed remarkable that likelihood inferences and model selection applied to the GCM(λ, K) model, point to the same results obtained by the FEXP(0) model discussed above. We interpret these results as further confirming the long memory nature of the series.

7.2 Simulated AR(4) Process

As our second example we consider $n = 1024$ observations simulated from the AR(4) stochastic process $y_t = 2.7607y_{t-1} - 3.8106y_{t-2} + 2.6535y_{t-3} - 0.9238y_{t-4} + \xi_t, \xi_t \sim \text{NID}(0, 1)$. The series is obtained from Percival and Walden (1993) and constitutes a test case for spectral estimation method, as the data generating process features a bimodal spectrum, with the peaks located very closely. In fact, the AR polynomial features two pairs of complex conjugate roots with modulus 1.01 and 1.02 and phases 0.69 and 0.88, respectively. As in Percival and Walden the series is preprocessed by a dpss data taper (with bandwidth parameter $W = 2/n$, see Percival and Walden, sections 6.4 and 6.18, for more details).

The specifications of the class GCM(λ, K) selected by AIC and BIC differ slightly. While the latter selects the true generating model, that is $\lambda = -1$ and $K = 4$, AIC selects $\lambda = -1$ and $K = 6$. However, the likelihood ratio test of the null that $K = 4$ is a mere 4.8.

The estimated coefficients of the GCM(-1, 4) model and their estimation standard errors are

Table 2: GCM(λ, 1) models: Whittle likelihood estimates of b_1, the GCC $c_{\lambda 1}$; value of log-likelihood at the maximum, $\ell(\hat{\vartheta})$, and prediction error variance (p.e.v.).

	GCM(λ, 1)			
λ	b_1	$c_{\lambda 1}$	$\ell(\hat{\vartheta})$	p.e.v
-2.50	-1.000	124.952	309.399	0.100
-2.29	-1.000	84.230	309.609	0.100
-2.25	-1.000	78.112	309.602	0.101
-2.00	-1.000	48.654	309.070	0.101
-1.50	-0.840	16.878	305.880	0.103
-1.00	-0.578	5.363	301.346	0.108
-0.50	-0.274	1.631	296.491	0.113
0.00	-	0.496	292.100	0.117
0.50	0.221	0.154	288.432	0.122
1.00	0.393	0.049	285.433	0.126

Figure 4: Periodogram and log-spectra estimated by the GCM(-1, 4), selected by BIC, and EXP(5) models.
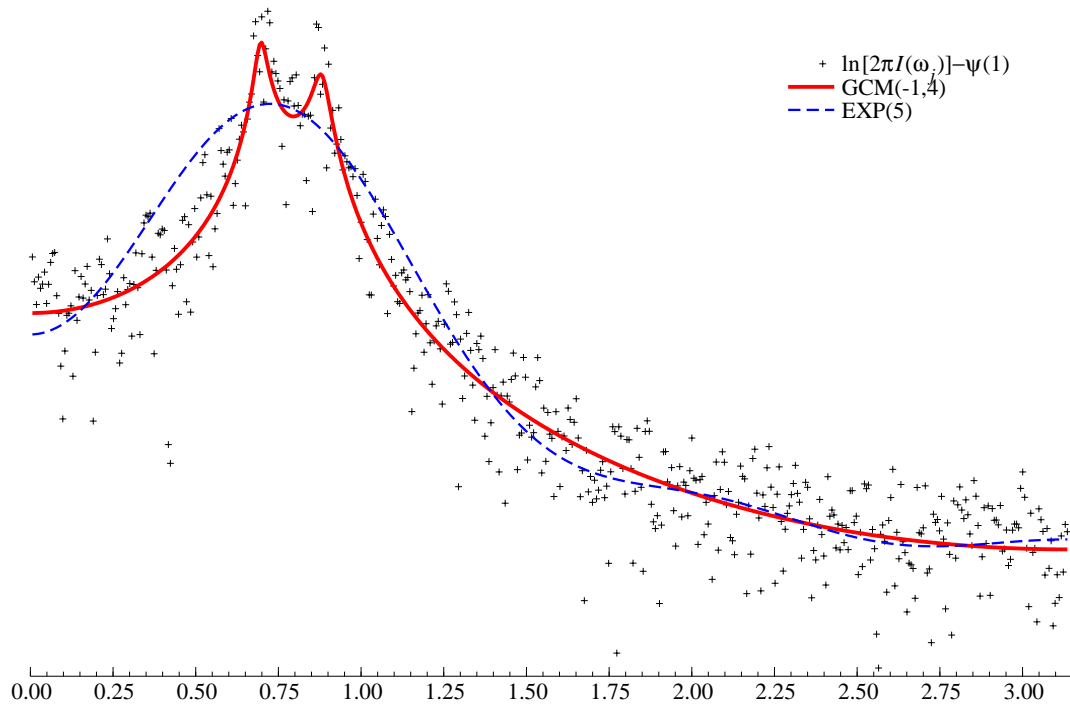
b_k	std. err.	true value
-2.7490	0.0007	-2.7607
3.7901	0.0016	3.8106
-2.6353	0.0007	-2.6535
0.9201	0.0025	0.9238

The comparison with the true autoregressive coefficients (reported in the last column) stresses that they are remarkably accurate. Figure 4 displays the centered periodogram and compares the log-spectra fitted by the selected GCM(-1,4) model and the EXP(5) model, which emerges if Box-Cox transformation parameter is set equal to zero. The latter fit is clearly suboptimal, as it fails to capture the two spectral modes.

7.3 Monthly Sunspot Series

The monthly averages of the sunspot numbers visible on the sun and can be considered a testbed for time series methods, although most often annual sunspot series are analysed (see Xia and Tong, 2011, section 6.2, and the references therein). The series is compiled by the Solar Influences Data Analysis Center in Belgium and it is known as the International Sunspot Number series; its construction is documented in Hathaway (2010) and the webpage http://solarscience.msfc.nasa.gov/SunspotCycle.shtml; we consider the monthly observations from January 1848 to December 2012, for a total of 1980 observations. The series is plotted in the top panel of figure 5, along with its sample correlogram. The most prominent feature is the presence of an approximate 11-year cycle (corresponding to 132 monthly observations and to the frequency $\varpi = 0.048$). Hathaway (2010) provides an excellent discussion of the main periodic features and discusses also other shorter cycles that are present in the series, among which the 154-day, annual and biannual cycles.

When the EXP model is fitted to the series, the estimated spectral density has a global maximum at the zero frequency, so that the implied autocorrelation function is not periodic and an important feature of the series is missed. As for the FEXP(K), which includes $-2\ln\left|4\sin\left(\frac{\omega+\varpi}{2}\right)\sin\left(\frac{\omega-\varpi}{2}\right)\right|$ among the regressors (here $\varpi = 2\pi/132$ is the frequency corresponding to a period of 11 years of monthly data), AIC selects $K = 26$, whereas the minimum BIC occurs for $K = 3$. The parameter estimates, as well as their standard error, the maximised likelihood and the estimated prediction error variance, are given in the following table for the two different specifications:

	FEXP(26)	FEXP(3)
$\tilde{d}$	0.392	0.454
s.e.	0.080	0.022
loglik	-4581.5	-4618.1
p.e.v.	238.0	246.8

23

Figure 5: International Sunspot Number series, January 1848 - December 2012, monthly averages. Series and sample autocorrelation function.
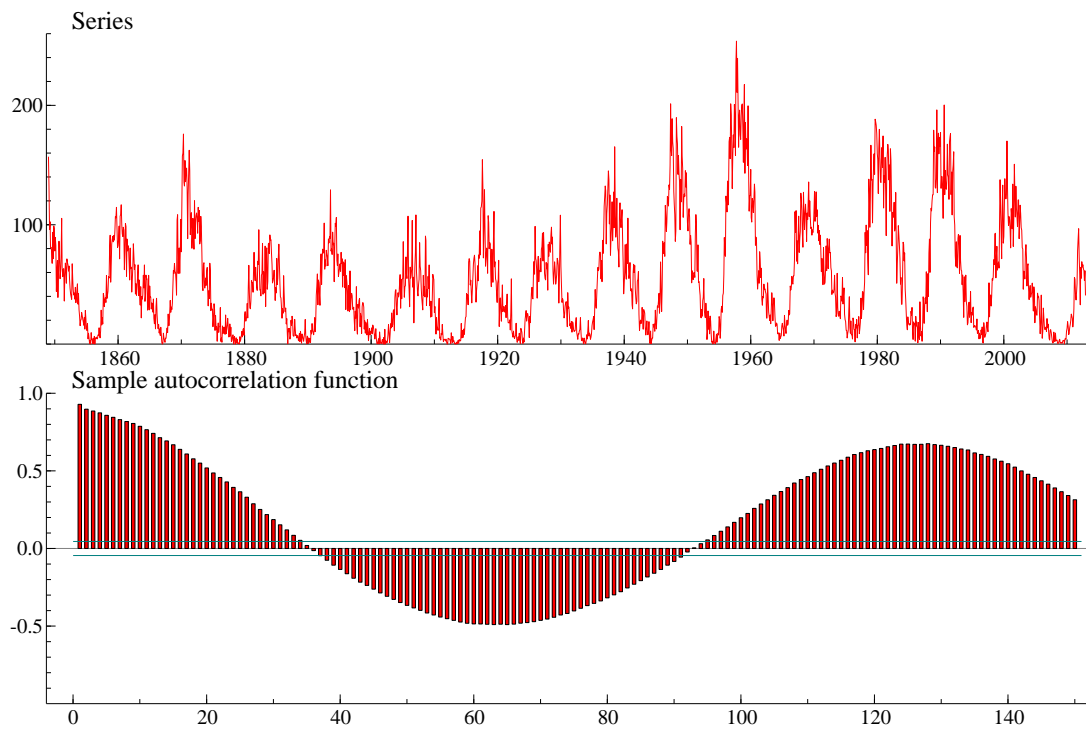
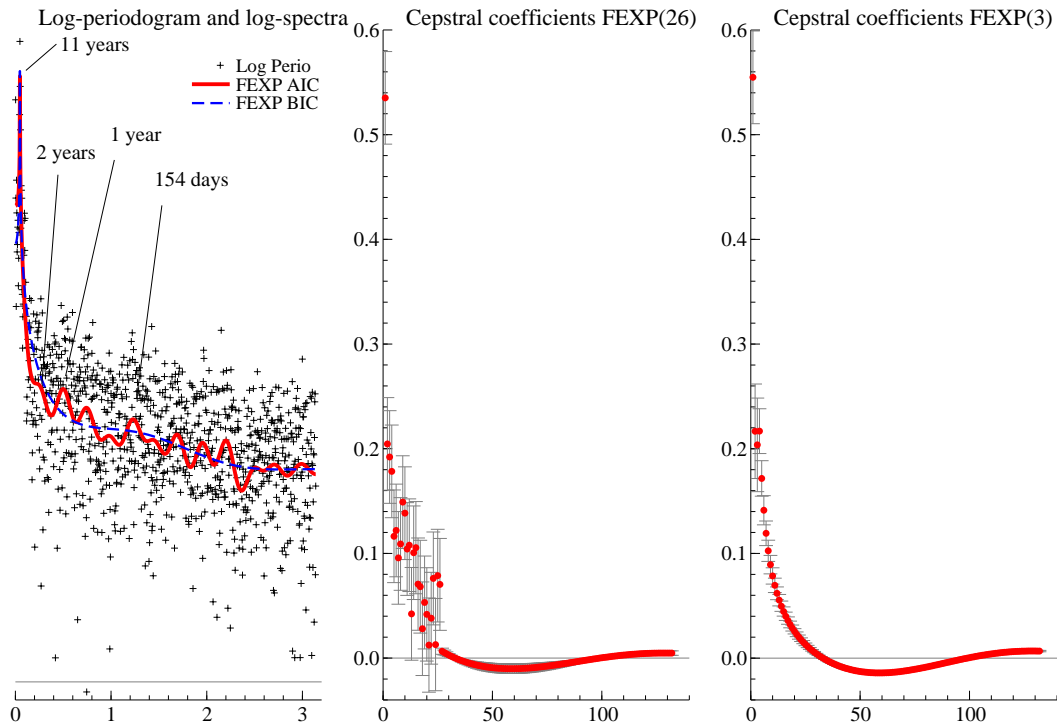Figure 6: Log-spectra of FEXP(26) and FEXP(3) models selected by AIC and BIC (first plot). Interval estimates of the cepstral coefficients $2d\cos(\varpi k)/k + c_k$ obtained from the FEXP(26) model (central plot) and the FEXP(3) model (right plot).

The log-spectra implied by the two different FEXP models are compared in the first plot of figure 6. While the 11 year cycle is captured by both models, the shorter cycles are fitted only by the FEXP(26) model. An important open issue is whether all the secondary peaks are a significant feature, or they originate from overfitting. The overall cepstral coefficients of the series, $2d\cos(\varpi k)/k + c_k$, resulting from the sum of the long-memory and the short memory component, are displayed in the second and the third plot of the figure, respectively for the FEXP(26) and the FEXP(3) models, along with their 95% confidence interval.

The smoothness prior used by Wahba and Pawitan and O'Sullivan for the short run coefficients c_k's is not suitable in this context, as the log-spectrum can feature local maxima. While the very first coefficients capture the behaviour around the long run frequency, higher order coefficients are responsible for the shorter cycles and shrinking them to zero would amount to ruling out shorter cycles.

We believe that gradient boosting can add valuable insight for selecting a model for the spectrum of the time series. Its componentwise approach, by which at each iteration only the contribution of a single

variable is determined, produces a set of solutions that very closely resembles the solution path for the lasso (see Bühlmann, 2006 and Hastie, Tibshirani and Friedman, 2009), thereby enabling variable selection and shrinkage. The total sample of 1980 time series observations is divided into two parts. The first half of the series is used as a training sample and the rest as a validation sample. The explanatory variables include $\cos(\omega k), k = 0, \ldots, 54$ and $-2 \ln \left| 4 \sin \left(\frac{\omega + \varpi}{2} \right) \sin \left(\frac{\omega - \varpi}{2} \right) \right|$. The boosting iterations provide a sequence of log-spectral estimates $C_m(\omega), m = 0, 1, \ldots, M$. The learning rate is set equal to $\nu = 0.1$, as it is customary to prevent the algorith from becoming too greedy. The coefficient profiles arising from $M = 500$ iterations of the algorithm are plotted in the left graph of figure 7.

If we let $D(C_m)$ denote the deviance associated to the log-spectrum fitted at the m-th boosting iteration (see expression (7.3)), the stopping iteration, m_{stop}, can be chosen according to the value that minimises $AIC = D(C_m) + 2df_m$ or $BIC = D(C_m) + \ln(n)df_m$. Alternatively, m_{stop} can be determined by out-of-sample validation, which amounts to evaluating the log-spectral fit $C_m(\omega)$ which minimises the deviance for the hold-out sample,

$$D^*(C_m) = 2 \sum_{j=1}^{N} \left[C(\omega_j) + 2\pi I^*(\omega_j) / \exp\{C(\omega_j)\} \right],$$

where $I^*(\omega_j)$ is the periodogram of the validation sample. The stopping iteration selected by AIC is $m_{stop} = 215$, whereas BIC has $m_{stop} = 108$. The crossvalidation (CV) criterion selects $m_{stop} = 259$, and, as a result, the log-spectrum selected by CV is very similar to that selected by AIC. The values of the long memory parameter and of the coefficients c_k's corresponding to these choices can be obtained from the left plot of figure 7. It emerges that the estimated value for d is not at all sensitive to the selection criterion, and it is about 0.25; the differences arise in the values of the c_k coefficients, though, with the AIC and CV solutions being characterised by larger high order coefficients, which implies that shorter cycles will be more represented. In conclusion, the selection criteria provide different characterizations of the short memory component.

The middle plot of figure 7 displays the iteration profile of the overall cepstral coefficients $2d \cos(\varpi k)/k + c_k$; the values selected at m_{stop} for the different criteria can be compared with those plotted in 6. The value of the first cepstral coefficient is larger for BIC, but all the remaining coefficients are smaller. The right plot of figure 7 displays the log-cepstral fits $C_{m_{stop}}(\omega)$ corresponding to the three criteria. Despite the fact that the short memory component is represented differently, the differences are minor, and we may tentatively conclude that we find statistical support for the presence of shorter cycles in conjunction with the 11 year cycle.

The next section will confirm that the boosting estimate of the long memory parameter tends to be stable across the selection criteria; moreover, it will show that it is more reliable than the maximum likelihood estimate arising from fitting the FEXP model.

Figure 7: Boosting for cepstral estimation. Coefficient profiles for the parameters d and c_k (left plot). Iteration profile of the overall cepstral coefficients, $2d\cos(\varpi k)/k + c_k$ (middle plot) and log-spectra selected by AIC, BIC and crossvalidation (CV) (right plot).
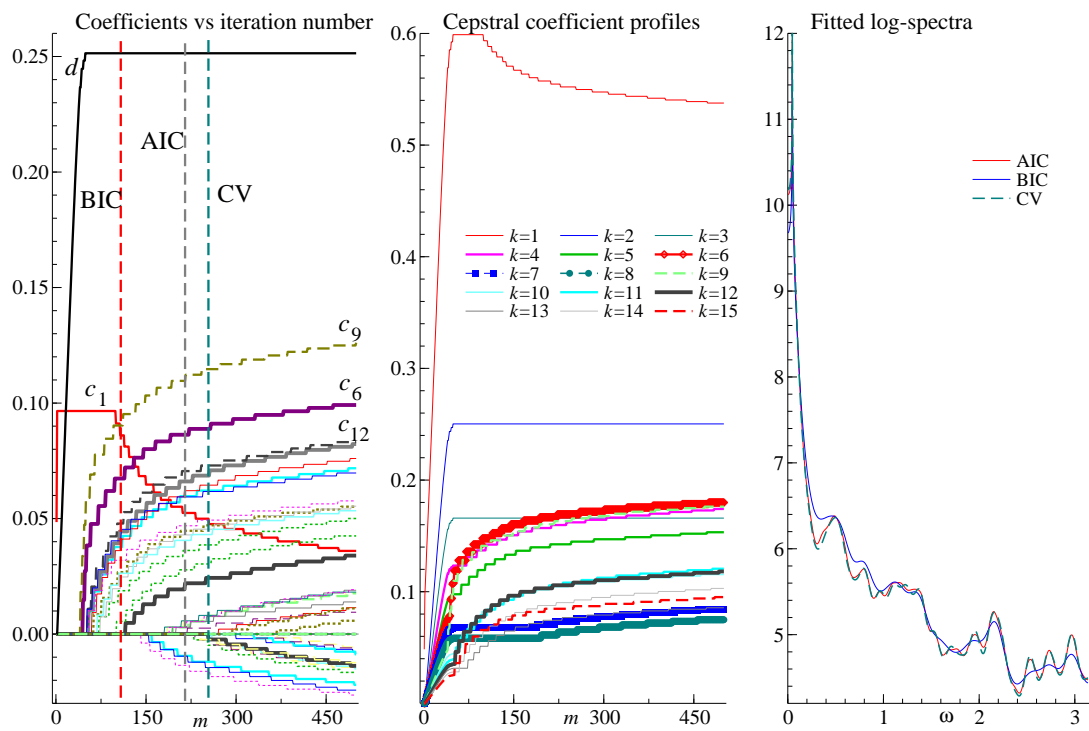
7.4 Separating Long Memory from Short

Assessing the contribution of the long memory component to the total log-spectrum can be very problematic when the short run component is very persistent. The difficulty is illustrated by a Monte Carlo simulation exercise, which also shows that boosting is remarkably much more accurate in this respect. We focus, in particular, on the long memory process $(1 - \phi B)(1 - B)^d y_t = \xi_t, \xi_t \sim \text{WN}(0, \sigma^2)$ with $\phi = 0.9$, $d = 0.4$ and $\sigma^2 = 1$; the cepstral coefficients of y_t for $k > 0$ result from the sum of $0.4/k$ and $0.9^k/k$ (see also section 7.4). In this situation it proves difficult to distill the contribution of the long memory component through the estimation of d in a FEXP framework.

We simulated 10,000 series of length 1000 from such a process and for each replication we estimated the FEXP(K) model, with K varying between 0 and 30, and with K selected by AIC and BIC. Due to the high correlation among the explanatory variables, the maximum likelihood estimator of d has a sampling distribution that depends substantially on K, as it can be seen from the top-left panel of figure 8, which displays the density of the MLE of the long memory paramenter, denoted $\tilde{d}_K$, as K varies. The bias-variance trade-off is evident: larger values of K imply less bias at the cost of a larger variance. The top right panel of figure 8 shows that the value $K = 13$ minimizes the mean square estimation error for d. This value does not concide with that minimising the Euclidean distance between the true log-spectrum and the estimated one: $\sum_{j=1}^{N} [\ln(2\pi f(\omega_j)) - z_j' \tilde{\theta}]^2$, displayed in the left bottom plot. Hence, if the criterion for assessing the quality of the FEXP fit is the mean square estimation error of the log-spectrum, then a smaller value of K, equal to 6, should be considered. Notice, however, that the density $\tilde{d}_6$ is highly concentrated in the nonstationarity region.

The last panel of figure 8 reports the percent frequency by which the FEXP order K is selected according to the BIC and AIC criteria. It seems that the AIC does a better job, in that its distribution is centered around the value that maximises the accuracy in the estimation of the log-spectrum.

When estimation is carried out by the boosting algorithm outlined in section 6, the picture changes quite dramatically: table 3 compares the mean, the standard deviation, the mean square error (multiplied by 100) of the estimates of the parameters c_0 (logarithm of p.e.v), c_1, c_2, c_3, and d obtained by the FEXP model with order selected by AIC and BIC with those obtained by boosting with stopping iteration chosen according to the two information criteria. It also compares the mean and the standard deviation of the distance of the estimated log-spectrum with the true one.

While it turns out that boosting achieves the same overall accuracy in estimating the log-spectrum (see the last column of the table), it is clearly superior in disentangling the contributions of the short and long memory components of the log-spectrum. The estimate of d suffers from a positive bias, but the distribution is concentrated in the stationary region and the mean square estimation error of the parameters is remarkably smaller.

Figure 8: Density and MSE of the MLE of d when y_t is genereated from $(1 - \phi B)(1 - B)^d y_t = \xi_t, \xi_t \sim$ WN$(0, \sigma^2)$ with $\phi = 0.9$, $d = 0.4$ and $\sigma^2 = 1$ (top panels). Log-spectrum distance as function of k (bottom left panel). Frequency distribution of K chosed by AIC and BIC (right bottom plot).
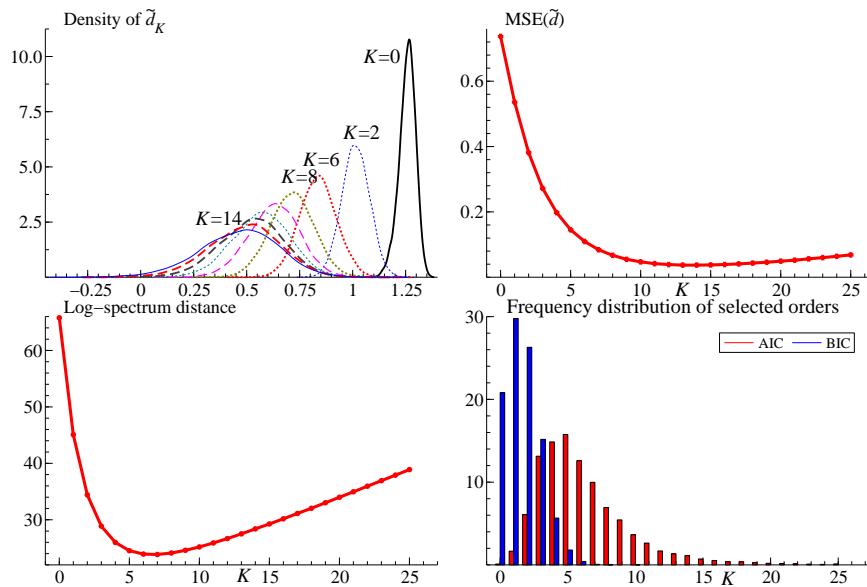
Table 3: Comparison of the sampling distributions of the estimators of the parameters $c_k, k = 0, 1, 2, 3$, and d of the FEXP model obtained by maximum likelihood and by the boosting algorithm.

	c_0	c_1	c_2	c_3	d	K	Distance
True value	0.000	0.900	0.405	0.243	0.400	∞	0
			FEXP - AIC				
Mean	0.090	0.609	0.266	0.152	0.670	6.156	31.520
St. Dev.	0.128	0.220	0.112	0.080	0.218	3.505	31.352
$100\times$ MSE	2.435	13.300	3.173	1.465	12.050		
			FEXP - BIC				
Mean	0.110	0.232	0.074	0.024	1.052	1.361	43.387
St. Dev.	0.128	0.182	0.095	0.055	0.169	1.173	35.642
$100\times$ MSE	2.841	47.941	11.863	5.104	45.397		
	c_0	c_1	c_2	c_3	d	m_{stop}	Distance
True value	0.000	0.900	0.405	0.243	0.400	∞	0
			Boosting - AIC				
Mean	-0.001	0.853	0.343	0.192	0.460	163.586	32.433
St. Dev.	0.012	0.050	0.035	0.034	0.027	16.356	34.294
$100\times$ MSE	0.015	0.473	0.509	0.369	0.436		
			Boosting - BIC				
Mean	0.001	0.884	0.326	0.171	0.459	106.775	43.864
St. Dev.	0.004	0.051	0.035	0.034	0.027	10.046	37.269
$100\times$ MSE	0.002	0.283	0.744	0.627	0.420		

Figure 9: Comparison of distribution of the estimates d, c_1 and c_2 arising from the FEXP model with AIC order selection and the boosting estimates with AIC step selection.
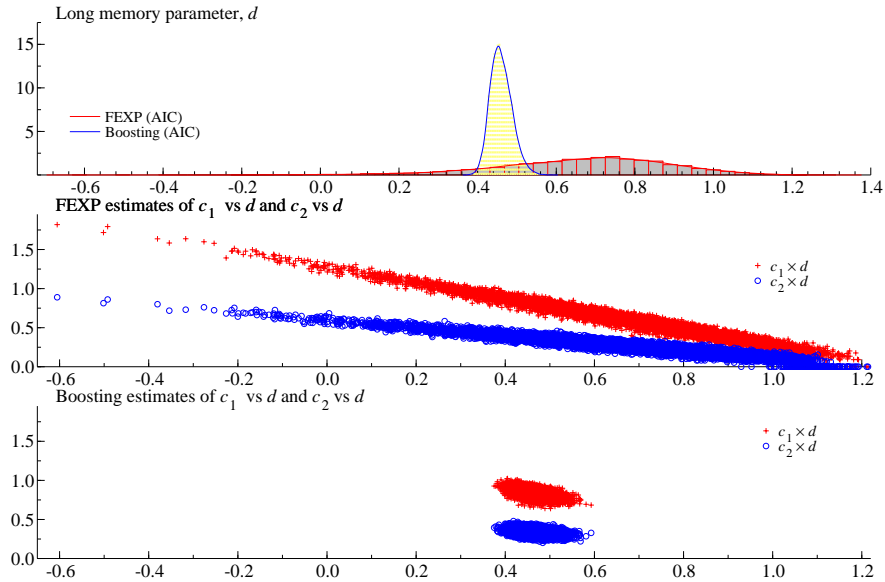
The differences and the efficiency gains can be better appreciated in figure 9, which compares the distributions of d when the selection criterion is the AIC. The central and bottom plot illustrate that the distribution of the estimators of c_1 and c_2 is strongly and negatively correlated with that of d; however their distribution is less sparse and the correlation is less strong when boosting is adopted.

8   Conclusions

Modelling the log-spectrum has a long tradition in the analysis of univariate time series and leads to computationally attractive likelihood based methods. We have devised a general frequency domain estimation framework within which nests the exponential model for the spectrum as a special case and allows for any power transformation of the spectrum to be modelled, so that alternative spectral fits can be encompassed. Another extension has dealt with a boosting algorithm for fitting the spectrum and we have discussed its potential for the estimation of long memory models. As a direction for future research we think that the exponential framework can have successful applications for modelling the time-varying spectrum of a locally stationary processes (Dahlhaus, 2012), by allowing the cepstral coefficients to vary over time, e.g. with autoregressive dynamics. Finally, a multivariate extension, the matrix-logarithmic spectral model for the spectrum of a vector time series, could be envisaged, along the lines of the model formulated by Chiu,

Leonard and Tsui (1996) for covariance structures.

References

[1] Andrews, D.W.K. and Guggenberger P.(1973), A Bias-Reduced Log-Periodogram Regression Estimator for the Long-Memory Parameter, *Econometrica*, 71, 675–712.

[2] Abramowitz, M., Stegun, I.A. (1972), *Handbook of Mathematical Functions with Formulas*, Courier Dover Publications.

[3] Battaglia, F. (1983), Inverse Autocovariances and a Measure of Linear Determinism For a Stationary Process, *Journal of Time Series Analysis*, 4, 7987 .

[4] Barndorff-Nielsen, O.E., Shou, G. (1973), On the Parametrization of Autoregressive Models by Partial Autocorrelations, *Journal of Multivariate Analysis*, 3, 408–419.

[5] Beran, J. (1993), Fitting Long-Memory Models by Generalized Linear Regression, *Biometrika*, 80, 4, 817–822.

[6] Bloomfield, P. (1973), An Exponential Model for the Spectrum of a Scalar Time Series, *Biometrika*, 60, 2, 217–226.

[7] Bogert B.P., Healy M.J.R. and Tukey J.W., (1963) The quefrency alanysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking, in Rosenblatt M. Ed. *Proceedings of the Symposium on Time Series Analysis* Chapter 15, 209–243. Wiley, New York.

[8] Box, G.E.P., and Cox, D.R. (1964), An analysis of transformations (with discussion), *Journal of the Royal Statistical Society, B*, 26, 211–246.

[9] Box, G.E.P., and Pierce, D.A. (1970), Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models, *Journal of the American Statistical Association*, 65, 1509-1526.

[10] Brillinger, D.R. (2001), *Time Series. Data Analysis and Theory* , SIAM, Holden Day, San Francisco.

[11] Brockwell, P.J. and Davis, R.A. (1991), *Time Series: Theory and Methods*, Springer-Verlag, New York.

[12] Bühlmann, P. (2006), Boosting for High Dimensional Linear Models, *The Annals of Statistics*, 34, 559–583.

[13] Bühlmann, P. and Hothorn, T. (2007), Boosting Algorithms: Regularization, Prediction and Model Fitting, *Statistical Science*, 22, 477–505.

[14] Cameron, M.A., Turner, T.R. (1987), Fitting Models to Spectra Using Regression Packages, *Journal of the Royal Statistical Society, Series C, Applied Statistics*, 36, 1, 47–57.

[15] Carter, C.K. and Kohn, R. (1997), Semiparametric Bayesian Inference for Time Series with Mixed Spectra, *Journal of the Royal Statistical Society, Series B*, 36, 1, 47–57.

[16] Childers, D.G. Skinner, D.P. Kemerait, R.C. (1977), The Cepstrum: A Guide to Processing, *Proceedings of the IEEE*, 65, 1428-1443.

[17] Chow, Y. and Grenander, U. (1985), A Sieve Method for the Spectral Density, *Annals of Statistics*, 13, 998-1010.

[18] Chiu, T.Y.M., Leonard, T. and Tsui, K-W. (1996), The Matrix-Logarithmic Covariance Model, *Journal of the American Statistical Association*, 91, 198-210.

[19] Cleveland, W.S. (1972), The Inverse Autocorrelations of a Time Series and Their Applications, *Technometrics*, 14, 2, 277–293.

[20] Dahlhaus, R. (1989), Efficient Parameter Estimation for Self Similar Processes. *The Annals of Statistics*, 17, 4, 1749-1766.

[21] Dahlhaus, R. (2012). Locally Stationary Processes. In *Handbook of Statistics*, *Time Series Analysis: Methods and Applications*, Volume, 30, chapter 13, Elsevier, p. 351–408.

[22] Davis, H.T. and Jones, R.H. (1968), Estimation of the Innovation Variance of a Stationary Time Series, *Journal of the American Statistical Association*, 63, 321, 141–149.

[23] Demmler, A. and Reinsch, C. (1975), Oscillation Matrices with Spline Smoothing. *Numerische Mathematik* , 24, 375–382.

[24] Doob, J.L. (1953), *Stochastic Processes*, John Wiley and Sons, New York.

[25] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression (with discussion). *The Annals of Statistics*. 32, 407-499.

[26] Eubank, R.L. (1999), . *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, New York.

[27] Fan, J. and Kreutzberger, E. (1998), Automatic Local Smoothing for Spectral Density Estimation. *Scandinavian Journal of Statistics*, 25, 359–369.

[28] Faÿ, G., Moulines, E. and Soulier, P. (2002), Nonlinear Functionals of the Periodogram, *Journal of Time Series Analysis*, 23, 523–551.

[29] Fokianos, K., Savvides, A. (2008), On Comparing Several Spectral Densities, *Technometrics*, 50, 317–331.

[30] Friedman, J., Hastie, T. and Tibshirani, R. (2000), Additive Logistic Regression: a Statistical View of Boosting, *The Annals of Statistics*, 28, 337- 407.

[31] Früwirth-Schnatter, S. and Früwirth-Schnatter, S. (2007), Auxiliary Mmixture Sampling with Applications to Logistic Models. *Computational Statistics and Data Analysis*, 51, 3509–3528-783.

[32] Geweke, J. and Porter-Hudak, S. (1983), The Estimation and Application of Long-Memory Time Series Models, *Journal of Time Series Analysis*, 4, 221–238.

[33] Gould, H.W. (1974), Coefficient Identities for Powers of Taylor and Dirichlet Series, *The American Mathematical Monthly*, 81, 1, 3–14.

[34] Gradshteyn, I.S. and Ryzhik, I.M. (1994) *Table of Integrals, Series, and Products* Jeffrey A. and Zwillinger D. Editors, Fifth edition, Academic Press.

[35] Gray, H.L, Zhang, N.F. and Woodward, W.A. (1989), On Generalized Fractional Processes, *Journal of Time Series Analysis*, 10, 233–257.

[36] Grenander, U. and Rosenblatt, M. (1957), *Statistical Analysis of Stationary Time Series*, John Wiley and Sons, New York.

[37] Hastie, T., Tibshirani, R. and Friedman, J. (2009), *The Elements of Statistical Learning*, Second Edition, Springer, New York.

[38] Hathaway, D.H. (2010), The Solar Cycle, *Living Reviews in Solar Physics*, 7, 1.

[39] Hosking, J.R.M. (1982), Fractional Differencing, *Biometrika*, 68, 165–176.

[40] Hsu N-J. and Tsai H. (2009), Semiparametric Estimation for Seasonal Long-Memory Time Series Using Generalized Exponential Models, *Journal of Statistical Planning and Inference*, 139, 1992–2009.

[41] Hurvich, C.M. (2002), Multistep forecasting of long memory series using fractional exponential models, *International Journal of Forecasting*, 18, 167–179.

[42] Janaceck, G.J. (1982), Determining the Degree of Differencing for Time Series via the Log Spectrum, *Journal of Time Series Analysis*, 3, 3, 177–183.

[43] Koenker R. and Yoon J. (2009), Parametric Links for Binary Choice Models: A Fisherian-Bayesian Colloquy, *Journal of Econometrics*, 152, 120–130.

[44] Koopmans, L.H. (1974), *The Spectral Analysis of Time Series*, Academic Press.

[45] Li, L.M. (2005), Some Notes on Mutual Information between Past and Future, *Journal of Time Series Analysis*, 27, 309–322.

[46] Luati, A. and Proietti, T. (2010), Hyper-spherical and Elliptical Stochastic Cycles, *Journal of Time Series Analysis*, 31, 169–181.

[47] Luati, A., Proietti, T. and Reale, M. (2012), The Variance Profile, *Journal of the American Statistical Association*, 107, 498, 607–621.

[48] Mardia, K.V. and Jupp, P.E. (2000), *Directional Statistics*, Wiley, Chichester, UK.

[49] McCullagh, T.S. and Nelder, J.A. (1989), *Generalized Linear Models*, Chapman&Hall Cambridge, UK.

[50] McElroy, T.S. and Holan, S.H. (2012), On the Computation of Autocovariances for Generalized Gegenbauer Processes , *Statistica Sinica*, to appear.

[51] Monahan, J.F. (1984), A Note Enforcing Stationarity in Autoregressive-Moving Average Models, *Biometrika*, 71, 2, 403–404.

[52] Moulines, P. and Soulier, E. (1999), Broadband Log-Periodogram Regression of Time Series with Long-Range Dependence, *Annals of Statistics*, 27, 1415–1439.

[53] Narukawa, M. and Matsuda, Y. (2011), Broadband Semi-Parametric Estimation of Long-Memory Time Series by Fractional Exponential Models, *Journal of Time Series Analysis*, 32, 175–193.

[54] Oppenheimen A.V. and Schafer R.W. (2010) *Discrete-Time Signal Processing*, Third Edition. Pearson Eucation, Upper Saddle River.

[55] Pawitan, Y. and O'Sullivan, F. (1994), Nonparametric Spectral Density Estimation Using Penalized Whittle Likelihood, *Journal of the American Statistical Association*, 89, 600–610.

[56] Percival, D.B. and Walden, A.T. (1993), *Spectral Analysis for Physical Applications*, Cambridge University Press.

[57] Pourahmadi M. (1983), Exact factorization of the spectral density and its application to forecasting and time series analysis *Communications in Statistics*, 12, 18, 2085–2094.

[58] Pourahmadi M. (2001), *Foundations of Time Series Analysis and Prediction Theory*, Wiley Series in Probability and Statistics, John Wiley and Sons.

[59] Proietti, T., and Luati , A. (2012), The Generalised Autucovariance Function, *MPRA working paper*, 47311.

[60] Robinson, P. (1991). Nonparametric function estimation for long memory time series. In *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proc. of the 5th Int. Symposium in Economic Theory and Econometrics*, 437–457.

[61] Robinson, P. (1995). Log-Periodogram Regression of Time Series with Long Range Dependence, *Annals of Statistics*, 23, 1048–1072.

[62] Rosen, O., Stoffer, D.S., Wood, S. (2009). Local Spectral Analysis via a Bayesian Mixture of Smoothing Splines, *Journal of the American Statistical Association*, 104, 249–262.

[63] Rosen, O., Stoffer, D.S., Wood, S. (2012). AdaptSPEC: Adaptive Spectral Estimation for Nonstationary Time Series, *Journal of the American Statistical Association*, 107, 1575–1589.

[64] Wahba, G. (1980). Automatic Smoothing of the Log-Periodogram, *Journal of the American Statistical Association*, 75, 122–132.

[65] Walker, A.M. (1964). Asymptotic Properties of Least Squares Estimates of Parameters of the Spectrum of a Stationary Non-Deterministic Time Series, *Journal of the Australian Mathematical Society*, 4, 363–384.

[66] Whittle, P. (1953). Estimation and Information in Stationary Time Series, *Arkiv för Matematik*, 2, 423–434.

[67] Xia Y. and Tong, H. (2011), Feature Matching in Time Series Modeling. *Statistical Science*, 26, 1, 21-46.