

**Statistica Sinica Preprint No: SS-2017-0322**

<b>Title</b>	Generalised Linear Cepstral Models for the Spectrum of a Time Series
<b>Manuscript ID</b>	SS-2017-0322
<b>URL</b>	<a href="http://www.stat.sinica.edu.tw/statistica/">http://www.stat.sinica.edu.tw/statistica/</a>
<b>DOI</b>	10.5705/ss.202017.0322
<b>Complete List of Authors</b>	Tommaso Proietti and Alessandra Luati
<b>Corresponding Author</b>	Tommaso Proietti
<b>E-mail</b>	tommaso.proietti@uniroma2.it

# GENERALIZED LINEAR CEPSTRAL MODELS FOR THE SPECTRUM OF A TIME SERIES

Tommaso Proietti and Alessandra Luati

*University of Rome 'Tor Vergata', CREATES, Aarhus, and University of Bologna*

*Abstract:* This paper introduces a class of generalized linear models with a Box–Cox link for the spectrum of a time series. The Box–Cox transformation of the spectral density is represented as a finite Fourier polynomial. Here, the coefficients of the polynomial, called generalized cepstral coefficients, provide a complete characterization of the properties of the random process. The link function depends on a power-transformation parameter, and can be expressed as an exponential model (logarithmic link), an autoregressive model (inverse link), or a moving average model (identity link). An advantage of this model class is the possibility of nesting alternative spectral estimation methods within the same likelihood-based framework. As a result, selecting a particular parametric spectrum is equivalent to estimating the transformation parameter. We also show that the generalized cepstral coefficients are a one-to-one function of the inverse partial autocorrelations of the process, which can be used to evaluate the mutual information between the past and the future of the process.

*Key words and phrases:* Generalised linear models, Box–Cox link, Whittle likelihood, mutual information.

## 1. Introduction

Analyses of stationary processes in the frequency domain have a long history in time series analyses. The spectral density decomposes the total variation of the process into the contributions of periodic components with different frequencies. It also provides a complete characterization of the serial correlation structure of the process, and, thus, all information needed for linear predictions and interpolations.

Several methods are available for estimating the spectrum. One of the most popular is the Whittle estimation of an autoregressive spectrum, which leads to the solution of an empirical Yule–Walker set of equations (see, e.g., (Percival and Walden, 1993, Chap. 9)). An alternative class of nonparametric estimates is obtained by taking the Fourier transform of a smoothed sample autocovariance function: if a truncated or rectangular smoothing kernel is applied to the autocovariances, this is equivalent to fitting a finite-order moving-average (MA) model using the method of moments. A third popular approach is the exponential model proposed by Bloomfield (1973). This model is derived by truncating the Fourier series expansion of the log-spectrum. The coefficients of the expansion are known as the cepstral coefficients, which are obtained from the discrete Fourier transform of the log-spectrum; their collection forms the cepstrum. This terminology was

## GENERALIZED CEPSTRAL MODELS

---

introduced by Bogert et al. (1963), with cepstral and cepstrum being anagrams of spectral and spectrum, respectively.

The exponential model is a generalized linear model of observations (a periodogram), asymptotically distributed as an exponential random variable. The model adopts a logarithmic link for the mean, which is the spectral density itself.

Refer to Oppenheim and Schaffer (2010, Chap. 13), Brillinger (1981) and Childers et al. (1977) for historical reviews of the cepstrum and its applications in signal processing. Solo (1986) extended the cepstral approach in order to model bivariate random fields. Local likelihood methods using a logarithmic link for the spectral estimation were considered by Fan and Kreutzberger (1998). The exponential model has also played an important role in regularized estimations of the spectrum (Wahba, 1980; Pawitan and O'sullivan, 1994), where smoothness priors are enforced by reducing higher-order cepstral coefficients toward zero, as well as being used to estimate time-varying spectra (Rosen et al., 2009, 2012). The exponential model has also been used for discrimination and clustering of time series, as in Fokianos and Savvides (2008). Holan et al. (2017) generalized the cepstral model to the multivariate case, and Gladish et al. (2014) modeled time-varying spatial error covariance matrices using a cepstral parameteri-

zation.

Against this background, this study contributes to the current literature by introducing a class of generalized linear models with a Box–Cox link, according to which, a linear model is formulated for the Box–Cox transformation ((Box and Cox, 1964)) of the spectral density. The link function depends on a power transformation parameter can be expressed as an exponential model, which corresponds to the case when the transformation parameter is equal to zero. Another special case is the inverse link, which can be used to model the inverse spectrum and, in our setting, is equivalent to an autoregressive estimation of the spectrum. A third special case is the identity link, which amounts to fitting an MA model. The idea of using power or cepstral correlation analyses in time series analyses dates back to Parzen (1992) and, in the context of speech recognition, to Kobayashi and Imai (1984). Here, the goal is to find a scale along which the transformed spectrum has a representation as a finite trigonometric polynomial of small order.

The coefficients of the trigonometric polynomial are related to the generalized autocovariances (Proietti and Luati, 2015), and are termed generalized cepstral coefficients. A likelihood inference is carried out in the frequency domain. To enforce the constraints needed to guarantee the pos-

---

GENERALIZED CEPSTRAL MODELS

---

itivity of the spectral density, we propose a reparameterization of the model based on a set of generalized inverse partial autocorrelations. These autocorrelations enable us to estimate the mutual information between the past and the future of a random process. We also prove the consistency and asymptotic properties of the parameter estimators.

As empirical applications, we examine the Southern Oscillation Index time series and perform a Monte Carlo simulation for a process characterised by a high dynamic range. These applications illustrate the flexibility of the class of generalized linear spectral models and show that an appropriate spectral estimation model (exponential, autoregressive, MA, etc.) can be selected in a likelihood-based framework.

The remainder of the paper is structured as follows. Section 2 introduces the class of generalized linear cepstral models (GLCMs) and discusses its time-series properties. Section 3 discusses the estimation based on the maximization of the Whittle likelihood and derives the asymptotic properties of the corresponding estimator. The empirical applications are provided in Section 4. Finally, in Section 5, we conclude the paper and suggest directions for further research.

## 2. GLCMS FOR THE SPECTRUM

---

### 2. GLCMs for the Spectrum

Let  $\{y_t\}_{t \in T}$  be a stationary real-valued stochastic process,  $T = \{0, \pm 1, \pm 2, \dots\}$ , with mean  $E(y_t) = \mu$  and covariance function  $\gamma_k = \int_{-\pi}^{\pi} e^{i\omega k} dF(\omega)$ ,  $k = 0, \pm 1, \pm 2, \dots$ , where  $F(\omega)$  is the spectral distribution function of the process and  $i$  is the imaginary unit. We assume that the spectral density function of the process exists,  $F(\omega) = \int_{-\pi}^{\omega} f(\lambda) d\lambda$ , the process is regular (Doob, 1953, p. 564) (i.e.  $\int_{-\pi}^{\pi} \ln f(\omega) d\omega > -\infty$ ), and  $\int_{-\pi}^{\pi} f(\omega)^\lambda d\omega < \infty$  for all  $\lambda \in \mathbb{R}$ . We let  $\Gamma_n = \{\gamma_{|s-t|}, s, t = 1, \dots, n\}$  denote the autocovariance matrix of  $y_t$  of order  $n$  (a Toeplitz matrix). The *cepstrum* of the process (Bogert et al., 1963) is defined as the sequence of *cepstral* coefficients

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln[2\pi f(\omega)] \cos(\omega k) d\omega, k = 0, 1, \dots, \quad (2.1)$$

which characterize the Fourier series  $\ln[2\pi f(\omega)] = c_0 + 2 \sum_{j=1}^{\infty} c_j \cos(\omega j)$ .

The Wold representation of  $\{y_t\}_{t \in T}$  is written as  $y_t = \psi(B)\xi_t$ , where  $\{\xi_t\}_{t \in T}$  is a white-noise process with zero mean and finite variance  $\sigma^2$ ,  $\xi_t \sim \text{WN}(0, \sigma^2)$ .

Here,  $2\pi f(\omega) = \sigma^2 |\psi(e^{-i\omega})|^2$ , where  $\psi(z) = 1 + \psi_1 z + \psi_2 z^2 + \dots$ ,  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ .

We consider the Box-Cox transform (Box and Cox, 1964) of the spectral-

## 2. GLCMS FOR THE SPECTRUM

---

generating function  $2\pi f(\omega)$ , with transformation parameter  $\lambda \in \mathbb{R}$ , and

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

We specify the following model for  $g(\omega)$ , under the assumption that it is represented as a finite trigonometric polynomial:

$$g(\omega) = c_{\lambda 0} + 2 \sum_{k=1}^K c_{\lambda k} \cos(\omega k). \quad (2.2)$$

The representation shown in (2.2) is linear in the coefficients  $c_{\lambda k}$ , which are equal to the inverse Fourier transform of the  $g(\omega)$  function:

$$c_{\lambda k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\omega) \cos(\omega k) d\omega, \quad k = 0, 1, \dots, K.$$

When  $\lambda$  is equal to zero, we obtain the Bloomfield (1973) exponential model as a special case. For  $k = 1, \dots, K$ , the coefficients  $c_{0k}$  are equal to the cepstral coefficients given in (2.1). Henceforth we refer to  $c_{\lambda k}$  in (2.2) as the generalized cepstral coefficient at lag  $k$ , and to  $\{c_{\lambda k}, k = 0, 1, \dots, K\}$  as the generalized cepstrum.

The spectral model with a Box-Cox link and mean function

$$f(\omega) = \begin{cases} \frac{1}{2\pi} [1 + \lambda g(\omega)]^{\frac{1}{\lambda}}, & \lambda \neq 0, \\ \frac{1}{2\pi} \exp[g(\omega)], & \lambda = 0, \end{cases} \quad (2.3)$$

is referred to as a GLCM for the spectrum with transformation parameter  $\lambda$  and order  $K$ , or GLCM( $\lambda, K$ ). Except for the case  $\lambda = 0$ , the coefficients

## 2. GLCMS FOR THE SPECTRUM

---

$c_{\lambda k}$  are subject to restrictions to ensure that  $f(\omega)$  in (2.3) is positive. These restrictions are discussed below.

**Remark 1** The GLCM assumes that the spectrum can be parsimoniously represented by a set of finite coefficients, representing the Fourier coefficient of the Box–Cox transformation of the spectrum. Consider, for simplicity, the GLCM(1, 1) model  $2\pi f(\omega) = 1 + c_{10} + 2c_{11} \cos \omega$ . The right-hand side is the spectral density of the first-order moving average (MA) process  $y_t = \xi_t + \psi \xi_{t-1}$ , such that  $1 + c_{10} = \sigma^2(1 + \psi^2)$  and  $c_{11} = \psi_1$ . More generally, the GLCM(1,  $K$ ) model is an MA( $K$ ) model. Conversely, the GLCM( $-1$ ,  $K$ ) process is an autoregressive (AR) process of order  $K$ . For instance, the spectral density of the GLCM( $-1$ , 1) process is  $f(\omega) = (2\pi)^{-1} [1 - c_{-1,0} - c_{-1,1} \cos \omega]^{-1}$ , which is also the spectral density of the AR(1) process  $y_t = \phi y_{t-1} + \xi_t$ , with  $1 - c_{-1,0} = \sigma^{-2}(1 + \phi^2)$ ,  $c_{-1,1} = \phi \sigma^{-2}$ .

**Remark 2** The dynamic range of the spectrum is defined as

$$10 \log_{10} \left( \frac{\max_{\omega} f(\omega)}{\min_{\omega} f(\omega)} \right);$$

see Percival and Walden (1993, sec. 6.3). This is related in the limit to the condition number of the autocovariance matrix  $\Gamma_n$  as  $n$  increases. A process

## 2. GLCMS FOR THE SPECTRUM

---

characterised by a high dynamic range, owing to the presence of spectral peaks, can be generated from a smooth function, such as  $g(\omega)$  given in (2.2), after the transformation (2.3) when the transformation parameter is negative. A short memory process has a finite dynamic range.

The next subsection shows that for stationary Gaussian processes,  $\{c_{\lambda k}, k = 1, \dots, K\}$  contain all information necessary to predict and extract the features of  $y_t$ .

### 2.1 Time-series properties

It follows immediately from (2.3) that  $\lambda$  and the corresponding generalized cepstrum  $\{c_{\lambda k}, k = 1, \dots, K\}$  uniquely characterize the spectral properties of the random process  $\{y_t\}_{t \in T}$ .

For  $\lambda = 0$ ,  $c_{00} = \ln \sigma^2$ , from the Szegő–Kolmogorov formula for the prediction error variance,

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

The coefficients of the Wold representation are obtained using the recursive formula  $\psi_j = j^{-1} \sum_{r=1}^j r c_{0r} \psi_{j-r}$ ,  $j = 1, 2, \dots$ ,  $\psi_0 = 1$ ; see, for example, Pourahmadi (1983).

For  $\lambda \neq 0$ , the generalized cepstral coefficients  $c_{\lambda k}$  are related to the gen-

## 2. GLCMS FOR THE SPECTRUM

---

eralized autocovariance function, introduced by Proietti and Luati (2015),

$$\gamma_{\lambda k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} [2\pi f(\omega)]^{\lambda} \cos(\omega k) d\omega \quad (2.4)$$

by the following relationships:

$$c_{\lambda 0} = \frac{1}{\lambda}(\gamma_{\lambda 0} - 1), \quad c_{\lambda k} = \frac{1}{\lambda} \gamma_{\lambda k}, \quad k \neq 0. \quad (2.5)$$

**Remark 3** For  $\lambda = 1$ ,  $c_{1k} = \gamma_k$ ,  $k > 0$ , the autocovariance function of the process is obtained. In the case  $\lambda = -1$  and  $k \neq 0$ ,  $c_{-1,k} = -\gamma_{i_k}$ , where  $\gamma_{i_k}$  is the inverse autocovariance of  $y_t$  (Cleveland, 1972). The intercept  $c_{\lambda 0}$  for  $\lambda = -1, 0, 1$ , is related to important characteristics of the stochastic process, because  $1/(1 - c_{-1,0})$  is the interpolation error variance,  $\exp(c_{0,0}) = \sigma^2$  is the prediction error variance, and  $c_{1,0} + 1 = \gamma_0$  is the unconditional variance of  $y_t$ .

The coefficients of the Wold and autoregressive representations of the process can be obtained using a spectral factorization that arises naturally after a reparameterization of the generalized cepstral coefficients. In fact,  $[2\pi f(\omega)]^{\lambda} = 1 + \lambda g_{\lambda}(\omega) = \gamma_{\lambda 0} + 2 \sum_{k=1}^K \gamma_{\lambda k} \cos(\omega k)$ , such that, by the representation theorem of Fejér and Riesz for non-negative trigonometric

## 2. GLCMS FOR THE SPECTRUM

polynomials (Grenander and Szegö, 1958, p. 20–21), we can write

$$[2\pi f(\omega)]^\lambda = \sigma_\lambda^2 b(e^{-i\omega})b(e^{i\omega}), \quad b(e^{-i\omega}) = 1 + b_1 e^{-i\omega} + \dots + b_K e^{-i\omega K}. \quad (2.6)$$

According to (2.6), when  $\lambda \neq 0$ , the generalized cepstral coefficients are obtained as

$$c_{\lambda 0} = \frac{1}{\lambda} [\sigma_\lambda^2 (1 + b_1^2 + \dots + b_K^2) - 1], \quad c_{\lambda k} = \frac{1}{\lambda} \sigma_\lambda^2 \sum_{j=k}^K b_j b_{j-k}. \quad (2.7)$$

Moreover,  $\sigma_\lambda^{2/\lambda}$  is the prediction error variance of the process, given as

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln(2\pi f(\omega)) d\omega = \frac{1}{\lambda} \ln \sigma_\lambda^2.$$

The coefficients  $b_k$  can be uniquely determined by imposing the condition that the roots of the polynomial  $b(z) = 1 + b_1 z + \dots + b_K z^K$  lie outside the unit circle, which, in turn, can be enforced by adopting the reparameterization of Barndorff-Nielsen and Schou (1973) and Monahan (1984). Given  $K$  coefficients  $\varsigma_k$ , where  $|\varsigma_k| < 1$ , for  $k = 1, \dots, K$ , which in the present setting, can be interpreted as generalized partial inverse autocorrelations (Bhansali, 1983), the coefficients of the polynomial  $b(z)$  are obtained from the last iteration of the Durbin–Levinson recursion

$$\begin{aligned} b_k^{(k)} &= \varsigma_k \\ b_j^{(k)} &= b_j^{(k-1)} + \varsigma_k b_{k-j}^{(k-1)}, \quad j = 1, \dots, k-1, \end{aligned} \quad (2.8)$$

for  $k = 1, \dots, K$ , positing  $b_j = b_j^{(K)}$ .

## 2. GLCMS FOR THE SPECTRUM

---

The rationale of (2.8) is as follows: let  $y_t^* = b(B)\xi_t^*$ ,  $\xi_t^* \sim \text{WN}(0, \sigma_\lambda^2)$ , such that the right-hand side of (2.6) represents the spectral density of  $y_t^*$ . The inverse process is  $b(B)x_t = \xi_t^*$ . The Durbin–Levinson algorithm maps the first  $K$  autocovariances of  $x_t$  onto the partial autocorrelations of the inverse process and obtains the coefficients  $b_j$  after  $K$  iterations. Equations (2.8) use the same recursions, starting with the partial autocorrelations as inputs, such that the coefficients  $\varsigma_k$ , for  $k = 1, \dots, K$ , are mapped onto the coefficients  $b_j$ , for  $j = 1, \dots, K$ . This parameterization ensures that  $b(z) \neq 0 \iff |z| \leq 1$ . The role of the coefficients  $\varsigma_k$  is discussed further in subsection 2.2.

The coefficients of the Wold representation are obtained recursively, as follows (Gould, 1974):

$$\psi_j = j^{-1} \sum_{r=1}^j \left( r \frac{\lambda + 1}{\lambda} - j \right) b_r \psi_{j-r}, \quad j > 0, \quad \psi_0 = 1.$$

The coefficients of the infinite AR representation are derived similarly. In summary, all information needed for the prediction is available from  $K + 1$  bits of information.

**Remark 4** Fractionally integrated processes (Giraitis et al., 2012) arise as limiting cases. Consider, for instance, the case  $b(z) = 1 - z$ , such that  $[2\pi f(\omega)]^\lambda = \sigma_\lambda^2 |1 - e^{-i\omega}|^2$  is the spectrum of a first-order noninvertible MA

## 2. GLCMS FOR THE SPECTRUM

process. If  $\lambda = -d^{-1}$ ,  $d \in (0, 0.5)$ , then a fractionally integrated process arises, the spectral density of which is unbounded at the origin:  $f(\omega) = \frac{\sigma^2}{2\pi} |2 \sin(\omega/2)|^{-2d}$ . More generally, if  $\lambda < -2$  and  $b(z)$  can be factorized as  $b(z) = (1 - z)b^*(z)$ , with  $b^*(z) = 1 + b_1^*z + \dots + b_{K-1}^*z^{K-1}$ , for  $b^*(z) \neq 0 \iff |z| \leq 1$ , the process is fractionally integrated of order  $d = -\lambda^{-1}$ . When  $b(z) = 1 - 2 \cos \omega_c z + z^2$  and  $\lambda = -2d^{-1}$ ,  $\omega_c \in (0, \pi)$ , the process is a Gegenbauer process; see Hosking (1981) and Gray et al. (1989).

### 2.2 The mutual information between past and future

The following theorem expresses the strong Szegő theorem (see theorem. 6 in Bingham, 2012, and the references therein) in terms of the generalized partial inverse autocorrelations introduced in (2.8).

**Theorem 1.** *Let  $\{y_t\}_{t \in T}$  be a purely nondeterministic Gaussian process with cepstral coefficients  $c_j$ ,  $j = 1, \dots, \infty$ , and generalized cepstral coefficients  $c_{\lambda k}$ ,  $k = 1, \dots, K$ . Then, for  $\lambda \neq 0$ ,*

$$\sum_{j=1}^{\infty} j c_j^2 = \frac{1}{\lambda^2} \sum_{k=1}^K k \ln(1 - \varsigma_k^2), \quad (2.9)$$

where  $\varsigma_{\lambda k}$  are the generalized partial inverse autocorrelations of the process and are related to the generalized cepstral coefficients by equations (2.7) and (2.8).

## 2. GLCMS FOR THE SPECTRUM

---

The term  $\sum jc_j^2$  has several important uses. The mutual information between the past and the future of a Gaussian process is defined as

$$I_{p-f} = \frac{1}{2} \sum_{j=1}^{\infty} jc_j^2. \quad (2.10)$$

Here,  $I_{p-f}$  measures the reduction in uncertainty about the future  $\mathcal{F}_t$ , the sigma-algebra generated by  $\{y_{t+j}, j = 1, 2, \dots\}$ , when the past  $\mathcal{P}_t$ , the sigma-algebra generated by  $\{y_{t-j}, j = 0, 1, 2, \dots\}$ , is known (Ibragimov and Rozanov, 2012; Li, 2006). If  $\{y_t\}_{t \in T}$  is defined on the probability space  $(\Omega, \mathfrak{J}, P)$ ,  $\mathfrak{J} = \mathcal{P}_t \cup \mathcal{F}_t$ , the amount of information about  $\{y_s\}_{s>t}$  contained in the past  $\{y_s\}_{s \leq t}$  is

$$I_{p-f} = \sup \sum P(A_i \cap B_j) \ln \frac{P(A_i \cap B_j)}{P(A_i)P(B_j)}. \quad (2.11)$$

Here, the supremum is taken over all possible finite partitions of  $\Omega$  in the nonintersecting events  $(A_i)_{i=1, \dots, n}, (B_j)_{j=1, \dots, m}$ , such that  $A_i \in \mathcal{P}_t$  for all  $i = 1, \dots, n$  and  $B_j \in \mathcal{F}_t$  for all  $j = 1, \dots, m$ . This is a mutual measure because the expression is symmetric about the past and future events. Heuristically, expression (2.11) measures the maximum expected deviation of the joint distribution of past and future events from the product of their marginal distributions. The equivalence between (2.11) and (2.10) is shown in Ibragimov and Rozanov (2012, Chap. 4).

Here,  $I_{p-f} = 0$  for a Gaussian white noise process, and  $I_{p-f} < \infty$  for an

### 3. WHITTLE LIKELIHOOD ESTIMATION

---

absolutely regular process Ibragimov and Rozanov (2012, Chap. 4). Theorem 1 also shows that  $I_{p-f}$  is infinite if  $|\zeta_k| = 1$ , for some  $k$ , which occurs in the case of fractionally integrated processes (see Remark 4). Finally, according to the strong Szegö limit theorem (Bingham, 2012, p. 305),  $\sum j c_j^2$  is the limit of  $[\ln |\Gamma_n| - n \ln \sigma^2]$  as  $n \rightarrow \infty$ . This can be used to approximate the log-determinant term of the Gaussian likelihood of the sample time series  $\{y_t, t = 1, \dots, n\}$ . Note that the evaluation of  $\sum_{j=1}^{\infty} c_j^2$  is not trivial, because it entails the sum of infinite terms, except for the class of generalized cepstral models. In the latter case, Theorem 1 states that the sum can be computed from the  $K$  inverse partial autocorrelation coefficients, as in (2.9).

#### 3. Whittle likelihood estimation

The main tool used to estimate the spectral density function and its functionals is the periodogram. Let  $\{y_t, t = 1, 2, \dots, n\}$  denote a time series, which is assumed to be a sample realization from a stationary short-memory Gaussian process, characterized by an autocovariance sequence satisfying  $\sum_{k=1}^{\infty} k \gamma_k^2 < \infty$ . In addition, let  $\omega_j = \frac{2\pi j}{n}$ , for  $j = 1, \dots, [n/2]$ , denote the Fourier frequencies, where  $[\cdot]$  denotes the integer part of the argument. The

### 3. WHITTLE LIKELIHOOD ESTIMATION

periodogram, or sample spectrum, is defined as

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

where  $\bar{y} = n^{-1} \sum_{t=1}^n y_t$ . In large samples (Brockwell and Davis, 1991, Chap. 10)

$$\frac{I(\omega_j)}{f(\omega_j)} \sim \text{IID } \frac{1}{2} \chi_2^2, \quad \omega_j = \frac{2\pi j}{n}, \quad j = 1, \dots, [(n-1)/2], \quad (3.12)$$

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

The above distributional results are the basis for the approximate or Whittle maximum likelihood inference for the generalized cepstral model: writing  $f(\omega)$  as in (2.3), and denoting as  $\theta_\lambda = [c_{\lambda 0}, c_{\lambda 1}, \dots, c_{\lambda K}]'$  the vector containing the generalized cepstral coefficients, where  $\theta_\lambda \in \Theta \subset \mathbb{R}^{K+1}$ , the log-likelihood of  $\{I(\omega_j), \text{ for } j = 1, \dots, N = [(n-1)/2]\}$ , is:

$$\ell(\lambda, \theta) = - \sum_{j=1}^N \left[ \ln f(\omega_j) + \frac{I(\omega_j)}{f(\omega_j)} \right]. \quad (3.13)$$

Letting  $z(\omega) = [1, 2 \cos(\omega), 2 \cos(2\omega), \dots, 2 \cos(K\omega)]'$  and writing  $g(\omega_j) = z(\omega_j)' \theta$ , (3.13) can be expressed as  $\ell(\lambda, \theta) = - \sum_{j=1}^N \ell_j(\lambda, \theta)$ , where, for  $1 + \lambda z(\omega_j)' \theta > 0$ ,

$$\ell_j(\lambda, \theta) = \begin{cases} \frac{1}{\lambda} \ln(1 + \lambda z(\omega_j)' \theta) + \frac{2\pi I(\omega_j)}{(1 + \lambda z(\omega_j)' \theta)^{\frac{1}{\lambda}}}, & \lambda \neq 0, \\ z(\omega_j)' \theta + \frac{2\pi I(\omega_j)}{\exp(z(\omega_j)' \theta)}, & \lambda = 0. \end{cases}$$

### 3. WHITTLE LIKELIHOOD ESTIMATION

---

Note that we have excluded the frequencies  $\omega = 0, \pi$  from the analysis; the latter may be included with little effort, and their effect on the inferences is negligible in large samples.

The profile likelihood of the model as  $\lambda$  varies can be used to select the spectral model for  $y_t$ . A similar idea is used by Koenker and Yoon (2009) to select an appropriate link function for binomial data. Let  $\ell_{\max}(\lambda)$  denote the partially maximized (or profile) Whittle likelihood  $\ell_{\max}(\lambda) = \max_{\theta \in \Theta} \ell(\lambda, \theta)$ , or equivalently,  $\ell_{\max}(\lambda) = \ell(\lambda, \tilde{\theta}_\lambda)$ , where  $\tilde{\theta}_\lambda = \operatorname{argmax}_{\theta \in \Theta} \ell(\lambda, \theta)$ . The maximum likelihood estimate (MLE) of the transformation parameter is obtained as the value of  $\lambda$  that maximizes the profile Whittle likelihood.

The truncation parameter,  $K$ , is chosen as the value that minimizes an information criterion, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), given, respectively, by:

$$AIC(K, \lambda) = -2\ell(\lambda, \tilde{\theta}_\lambda) + 2K, \quad BIC(K, \lambda) = -2\ell(\lambda, \tilde{\theta}_\lambda) + \ln(N)K. \quad (3.14)$$

#### 3.1 Reparameterization

The main difficulty with an MLE of the GLCM in the case  $\lambda \neq 0$  is enforcing the condition  $1 + \lambda z(\omega_j)' \theta > 0$ . This problem is well known in the literature on generalized linear models for gamma-distributed observations, for which

### 3. WHITTLE LIKELIHOOD ESTIMATION

---

the canonical link is the inverse link (McCullagh and Nelder, 1989).

The most appropriate solution that ensures the positive definiteness and the regularity of the spectral density is to reparameterize the generalized cepstral coefficients as shown in (2.7). For  $k = 1, \dots, K$ , the parameters  $b_k$  are expressed as a function of the the generalized inverse partial autocorrelations  $|\varsigma_k| < 1$ . In practice, we estimate  $K$  unconstrained real parameters,  $\vartheta_k$ , that are mapped onto the interval  $(-1, 1)$  by the Fisher inverse transformations  $\varsigma_k = \frac{\exp(2\vartheta_k)-1}{\exp(2\vartheta_k)+1}$ , for  $j = 1, \dots, K$ . Furthermore, we set  $\vartheta_0 = \ln(\sigma_\lambda^2)$ .

#### 3.2 Asymptotic properties

In this section, we prove the consistency and asymptotic normality of the Whittle MLE of the vector  $\theta$ , which we denote as  $\tilde{\theta}$ , and the transformation parameter  $\lambda$ . In practice, the transformation parameter is estimated by maximizing the profile likelihood  $\ell_{\max}(\lambda)$ . We denote as  $(\lambda_0, \theta'_0)'$  the true parameter value of  $(\lambda, \theta)'$ .

When  $\lambda_0 \neq 0$ ,  $[2\pi f(\omega)]^\lambda = 1 + \lambda g(\omega)$  and  $g(\omega) = z(\omega)' \theta$ . The asymptotic theory for the Whittle MLE of  $\theta_\lambda$  is based on the following assumptions.

A1.  $\{y_t\}_{t \in \mathcal{T}}$  is a stationary Gaussian process and  $\exists m > 0$ , such that

$$1 + \lambda g(\omega) \geq m, \forall \omega \in [-\pi, \pi].$$

### 3. WHITTLE LIKELIHOOD ESTIMATION

- A2.  $\lambda_0 \in \Lambda \subset \mathbb{R}$  and  $\theta_0 \in \Theta \subset \mathbb{R}^{K+1}$  are the true parameter values,  $(\lambda_0, \theta_0)' = \operatorname{argmax}_{\lambda, \theta} \mathbb{E}\{N^{-1}\ell(\lambda, \theta)\}$ , and  $\Lambda \times \Theta$  is a closed subset of the  $K + 2$  Euclidean space, containing all points  $(\lambda, \theta)'$  for which  $\lambda z(\omega)' \theta > -1, \forall \omega \in [-\pi, \pi]$ .
- A3.  $\{y_t\}_{t \in \mathcal{T}}$  is an absolutely regular process, that is, its cepstral coefficients satisfy  $\sum_{j=1}^{\infty} j c_j^2 < \infty$ .

**Theorem 2.** *Under conditions A1–A3,*

$$\tilde{\lambda} \rightarrow_p \lambda_0, \quad \tilde{\theta} \rightarrow_p \theta,$$

and

$$\sqrt{n} \begin{pmatrix} \tilde{\lambda} - \lambda_0 \\ \tilde{\theta} - \theta_0 \end{pmatrix} \rightarrow_d N(0, V(\lambda_0, \theta_0)),$$

with

$$[V(\lambda, \theta)]^{-1} = \begin{pmatrix} V(\lambda) & V(\lambda, \theta) \\ V(\theta, \lambda) & V(\theta) \end{pmatrix},$$

where

$$V(\lambda) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{1}{[2\pi f(\omega)]^{2\lambda}} [z(\omega)' \theta]^4 [m(\omega)]^2 d\omega,$$

$$V(\lambda, \theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{1}{[2\pi f(\omega)]^{2\lambda}} [z(\omega)' \theta]^2 [m(\omega)] z(\omega) d\omega, \quad V(\theta, \lambda) = V(\lambda, \theta)'$$

$$V(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{1}{[2\pi f(\omega)]^{2\lambda}} z(\omega) z(\omega)' d\omega,$$

and

$$m(\omega) = \frac{\lambda z(\omega)' \theta - (1 + \lambda z(\omega)' \theta) \ln(1 + \lambda z(\omega)' \theta)}{[\lambda z(\omega)' \theta]^2}.$$

### 3. WHITTLE LIKELIHOOD ESTIMATION

---

In the exponential case, when  $\lambda_0 = 0$ , A1 and A2 are not needed, the finiteness of  $\sum_{j=1}^K jc_j^2$  implies  $\sum_{j=1}^{\infty} j\gamma_j^2$ , and the remaining conditions of Theorems II.2.1 and II.2.2 in Dzhaparidze (1986) are fulfilled. Hence, under assumption A2, the Whittle estimates of the transformation parameter and the cepstral coefficients are consistent, and the matrix  $[V(0, \theta)]^{-1}$  has elements

$$V^{(\lambda)} = \frac{1}{16\pi} \int_{-\pi}^{\pi} [z(\omega)' \theta]^4 d\omega, \quad V^{(\theta)} = \frac{1}{4\pi} \int_{-\pi}^{\pi} z(\omega) z(\omega)' d\omega,$$

$$V^{(\lambda, \theta)} = -\frac{1}{8\pi} \int_{-\pi}^{\pi} [z(\omega)' \theta]^2 z(\omega) d\omega.$$

These expressions are obtained by taking the limit for  $\lambda \rightarrow 0$  of  $V^{-1}(\lambda, \theta)$ .

**Remark 5** The restrictions on the parameter space  $\Lambda \times \Theta$  arise from A1, because the set of  $(\lambda, \theta)'$ , for which  $1 + \lambda g(\omega) > 0$ ,  $\forall \omega \in [-\pi, \pi]$ , is the set of points for which  $z(\omega)' \theta > -\lambda^{-1}$  when  $\lambda > 0$ , whereas, for  $\lambda < 0$ ,  $z(\omega)' \theta < -\lambda^{-1}$ . Hence, as  $|\lambda|$  increases, the positivity condition becomes more stringent for the set of admissible  $\theta$ . A1 also rules out the long-memory case; recall from Remark 4 that a fractionally integrated process arises when  $\lambda = -d^{-1}$ , for  $0 < d < 0.5$ , and  $1 + \lambda g(\omega) = 0$  for  $\omega = 0$ .

**Remark 6** The function that maps the partial autocorrelation coefficients to the model parameters is one-to-one and smooth (see Barndorff-Nielsen

---

#### 4. EMPIRICAL EXAMPLES

and Schou, 1973, Theorem. 2), such that the asymptotic properties of the Whittle estimator continue to hold.

#### 4. Empirical examples

Here, we apply the proposed generalizations to time series that have been analyzed extensively in the literature and that provide useful test beds for the class of generalized linear spectral models. In the applications, we use the tapered periodogram  $I(\omega_j) = \frac{1}{2\pi n} |\sum_{t=1}^n h_t(y_t - \bar{y})e^{-i\omega_j t}|^2$ , where  $h_t \geq 0, t = 1, \dots, n$ , and  $\sum_{t=1}^n h_t^2 = 1$ . We use a taper formed for zero-order discrete prolate spheroidal sequences (DPSS), obtained as the eigenvector corresponding to the largest eigenvalue of the matrix  $A(\nu)$ , with elements

$$a_{ij}(\nu) = \begin{cases} \frac{\sin(2\pi\nu(i-j))}{\pi(i-j)}, & \text{for } i \neq j, \\ 2\nu, & \text{for } i = j, \end{cases}$$

for  $i, j = 1, \dots, n$ . The matrix depends on the bandwidth parameter  $\nu$ , which is often set equal to  $\nu = 2/n$ . Refer to Percival and Walden (1993, sec. 3.9 and ch. 7) for further details. Brillinger (1981, Theorem. 5.2.7) showed that for the tapered periodogram, the same distributional result as that given in (3.12) holds.

In the following we refer to the GLCM with transformation parameter  $\lambda$  and order  $K$  as  $\text{GLCM}(\lambda, K)$ .

## 4. EMPIRICAL EXAMPLES

---

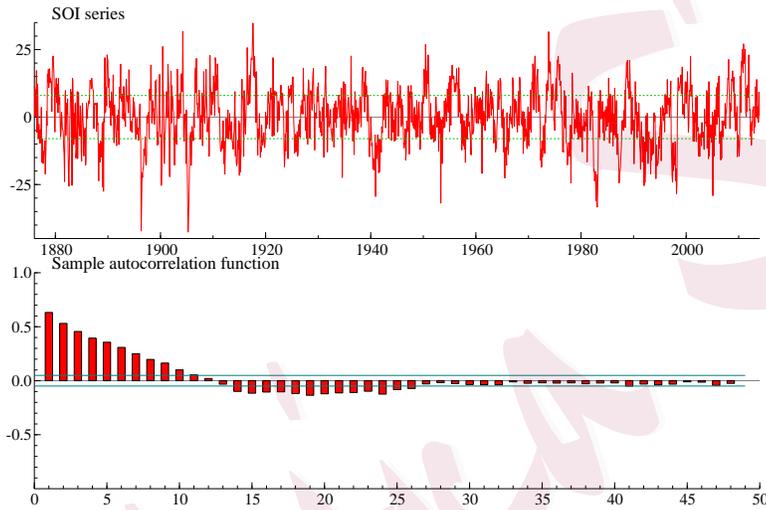
### 4.1 Southern Oscillation Index

The Southern Oscillation Index (SOI) measures the difference in surface air pressure between Tahiti and Darwin, and is an important indicator of the strength of El Niño and La Niña events. Here, values below -8 indicate an El Niño event, and positive values above 8 indicate a La Niña event. The index reflects the cyclic warming (negative SOI) and cooling (positive SOI) of the eastern and central Pacific, which affect the sea-level pressure at the two locations. The monthly series from January 1876 to December 2013 is plotted in Figure 1, along with the autocorrelation function. The series exhibits periodic behavior: often, the El Niño and La Niña episodes alternate, giving the SOI a cyclical feature with an irregular period of about 3–7 years (see, e.g., <http://earthobservatory.nasa.gov>).

We investigate the  $\text{GLCM}(\lambda, K)$  representation that provides the best fit to the sample spectrum of the time series. This depends on two crucial parameters: the truncation parameter  $K$ , and the power parameter  $\lambda$ . These can be selected according to the information criteria given in (3.14). Estimating the  $\text{GLCM}(\lambda, K)$  on a grid of values for  $\lambda$  in the range  $[-2.50, 1.00]$ , with step 0.01, and for  $K$  ranging from 0 to 10, and computing the AIC and BIC criteria, leads to selecting  $K = 7$  and  $\tilde{\lambda} = -2.28$ . Figure 2 displays the prediction error variance and the profile Whittle likelihood of

#### 4. EMPIRICAL EXAMPLES

Figure 1: Southern Oscillation Index. Time series and sample autocorrelation function. In the first plot, the horizontal lines are drawn at  $\pm 8$ .



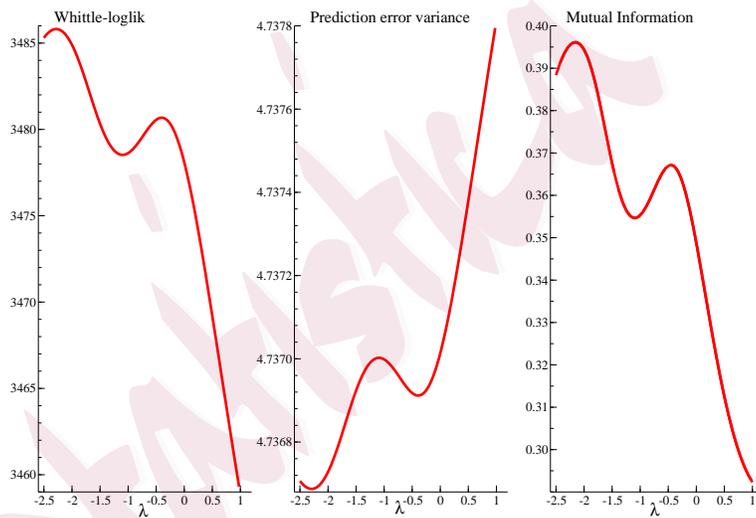
the  $\text{GLCM}(\lambda, 7)$  models as a function of  $\lambda$ , which also shows that the optimal value of the power-transformation parameter is  $\tilde{\lambda} = -2.28$ . The third plot displays the corresponding estimates of  $I_{p-f}$ , which peaks at around  $\lambda = -2$ . The plot illustrates that the pair  $(\lambda, K)$  that minimizes the AIC does not necessarily maximize  $I_{p-f}$ .

An approximate  $(1 - \alpha)\%$  confidence interval for  $\lambda$  can be constructed using the asymptotic distribution of the likelihood ratio test (see Pawitan

#### 4. EMPIRICAL EXAMPLES

---

Figure 2: Southern Oscillation Index. Whittle likelihood, prediction error variance, and mutual Information as a function of  $\lambda$  for GLMS( $\lambda, K$ ) models with  $K = 7$ .



#### 4. EMPIRICAL EXAMPLES

(2013, section 9.11)), for the set of values

$$\left\{ \lambda : \ell_{\max}(\lambda) \geq \ell_{\max}(\tilde{\lambda}) - \frac{1}{2}q_{1-\alpha} \right\},$$

where  $q_{1-\alpha}$  is the  $(1 - \alpha)$ -quantile of the  $\chi_1^2$  distribution. For  $\alpha = 0.05$ ,  $q_{0.95} = 3.84$ .

The estimated spectrum is  $\tilde{f}(\omega) = \frac{1}{2\pi} \left[ \tilde{\sigma}_\lambda^2 \tilde{b}(e^{-i\omega}) \tilde{b}(e^{i\omega}) \right]^{-1/2.278}$ , with  $\tilde{\sigma}_\lambda^2 = 2161.74$  and

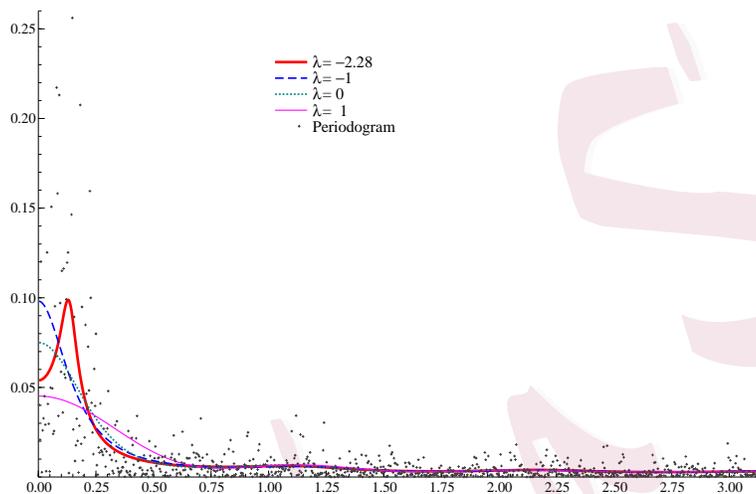
$$\tilde{b}(e^{-i\omega}) = 1 - 1.02e^{-i\omega} - 0.03e^{-i2\omega} - 0.05e^{-i3\omega} - 0.08e^{-i4\omega} + 0.04e^{-i5\omega} - 0.02e^{-i6\omega} + 0.23e^{-i7\omega}.$$

From the second panel of Figure 2, it is evident that the likelihood ratio test of  $\lambda = -2$  for a GLCM( $\lambda, K$ ) model with  $K = 7$  is not statistically significant. As a result, the estimated spectrum does not differ that using an autoregressive model of order 14, such that the autoregressive polynomial is the square of a polynomial of order 7. This polynomial has three pairs of complex conjugate roots and a real root.

Figure 3 plots the periodogram of the SOI series and superimposes the spectral densities fitted by the GLCM( $\lambda, K$ ) model with  $K = 7$ ,  $\lambda$  set equal to 1, 0,  $-1$ , and  $\tilde{\lambda} = -2.28$ . The case when  $\lambda$  is equal to one corresponds to fitting an MA(7) model to the series, whereas the case  $\lambda = 0$  corresponds to fitting Bloomfield's exponential model of order  $K = 7$ ;  $\lambda = -1$  corresponds to fitting an AR(7). Note that in none of these cases does a spectral peak

#### 4. EMPIRICAL EXAMPLES

Figure 3: Southern Oscillation Index. Comparison of the spectral density estimates arising from different  $GLCM(\lambda, K)$  models with  $K = 7$ .



arise at a frequency other than zero. In contrast, The spectrum fitted by the maximum likelihood has a clear mode at a frequency corresponding to a period of about four years.

#### 4.2 Simulated AR(4) process

This example estimates the power spectrum of the AR(4) stochastic process

$$y_t = 2.7607y_{t-1} - 3.8106y_{t-2} + 2.6535y_{t-3} - 0.9238y_{t-4} + \xi_t, \xi_t \sim \text{NID}(0, 1). \quad (4.15)$$

#### 4. EMPIRICAL EXAMPLES

---

The interest here lies in the bimodal nature of the spectral density to be estimated, featuring two peaks located very close together in the frequency range. 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. The closeness of the two modes renders the estimation of the spectrum rather problematic. Thus, this process constitutes a test case for spectral estimation methods; see Percival and Walden (1993).

A sample time series of length  $n = 1,024$  from the above AR(4) process is obtained from Percival and Walden (1993). A detailed analysis is presented here for illustrative purposes.

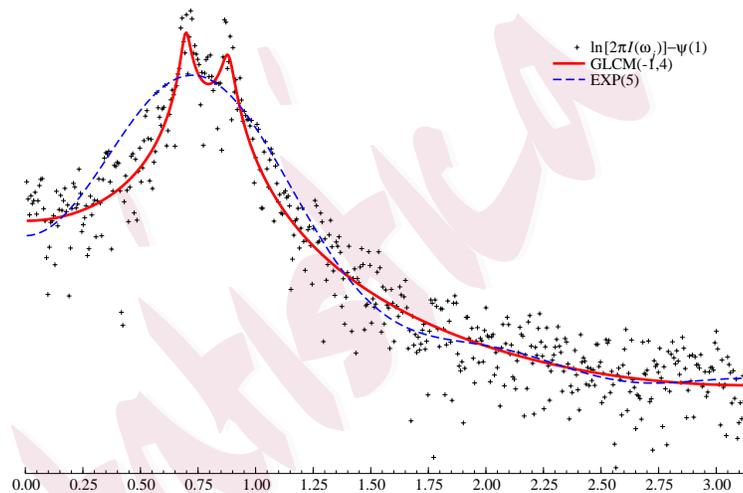
The specifications of the class  $\text{GLCM}(\lambda, K)$  selected by the AIC and BIC differ slightly. While the latter selects the true generating model, that is  $\lambda = -1$  and  $K = 4$ , the AIC selects  $\lambda = -1$  and  $K = 6$ . However, the likelihood ratio test of the null that  $K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

The estimated coefficients of the  $\text{GLCM}(-1, 4)$  model and their estimation standard errors are given as follows:

#### 4. EMPIRICAL EXAMPLES

---

Figure 4: Periodogram and log-spectra estimated using the GLCM(-1, 4), selected by the BIC, and the exponential model with  $K = 5$ , EXP(5).



---

#### 4. EMPIRICAL EXAMPLES

$\tilde{b}_k$	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) highlights that the coefficient estimates are remarkably accurate. Figure 4 displays the centered periodogram and compares the log-spectra fitted by the selected GLCM( $-1, 4$ ) model and the exponential model with  $K = 5$ , which emerges if the Box-Cox transformation parameter is set equal to zero. The latter fit is clearly suboptimal, because it fails to capture the two spectral modes.

In order to evaluate whether the above results are generalizable, we conduct a Monte Carlo experiment in which 5,000 replications of length  $n = 1,024$  are generated independently according to the AR(4) Gaussian process (4.15). For each replication, we estimate a sequence of GLCM( $\lambda, K$ ) models, with  $\lambda$  taking values on a grid from  $-1.75$  to  $1$ , with step  $0.05$ , and for  $K$  ranging from  $0$  to  $8$ . This range of values covers the subset of interest of the parameter space and that in which the maximization of the likelihood by a numerical quasi-Newton algorithm is successful. Model

#### 4. EMPIRICAL EXAMPLES

---

selection is carried out using the AIC and BIC. We record the MLEs of the parameters,  $\tilde{\sigma}_\lambda^2$  and  $\tilde{b}_k$ , for  $k = 1, \dots, K$ , as well as the estimated spectral density.

The main results are presented in Table 1, which reports the summary statistics of the distribution of the MLEs of the GLCM( $\lambda, K$ ) models separately for the two cases when the model selection is carried out using the AIC and the BIC. The summary statistics are the median, mean, standard deviation (St. dev.), and root mean square estimation error (RMSE). The main findings are summarized as follows. Model selection by the BIC yields more accurate estimates of the parameters and the order of the model. The RMSE is systematically smaller, and both the variance and the bias of the estimator are smaller. The distribution of the selection frequency for the order of the GLCM( $\lambda, K$ ) is given as follows:

$K$	0	1	2	3	4	5	6	7	8
AIC	0.00	0.00	0.00	0.00	0.36	0.17	0.14	0.16	0.17
BIC	0.00	0.00	0.00	0.00	0.78	0.16	0.04	0.02	0.01

Finally, the standard error in the estimation of the log-spectrum is computed by averaging across the simulations

$$\left[ \frac{1}{N} \sum_{j=1}^N \left( \ln \tilde{f}(\omega_j) - \ln f(\omega_j) \right)^2 \right]^{1/2},$$

where  $\tilde{f}(\omega_j)$  is the fitted spectral density. These errors are 0.1788 and

---

## 5. CONCLUSIONS

0.1520, respectively, when the model is selected by the AIC and BIC. Therefore, the AIC results in an efficiency loss of about 18% when estimating the log-spectrum of the series.

The overall conclusion is that, despite the differences due to the selection criterion, the  $\text{GLCM}(\lambda, K)$  is an effective spectral estimation method, yielding an autoregressive spectral estimate ( $\lambda = -1$ ) or a neighboring estimate in all cases.

### 5. Conclusions

We have proposed a general frequency domain-estimation framework that nests the exponential model for the spectrum as a special case and allows for any power transformation of the spectrum to be modeled. As a result, alternative spectral fits can be encompassed.

The methods are potentially not robust to outliers; however, a robust periodogram can be used as the basis for inferences. Robustness may occur in the frequency domain, as in McCloskey and Hill (2017), or by applying a data-cleaning algorithm in the time domain, supported by a robust Kalman filter, as in Masreliez and Martin (1977) and Martin and Thomson (1982).

In future research, we would like to apply the GLCM to model the time-varying spectrum of a locally stationary processes (Dahlhaus, 2012)

5. CONCLUSIONS

Table 1: Summary statistics of the Monte Carlo distribution of the MLEs of the parameters of the GLCM( $\lambda, K$ ) model, based on 5,000 replications from the AR(4) model (4.15).

Parameter	AIC				BIC				True values
	Median	Mean	St. dev.	RMSE	Median	Mean	St. dev.	RMSE	
$\tilde{\lambda}$	-1.0000	-1.0242	0.0903	0.0242	-1.0000	-1.0016	0.0305	0.0016	-1
$\tilde{K}$	5.0000	5.5968	1.5177	1.5968	4.0000	4.3072	0.6752	0.3072	4
$\tilde{b}_1$	-2.7581	-2.8169	0.2783	0.0562	-2.7554	-2.7579	0.0888	0.0028	-2.7607
$\tilde{b}_2$	3.8026	4.0037	0.7305	0.1931	3.7974	3.8071	0.2496	0.0035	3.8106
$\tilde{b}_3$	-2.6486	-2.9514	1.0288	0.2979	-2.6388	-2.6576	0.3447	0.0041	-2.6535
$\tilde{b}_4$	0.9224	1.1586	0.7674	0.2348	0.9176	0.9337	0.2547	0.0099	0.9238
$\tilde{b}_5$	0.0000	-0.0847	0.3370	0.0847	0.0000	-0.0060	0.1228	0.0060	0
$\tilde{b}_6$	0.0000	-0.0015	0.1979	0.0015	0.0000	-0.0008	0.0770	0.0008	0
$\tilde{b}_7$	0.0000	0.0034	0.1089	0.0034	0.0000	0.0010	0.0361	0.0010	0
$\tilde{b}_8$	0.0000	0.0048	0.0391	0.0048	0.0000	0.0002	0.0089	0.0002	0
$\tilde{\sigma}_\lambda$	1.0161	1.5414	1.5363	0.5414	1.0044	1.0535	0.5061	0.0535	1

---

## 5. CONCLUSIONS

by allowing the cepstral coefficients to vary over time, for example, with autoregressive dynamics. Finally, a multivariate extension, namely the matrix-logarithmic spectral model for the spectrum of a vector time series, can be investigated, along the lines of the model formulated by Chiu et al. (1996) for covariance structures. A further alternative is the vector exponential model considered recently by Holan et al. (2017).

### Acknowledgments

The authors wish to thank Francesco Battaglia, Victor Solo, Scott Holan, Tucker McElroy, and the referees for their comments, suggestions, and bibliographic references that led to several improvements in the paper. Preliminary versions of this paper circulated under the title *The Exponential Model for the Spectrum of a Time Series: Extensions and Applications*. The authors gratefully acknowledge the financial support of the Italian Ministry of Education, University and Research (MIUR), PRIN Research Project 2010-2011 - prot. 2010J3LZEN, titled “Forecasting economic and financial time series.”. Tommaso Proietti gratefully acknowledges support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

## Appendix

### A. Proof of Theorem 1

The coefficients  $\{\zeta_k, k = 1, \dots, K\}$  are the partial autocorrelations of the process  $b(B)x_t = \xi_t, \xi_t \sim \text{WN}(0, \sigma_\lambda^2)$  (see Bhansali, 1983), whose spectrum  $f_x(\omega)$  is proportional to  $[f(\omega)]^{-\lambda}$ . We denote the cepstrum of the process  $x_t$  by  $\{\kappa_j, j = 0, 1, \dots\}$ ,

$$\kappa_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln[2\pi f_x(\omega)] \cos(\omega j) d\omega.$$

According to the strong Szegő theorem, see Bingham (2012, Theorem 6), applied to the process  $x_t$ ,

$$\prod_{k=1}^K (1 - \zeta_k^2)^{-k} = \exp\left(\sum_{j=1}^{\infty} j \kappa_j^2\right).$$

For  $j > 0$ , the cepstral coefficients of  $x_t$  are related to those of  $y_t$  by  $\kappa_j = -\lambda c_j$ , such that

$$\frac{1}{\lambda^2} \sum_{k=1}^K k \ln(1 - \zeta_k^2) = \sum_{j=1}^{\infty} j c_j^2.$$

### B. Proof of Theorem 2

Under the stated assumptions, Theorems II.2.1 (consistency) and II.2.2 of Dzhaparidze (1986, chapter II, pp. 105 and 109, respectively) hold, as we demonstrate below.

## B. PROOF OF THEOREM 2

Condition A1 ensures that  $1 + \lambda g(\omega) = \gamma_{\lambda 0} + 2 \sum_{k=1}^K \gamma_{\lambda k} \cos(\omega k)$  is positive. In fact,  $1 + \lambda g(\omega)$  is a proper spectral density and, under A1, the generalized autocovariances  $\{\gamma_{\lambda k}\}$  form a positive definite sequence. In addition, because  $g(\omega) = z(\omega)' \theta$ ,  $1 + \lambda g(\omega)$  is a smooth and symmetric function of  $\omega \in [-\pi, \pi]$ . Assumption A2 states that the true parameter vector is interior to the parameter space. Note that our model is identified, that is  $(\lambda_1, \theta_1)' \neq (\lambda_2, \theta_2)' \rightarrow f_1(\omega) \neq f_2(\omega)$  for almost all  $\omega$  and  $(\lambda_1, \theta_1)$ , for  $(\lambda_2, \theta_2) \in \Lambda \times \Theta$ , where  $f_r(\omega)$  denotes  $[1 + \lambda z(\omega)' \theta_r]^{1/\lambda_r} / (2\pi)$ ,  $r = 0, 1$ .

Assumption A3 implies that the autocovariances of  $y_t$  satisfy the condition  $\sum_{k=1}^{\infty} k \gamma_k^2 < \infty$ ; see Li (2006), Theorem 1. The summability condition on the squared autocovariances, along with  $f(\omega) > 0$ , by Corollary I.3.1 in Dzhaparidze (1986, p. 66) guarantees that the principal part of the Gaussian log-likelihood can be approximated by the Whittle likelihood.

The derivatives of the inverse of the spectral density are, respectively,

$$\begin{aligned} \frac{\partial}{\partial \lambda} f(\omega)^{-1} &= -\frac{1}{f(\omega)} \frac{1}{1 + \lambda z(\omega)' \theta} [z(\omega)' \theta]^2 m(\omega), \\ \frac{\partial}{\partial \theta} f(\omega)^{-1} &= -\frac{1}{f(\omega)} \frac{1}{1 + \lambda z(\omega)' \theta} z(\omega). \end{aligned}$$

Under the stated assumptions,  $1 + \lambda z(\omega)' \theta > 0$  and

$$\lim_{\lambda \rightarrow 0} \begin{pmatrix} \frac{\partial}{\partial \lambda} f(\omega)^{-1} \\ \frac{\partial}{\partial \theta} f(\omega)^{-1} \end{pmatrix} = -\frac{1}{f(\omega)} \begin{pmatrix} -[z(\omega)' \theta]^2 / 2 \\ z(\omega) \end{pmatrix}, \quad f(\omega) = \exp(z(\omega)' \theta) / (2\pi).$$

Hence, the derivatives exist and are continuous in  $(\lambda, \theta)$ . Hence, by Theorem

## B. PROOF OF THEOREM 2

II.2.1 in Dzhaparidze (1986, p. 105),  $\tilde{\lambda} \rightarrow_p \lambda_0$  and  $\tilde{\theta} \rightarrow_p \theta_0$ .

Furthermore, the spectral density  $f(\omega)$  is a twice-differentiable function of  $\lambda$  and  $\theta$ , and the second derivatives are

$$\begin{aligned}\frac{\partial^2 f(\omega)}{\partial \lambda^2} &= f(\omega) \frac{[z(\omega)' \theta]^3}{[1 + \lambda z(\omega)' \theta]^2} \left\{ z(\omega)' \theta [m(\omega)]^2 - \frac{1 + 2[1 + \lambda z(\omega)' \theta] m(\omega)}{\lambda z(\omega)' \theta} \right\}, \\ \frac{\partial^2 f(\omega)}{\partial \lambda \partial \theta} &= f(\omega) \frac{z(\omega)' \theta}{[1 + \lambda z(\omega)' \theta]^2} [z(\omega)' \theta m(\omega) - 1] z(\omega), \\ \frac{\partial^2 f(\omega)}{\partial \theta \partial \theta'} &= \frac{f(\omega)}{[1 + \lambda z(\omega)' \theta]^2} (1 - \lambda) z(\omega) z(\omega)'.\end{aligned}$$

Under the stated assumptions,  $1 + \lambda z(\omega)' \theta > 0$ , the second partial derivatives are continuous in  $\omega$ . Furthermore,

$$\begin{aligned}\lim_{\lambda \rightarrow 0} \frac{\partial^2 f(\omega)}{\partial \lambda^2} &= f(\omega) [z(\omega)' \theta]^3 \left[ \frac{1}{4} z(\omega)' \theta + \frac{2}{3} \right], \\ \lim_{\lambda \rightarrow 0} \frac{\partial^2 f(\omega)}{\partial \lambda \partial \theta} &= -f(\omega) [z(\omega)' \theta] \left[ \frac{1}{2} z(\omega)' \theta + 1 \right] z(\omega), \\ \lim_{\lambda \rightarrow 0} \frac{\partial^2 f(\omega)}{\partial \theta \partial \theta'} &= f(\omega) z(\omega) z(\omega)',\end{aligned}$$

where  $f(\omega) = \exp(z(\omega)' \theta) / (2\pi)$ .

Letting  $\delta = (\lambda, \theta)'$ , the score vector and the Hessian matrix associated with the Whittle likelihood in (3.13) are, respectively,

$$\begin{aligned}\mathcal{S}(\delta) &= \sum_{j=1}^N \left( \frac{I(\omega_j)}{f(\omega_j)} - 1 \right) \frac{\partial \ln f(\omega)}{\partial \delta}, \\ \mathcal{H}(\delta) &= - \sum_{j=1}^N \left( \frac{I(\omega_j)}{f(\omega_j)} - 1 \right) \frac{\partial^2 \ln f(\omega)}{\partial \delta \partial \delta'} - \sum_{j=1}^N \frac{I(\omega_j)}{f(\omega_j)} \frac{\partial \ln f(\omega)}{\partial \delta} \frac{\partial \ln f(\omega)}{\partial \delta'}.\end{aligned}$$

GENERALISED CEPSTRAL MODELS

Hence, the information matrix  $\mathcal{I}(\alpha) = E[-\mathcal{H}(\alpha)]$  is such that

$$\frac{1}{N}\mathcal{I}(\delta) \rightarrow_p \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial \ln f(\omega)}{\partial \delta} \frac{\partial \ln f(\omega)}{\partial \delta'} d\omega,$$

where

$$\frac{\partial \ln f(\omega)}{\partial \delta} = \begin{pmatrix} \frac{\partial \ln f(\omega)}{\partial \lambda} \\ \frac{\partial \ln f(\omega)}{\partial \theta} \end{pmatrix} = \frac{1}{1 + \lambda z(\omega)' \theta} \begin{pmatrix} [z(\omega)' \theta]^2 m(\omega) \\ z(\omega) \end{pmatrix}.$$

Thus, by Theorem II.2.2 in Dzhaparidze (1986, p. 109),  $\tilde{\lambda}$  and  $\tilde{\theta}$  are asymptotically normal, as stated in the theorem with the asymptotic covariance matrix  $V(\lambda_0, \theta_0)$ , where

$$V(\lambda, \theta) = \left[ \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial \ln f(\omega)}{\partial \delta} \frac{\partial \ln f(\omega)}{\partial \delta'} d\omega \right]^{-1}.$$

## References

- Barndorff-Nielsen, O. and Schou, G. (1973). On the parametrization of autoregressive models by partial autocorrelations. *Journal of multivariate Analysis*, 3(4):408–419.
- Bhansali, R. (1983). The inverse partial correlation function of a time series and its applications. *Journal of Multivariate Analysis*, 13(2):310–327.
- Bingham, N. (2012). Szegő’s theorem and its probabilistic descendants. *Probability Surveys*, 9.
- Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika*, 60(2):217–226.

---

## REFERENCES

- Bogert, B. P., Healy, M., and Tukey, J. (1963). The quefrency analysis of time series for echoes: Cepstrum pseudo-autocovariance, cross-cepstrum, and saphe cracking. *Time series analysis*, pages 209–243.
- Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252.
- Brillinger, D. R. (1981). *Time series: data analysis and theory*, volume 36. Siam.
- Brockwell, P. J. and Davis, R. A. (1991). *Time series: theory and methods*. Springer Series in Statistics.
- Childers, D. G., Skinner, D. P., and Kemerait, R. C. (1977). The cepstrum: A guide to processing. *Proceedings of the IEEE*, 65(10):1428–1443.
- Chiu, T. Y., Leonard, T., and Tsui, K.-W. (1996). The matrix-logarithmic covariance model. *Journal of the American Statistical Association*, 91(433):198–210.
- Cleveland, W. S. (1972). The inverse autocorrelations of a time series and their applications. *Technometrics*, 14(2):277–293.
- Dahlhaus, R. (2012). Locally stationary processes. *Handbook of statistics*, 30:351–412.
- Doob, J. L. (1953). *Stochastic processes*, volume 7. Wiley New York.
- Dzhaparidze, K. (1986). *Parameter estimation and hypothesis testing in spectral analysis of stationary time series*. Springer Science & Business Media.
- Fan, J. and Kreutzberger, E. (1998). Automatic local smoothing for spectral density estimation.

## REFERENCES

---

- Scandinavian Journal of Statistics*, 25(2):359–369.
- Fokianos, K. and Savvides, A. (2008). On comparing several spectral densities. *Technometrics*, 50(3):317–331.
- Giraitis, L., Koul, H. L., and Surgailis, D. (2012). *Large sample inference for long memory processes*. World Scientific Publishing Company.
- Gladish, D., Wikle, C., and Holan, S. (2014). Covariate-based cepstral parameterizations for time-varying spatial error covariances. *Environmetrics*, 25(2):69–83.
- Gould, H. (1974). Coefficient identities for powers of taylor and dirichlet series. *American Mathematical Monthly*, pages 3–14.
- Gray, H. L., Zhang, N.-F., and Woodward, W. A. (1989). On generalized fractional processes. *Journal of Time Series Analysis*, 10(3):233–257.
- Grenander, U. and Szegő, G. (1958). *Toeplitz forms and their applications*, volume 321. Univ of California Press.
- Holan, S. H., McElroy, T. S., and Wu, G. (2017). The cepstral model for multivariate time series: The vector exponential model. *Statistica Sinica*, 27:23–42.
- Hosking, J. R. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.
- Ibragimov, I. A. and Rozanov, Y. A. (2012). *Gaussian random processes*, volume 9. Springer Science & Business Media.
- Kobayashi, T. and Imai, S. (1984). Spectral analysis using generalized cepstrum. *IEEE trans-*

---

## REFERENCES

- actions on acoustics, speech, and signal processing*, 32(5):1087–1089.
- Koenker, R. and Yoon, J. (2009). Parametric links for binary choice models: A fisherian–bayesian colloquy. *Journal of Econometrics*, 152(2):120–130.
- Li, L. M. (2006). Some notes on mutual information between past and future. *Journal of Time Series Analysis*, 27(2):309–322.
- Martin, R. D. and Thomson, D. J. (1982). Robust-resistant spectrum estimation. *Proceedings of the IEEE*, 70(9):1097–1115.
- Masreliez, C. and Martin, R. (1977). Robust bayesian estimation for the linear model and robustifying the kalman filter. *IEEE transactions on Automatic Control*, 22(3):361–371.
- McCloskey, A. and Hill, J. B. (2017). Parameter estimation robust to low-frequency contamination. *Journal of Business & Economic Statistics*, pages 1–13.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall,.
- Monahan, J. F. (1984). A note on enforcing stationarity in autoregressive-moving average models. *Biometrika*, 71(2):403–404.
- Oppenheim, A. V. and Schafer, R. W. (2010). *Discrete-time signal processing*. Pearson Education, Upper Saddle River.
- Parzen, E. (1992). Time series, statistics, and information. In Parzen, E. e. a., editor, *New Directions in Time Series Analysis*, pages 265–286. Springer Verlag: New York.
- Pawitan, Y. (1993). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*.

---

## REFERENCES

- Oxford University Press.
- Pawitan, Y. and O’Sullivan, F. (1994). Nonparametric spectral density estimation using penalized whittle likelihood. *Journal of the American Statistical Association*, 89(426):600–610.
- Percival, D. B. and Walden, A. T. (1993). *Spectral analysis for physical applications*. Cambridge University Press.
- Pourahmadi, M. (1983). Exact factorization of the spectral density and its application to forecasting and time series analysis. *Communications in Statistics-Theory and Methods*, 12(18):2085–2094.
- Proietti, T. and Luati, A. (2015). The generalised autocovariance function. *Journal of Econometrics*, 186(1):245–257.
- Rosen, O., Stoffer, D. S., and Wood, S. (2009). Local spectral analysis via a bayesian mixture of smoothing splines. *Journal of the American Statistical Association*, 104(485):249–262.
- Rosen, O., Wood, S., and Stoffer, D. S. (2012). Adaptspec: Adaptive spectral estimation for nonstationary time series. *Journal of the American Statistical Association*, 107(500):1575–1589.
- Solo, V. (1986). Modeling of two-dimensional random fields by parametric cepstrum. *IEEE transactions on information theory*, 32(6):743–750.
- Wahba, G. (1980). Automatic smoothing of the log periodogram. *Journal of the American Statistical Association*, 75(369):122–132.

## REFERENCES

---

Dipartimento di Economia e Finanza, Università di Roma 'Tor Vergata', Via Columbia 2, 00133

Rome, Italy, and CREATES - Center for Research in Econometric Analysis of Time Series,

Aarhus, Denmark.

E-mail: [tommaso.proietti@uniroma2.it](mailto:tommaso.proietti@uniroma2.it)

Dipartimento di Scienze Statistiche 'Paolo Fortunati', Via Belle Arti 41, 40126 Bologna, Italy.

E-mail: [alessandra.luati@unibo.it](mailto:alessandra.luati@unibo.it)