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

Tommaso Proietti                 Alessandra Luati
University of Sydney             University of Bologna

February 7, 2012- Preliminary and incomplete

Abstract

The exponential model is based on the Fourier expansion of the spectral density of a stationary process; when coupled with the sampling properties of the periodogram, a generalized linear model for gamma distributed data arises. After discussing the merits and the limitations of the exponential model for the spectrum of a stationary time series, with particular reference to the analysis of long memory processes, we propose two main generalisations: firstly, we propose maximum L_q likelihood estimation for the identification of spectral peaks; secondly, we consider replacing the logarithmic link with a more general Box-Cox link, which encompasses the identity, the logarithmic and the inverse link. From the empirical standpoint, we consider the case for pooling the periodogram and we discuss the problem of separating the long memory component from the short one.
Contents

1 Introduction 3

2 cepStral alAnysis and spCtal Analysis 4
   2.1 ARMA and Echo processes ............................................. 5
   2.2 Fractional Exponential (FEXP) processes .............................. 7

3 Exponential Models and Log-Periodogram Regression 8

4 Approximate Maximum Likelihood Estimation 9

5 Illustrations 10
   5.1 Box and Jenkins Series A .................................................. 10
   5.2 U.S. Gross Domestic Product growth ................................. 12

6 Pooling the periodogram 13

7 Maximum \( L_q \)-Likelihood Estimation 17

8 Generalized Spectral Models with Box-Cox Link 20
   8.1 Illustrations ............................................................. 21
   8.2 The Treatment of Missing Observations and the Sensitivity of the Spectral Estimates ... 22

9 Separating long memory from short memory 22

10 Bayesian Model Selection for the EXP model 26

11 Conclusions 29
1 Introduction

The Exponential Model (EXP model, a.k.a. *cepStral alAnysis*) provides a flexible and effective semiparametric tool for the estimation of the spectrum of a time series. It is based on the Fourier expansion of the logarithm of the spectral density function of a stationary stochastic process. The exponential model is most suitable 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.

Indeed, the idea of taking the logarithm of the spectral density function of a process has been developed for the analysis of series that are contaminated by echoes (Bogert, Healey and Tuckey, 1963), namely seismological data, whose spectral densities typically factorize as the product of two components. One component is the spectrum of the original series and the other is the contribution to the spectrum of the echo. To be more specific, if a series, \( z_t \), obeys the dynamic equation \( z_t = y_t + \alpha y_{t-k} \), where \( y_t \) is a stationary process, then its spectral density function, \( f_z(\omega) \), is the product \( f_z(\omega) = f_y(\omega)(1 + 2\alpha \cos(\omega k) + \alpha^2) \). To isolate the contribution of the echo, Bogert, Healy and Tuckey (1963) developed what they called cepstral analysis, based on the discrete Fourier transform of the logarithm of the spectral density function, that they called the cepstrum. The interpretation is direct: a peak at a certain point in the abscissa \( k \) (quefrency) of the cepstrum suggests the existence of an echo with the corresponding delay in the original series. The new terminology, which includes the words cepstrum, cepstral analysis and quefrency, was motivated to distinguish the time domain of the original series from the discrete-time index that is the argument of the cepstrum. We refer the reader to Oppenheim and Schafer (2010 ch. 13), Brillinger (2002) and Childers, Skinner and Kemerait (1977) for historical reviews on the cepstrum and its applications in signal processing.

Bloomfield (1973) exponential model originated from considering a finite number of terms in the Fourier expansion of the log-spectrum. Robinson (1991) and Beran (1993) generalised Bloomfield’s EXP model to the long memory case.

Maximum likelihood estimation of these models is based on Whittle (1951, 1952), Walker (1963), Cameron and Turner (1987), and Narukawa and Matsuda (2011). Among the uses of the EXP model we mention discrimination and clustering of time series, see Fokianos and Savvides (2008).

This paper aims at

- Illustrating the potential of cepStral alAnysis, also for long memory time series.
- Establishing a role for periodogram pooling.
- Extending the analysis to general power transformations of the spectrum.
- Introducing Bayesian cepStral alAnysis based on log-periodogram generalized regression.
The structure of the paper is as follows. Section 2 introduces the exponential model, with reference to ARMA and echo processes, and the fractional exponential model. Sections 3 and 4 consider cepstral estimation by log-periodogram regression and approximate maximum likelihood, respectively. Section 5 introduces some illustrative examples. Finite sample estimation using the pooled periodogram is considered in section 6, while the maximum $L_q$ likelihood estimation and the generalized regression model with Box-Cox link are developed in sections 7 and 8, respectively.

2 cepStral alAnysis and speCtAl Analysis

Let $\{y_t\}_{t \in T}$ be a stationary zero-mean stochastic process indexed by a discrete time set $T$, with covariance function $\gamma_k = \int_{-\pi}^{\pi} e^{i\omega k} dF(\omega)$, 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, $F(\omega) = \int_{-\pi}^{\pi} f(\lambda) d\lambda$, 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,

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

(1)

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

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

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 $\gamma_k = \int_{-\pi}^{\pi} f(\omega) \exp(\imath \omega k) d\omega$. The cepstral coefficients are related to the mutual information between the past and the future (Li, 2005) ... from which the relation between the cepstral coefficients and the partial autocorrelations follows.

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 \( \varphi(B) = 1 + \varphi_1 B + \varphi_2 B^2 + \ldots, \sum_j |\varphi_j| < \infty, \xi_t \simWN(0, \sigma^2) \), 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

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

with \( \varphi_0 = 1 \), or, equivalently,

\[
\varphi_j = j^{-1} (c_1 \varphi_{j-1} + 2 c_2 \varphi_{j-2} + \cdots + j c_j), \quad j = 1, 2, \ldots.
\]

The derivation, see Janacek (1982), Pourahmadi (1983) and Hurvich (2002), is based on the spectral factorization \( f(\omega) = \sigma^2 \varphi(e^{-i\omega}) \varphi(e^{i\omega}) \); setting the factor \( \varphi(z) = \exp\left(\sum_{k=1}^\infty c_k z^k\right) \) and equating the derivatives of both sides with respect to \( z \) at the origin gives the above result. For instance, if \( c_k = 0, k > 1 \), \( \varphi_j = (j!)^{-1} c_1, j > 0 \), 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) and that of a moving average (MA) process (no cutoff at a lag equal to the order).

The autoregressive representation \( \pi(B) y_t = \xi_t \), where \( \pi(B) = \sum_{j=0}^\infty \pi_j B^j = \varphi(B)^{-1} \) has \( \pi_0 = 1 \) and

\[
\pi_j = -j^{-1} \sum_{r=1}^j r c_r \pi_{j-r}, \quad j = 1, 2, \ldots.
\]

### 2.1 ARMA and Echo processes

If \( y_t \) is a white noise (WN) process, \( c_k = 0, k > 0 \). Figure III displays the cepstra of the AR(1) process \( y_t = \phi y_{t-1} + \xi_t, \xi_t \simWN(0, \sigma^2) \) with \( \phi = 0.9 \) and coefficients \( c_k = \phi^k/k \) (top left plot). 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 \( \theta = -0.9 \) are obtained by reversing the first 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 bottom plots concern the cepstra of two pseudo-cyclical processes: the AR(2) process \( y_t = 1.25 y_{t-1} - 0.95 y_{t-2} + \xi_t \) with complex roots, and the ARMA(2,1) process \( y_t = 1.75 y_{t-1} - 0.95 y_{t-2} + \xi_t + 0.5 \xi_{t-1} \). 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.
Figure 1: Cepstral coefficients $c_k, k = 1, \ldots, 20$ for selected ARMA models

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 factorized in terms of their roots

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

(see Brockwell and Davis, 1991, section 4.4),

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

where

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

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, $a_j = r^{-1}e^{i\lambda}, \overline{a}_j = r^{-1}e^{-i\lambda}$, their contribution to the cepstrum is via the coefficients $r^k \cos(\lambda k)/k$.

The cepstrum of an echo process $y_t + \alpha y_{t-m}$ has coefficients

\[
c_k = \frac{(-\alpha)^{k/m}}{k/m}.
\]
at lags (also referred as quefrecies) \( k = m, 2m, 3m, \ldots \), and \( c_k \) elsewhere, where \( c_k \) are the cepstral coefficients of the process \( y_t \).

As it can be argued from the cepstral analysis of ARMA and echo processes, there is very little that the analysis of the coefficients \( c_k \) adds to the analysis of the autocovariance function for the identification of a particular structure. The cepstra are in fact characterized by a pattern similar to the autocovariance function, converging to zero more rapidly, due to the factor \( 1/k \). However, the class of truncated EXP processes with is not embedded in the

### 2.2 Fractional Exponential (FEXP) processes

As stated in the introduction, cepstral analysis is most suitable for long memory processes (Beran, 1993), whose spectral density can be written as \(|2 \sin(\omega/2)|^{-2d} f(\omega)\), where the first factor is the Fourier transform of the fractional difference operator \((1 - B)^{-d}\), \(|1 - e^{-\omega}|^{-2d} = |2 \sin \frac{\omega}{2}|^{-2d}\), and \( f(\omega) \) is the spectral density function of a short memory process. The process is stationary if and only if \( d < 0.5 \).

The logarithm of the spectral generating function is thus linear in \( d \) and in the cepstral coefficients associated with \( \ln 2\pi f(\omega) \)

\[
c_0 + \sum_k c_k \cos(\omega k) - 2d \ln |2 \sin \frac{\omega}{2}|
\]

where \( c_0 \) retains its link to the p.e.v. \( \sigma^2 = \exp(c_0) \) as \( \int_{-\pi}^{\pi} \ln |2 \sin \frac{\omega}{2}| d\omega = 0. \)

For a fractional noise (FN) process, \((1 - B)^d y_t \sim \text{NID}(0, \sigma^2)\), the cepstral coefficients show an hyperbolic decline:

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

For more general processes the cepstral coefficients are additively decomposable as \( c_k^* + c_k \).

Notice also that

\[
- \ln |2 \sin(\omega_j/2)| = \sum_{k=1}^{\infty} \frac{\cos(\omega k)}{k}
\]

(see also Gradshteyn and Ryzhik, 1994, formula 1.441.2) which tends to infinity when \( \omega \to 0 \).

For long memory seasonal processes the cepstrum is linear in the memory parameters \((1 - B)^d \prod (1 - 2 \cos \lambda_j B + B^2)^{d_j} y_t = x_t\), where \( x_t \) is a short-memory process with spectral density \( f(\omega) \). The factor \( 1 - 2 \cos \lambda_j B + B^2 \) is known as a Gegenbauer polynomial and features a pair of complex conjugate roots with unit modulus and phase \( \lambda_j \). Each contributes to the cepstrum via the term

\[
-2d_j \ln \left| 4 \sin \frac{\omega + \lambda_j}{2} \sin \frac{\omega - \lambda_j}{2} \right|.
\]
3 Exponential Models and Log-Periodogram Regression

The class of models considered in this paper is based on Bloomfield (1973), who proposed truncating the Fourier representation of $\ln 2\pi f(\omega)$ to the $K$ term (EXP$(K)$ model):

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

Due to the sampling properties of the periodogram, a generalized linear model for gamma distributed errors with logarithmic link can be formulated for the spectral analysis of a time series. We start by reviewing the sampling distribution of the periodogram.

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

The periodogram (sample spectrum) is defined as

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

where $\bar{y} = n^{-1} \sum_t y_t$.

The following large sample distributional result holds (Koopmans, 1974, ch. 8):

$$\frac{I(\omega_j)}{f(\omega_j)} \sim \text{IID } 1/2 \chi^2_2, 0 < \omega_j < \pi$$  \hspace{1cm} (3)

whereas $\frac{I(\omega_j)}{f(\omega_j)} \sim \chi^2_1, \omega_j = 0, \pi$, where $\chi^2_m$ denotes a chi-square random variable with $m$ degrees of freedom.

When the model contains the long memory component

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

will be referred to as the FEXP$(K)$ model.

The strength of the approach lies in the linearity in the cepstral coefficients, which in turn have a convenient interpretation.

Unbiased, but inefficient, estimates of the parameters $c_k$ can be obtained by applying ordinary least squares to the log-periodogram regression:

$$\ln[2\pi I(\omega_j)] - \psi(1) = c_0 + \sum_{k=1}^{K} c_k \cos(\omega k) + \epsilon_j, \quad \omega_j = \frac{2\pi j}{n}, j = 1, \ldots, \left\lfloor \frac{n-1}{2} \right\rfloor.$$  \hspace{1cm} (5)

In fact $\epsilon_j$ is a centered log-chisquare variable and where $\psi(\cdot)$ is the digamma function ($\psi(1) = -\gamma$, where $\gamma = 0.57722$ is Euler’s constant and $\text{Var}(\epsilon_j) = \psi'(1) = \pi^2/6$). 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 \(-2 \ln |2 \sin \frac{\omega}{2}|\). An improved bias-reduced estimator based on log-periodogram regression is analysed in Andrews and Guggenberger (2003).

The unknown coefficients can be estimated by (approximate) maximum likelihood: in fact, the previous distributional results states that in large samples \(I(\omega_j)/f(\omega_j), j = 1, \ldots, [(n - 1)/2]\) are independent exponential random variables with unit mean (notice that we exclude the frequencies \(\omega = 0, \pi\) from the above representation).

4 Approximate Maximum Likelihood Estimation

Estimation by maximum likelihood of the EXP model see Cameron and ... For the FEXP The asymptotic theory presented in Narukawa and Matsuda (2011).

Denote by \(z_j\) the vector of explanatory variables at frequency \(\omega_j\) and by \(\theta\) the unknown coefficients, e.g. in the long memory case \(z'_j = (1, 2 \cos \omega_j, 2 \cos(2\omega_j), \ldots, 2 \cos(K\omega_j), -2 \ln |2 \sin(\omega_j/2)|), \theta' = (c_0, c_1, c_2, \ldots, c_K, d)\), 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 \(f(\omega_j)\), or equivalently, \(I(\omega_j)/f(\omega_j) \sim \frac{1}{2} \chi^2_2\), the log-likelihood of \(\{I(\omega_j), j = 1, \ldots, n^* = [(n - 1)/2]\}\), is:

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

where

\[
\ell_j(\theta) = z'_j \theta + \frac{2\pi I(\omega_j)}{\exp(z'_j \theta)}.
\]

The score vector and the Hessian matrix are

\[
g(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = -\sum_j z_j u_j, u_j = 1 - \frac{2\pi I(\omega_j)}{\exp(z'_j \theta)}
\]

\[
H(\theta) = \frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'} = -\sum_j W_j z_j z'_j, W_j = \frac{2\pi I(\omega_j)}{\exp(z'_j \theta)}
\]

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

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

\[
\tilde{\theta}_{i+1} = \tilde{\theta}_i - [H(\tilde{\theta}_i)]^{-1} g(\tilde{\theta}_i)
\]
or by the method of scoring:

\[ \hat{\theta}_{i+1} = \hat{\theta}_i + [I(\hat{\theta}_i)]^{-1} g(\hat{\theta}_i) \]

with starting value \( \hat{\theta}_0 = (\sum_z z'_j\hat{\theta}_i)^{-1} \sum_j z_j (\ln[2\pi I(\omega_j)] + \gamma) \). In the former case, \( \hat{\theta}_{i+1} \) is obtained equivalently by weighted least squares regression of \( z'_j \hat{\theta}_i - u_j/W_j \) on \( z_j \) with weights \( W_j, j = 1, \ldots, n^* \), where both \( u_j \) and \( W_j \) are evaluated at \( \hat{\theta}_i \). See Cameron and Turner (1987).

The estimator is asymptotically normal and efficient (Dahlhaus, 1989)

\[
\sqrt{n}(\theta - \hat{\theta}) \Rightarrow_d N(0, V), \quad V^{-1} = 2 \lim_{n \to \infty} \sum_{j=1}^{n^*} z_j z'_j = \frac{1}{4\pi} \int_{-\pi}^{\pi} z(\omega)z(\omega)'d\omega
\]

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

\[
V^{-1} = \begin{bmatrix}
\frac{1}{2} & 0 & 0 & \ldots & 0 & 0 \\
0 & 1 & 0 & \ldots & 0 & 1 \\
0 & 0 & 1 & \ddots & 0 & \frac{1}{2} \\
\vdots & \vdots & 0 & \ddots & \vdots & \vdots \\
0 & 0 & \ldots & \ldots & 1 & \frac{1}{K} \\
0 & 1 & \frac{1}{2} & \ldots & \frac{1}{K} & \frac{2}{\pi^2}
\end{bmatrix}
\]

Notice that the MLE of \( d \) is asymptotically independent of the intercept, but it is correlated with that of the cepstral coefficients (asymptotically, \( \text{corr}(2 \cos(\omega k), -2 \ln |2 \sin(\omega/2)|) = \frac{\sqrt{6}}{\pi K} \approx 0.78 \).

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 \( \frac{2\pi I(\omega_j)}{\exp(z_j')} - 1 = -u_j \).

Notice that we assume stationarity throughout. For the nonstationary case see Shimotsu and Phillips (2005) and the references therein.

## 5 Illustrations

### 5.1 Box and Jenkins Series A

Our first empirical illustration deals with *Box and Jenkins (BJ) Series A*, a sequence of 200 chemical process concentration readings. The series, plotted in figure 28 along with its first differences, was investigated in the original paper by Bloomfield (1973), with the intent of comparing ARMA models with the \( \text{EXP}(K) \).

Bloomfield fitted a model with \( K \) chosen so as to match the number of ARMA parameters. BJ (1970) had fitted an ARMA(1,1) model to the levels and an AR(1) to the differences. The estimated p.e.v. is 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 \( \sigma^2 \) out of the
likelihood function. He found that the estimated prediction error variance (p.e.v.) resulted in 0.146 and 0.164, respectively. He concluded that ARMA models are more flexible.

First and foremost, 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 carried out by IRLS, AIC selects an EXP(7) for the levels and an EXP(5) for the 1st differences. The p.e.v. estimates are 0.099 and 0.097. BIC selects an EXP(3) in both cases. The p.e.v. estimates are 0.103 and 0.103. The FEXP(0) provides an excellent fit: the $d$ parameter is estimated equal to 0.437 (s.e. 0.058), and the p.e.v. is 0.100.

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

When first differences are considered, the FEXP(0) applied to 1st differences yields the $d$ parameter is estimated 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 the possibility that the series has long memory, which was not explored in the literature.
5.2 U.S. Gross Domestic Product growth

Figure 4 displays the quarterly U.S. Gross Domestic Product (GDP) growth rate (in percentage points). When an ARMA(2,2) model is fitted by frequency domain maximum likelihood, the estimated AR polynomial features two complex roots which imply a peak in the fitted spectrum at frequency 0.68, corresponding to a period of 9.3 quarters. The estimated p.e.v. is 0.82.

Luati and Proietti (2010) identify two stochastic stationary cycles corresponding to the two spectral modes of figure 5. The first and minor mode corresponds to a cycle with a longer period of about four years (16 quarters), whereas the second higher mode is defined at the frequency identified by the ARMA model.

When an EXP model is selected according to AIC of BIC, an EXP(2) is selected, and as it can be evinced from figure 5, the fitted spectrum is unimodal with the mode located at the zero frequency. Hence, no pseudo-cyclical feature is identified. The estimated p.e.v. is 0.84. We need to set $K = 5$ to estimate a spectrum similar that implied by the ARMA model (p.e.v. reduces to 0.83). Furthermore, to be able a tale of two cycles we need to impose a very high truncation parameter, $K = 12$, which is clearly overfitting (p.e.v. 0.79).

This example illustrates that the EXP model may be less sensitive to cyclical features. A possible explanation lies with the logarithmic link embedded in the model specification: in fact, the logarithmic transform attenuates the separation of high periodogram ordinates from the others and boosts that of small ones. In a later section we are going to explore the use of alternative link functions based on the Box-Cox (Box and
Cox, 1964) power transformation.

One way to improve the resolution of the spectral fit at the low frequency is to use a weighted likelihood approach: a possibility in this direction is offered by maximum $L_q$ likelihood estimation, which revolves around a transformation of $L_j(\theta)$ that emphasize the fit of high periodogram ordinates with low density.

6 Pooling the periodogram

Bloomfield (1973) discusses the estimation of the exponential spectral model using the pooled periodogram ordinates over blocks of consecutive Fourier frequencies. The evidence he presents, based on a Monte Carlo experiment is far from conclusive. Pooling for log-periodogram regression is considered in Moulines and Soulier (1999). Fay, Moulines and Soulier (2002) illustrate its role when the data are non Gaussian.

Assuming that $n^* = Mm$, where $M$ denotes the number of adjacent blocks and $m$ is the number of Fourier frequencies in each block, if the spectrum is constant in each block, then, approximately

$$\sum_{k=1}^{m} 2\pi I(\omega_{jm+k}) = 2\pi f \left( \omega_{jm+\frac{m+1}{2}} \right) X_{j,m}$$

with $X_{j,m}$ representing a $G(m, 1)$ random variable, or chisquare with $2m$ degrees of freedom.

Recalling that $X_{j,m} \sim \text{Gamma}(m, 1)$ has probability density function $f(x) = [\Gamma(m)]^{-1}x^{m-1}e^{-x}$, the
Figure 5: Spectrum and log-spectrum estimation for U.S. GDP growth using an exponential model with $K$ selected by AIC (same as BIC) and $K = 15$.

Log-likelihood associated with the observed $\{\sum_{k=1}^{m} 2\pi I(\omega_{jm+k}), j = 0, \ldots, M\}$ is

$$L_m(\theta) = \kappa(m) - m \sum_{j=0}^{M} \left[ z_{j,m}^T \theta + \frac{2\pi I(\omega_{jm})}{\exp(z_{j,m}^T \theta)} \right]$$

(8)

where

$$\kappa(m) = -M \ln \Gamma(m) + (m - 1) \sum_{j=0}^{M-1} \ln \left[ \sum_{k=1}^{m} 2\pi I(\omega_{jm+k}) \right]$$

is the periodogram average over the range of frequency (the Daniell type estimate of the spectral density at the frequency $\omega_{jm} = \omega_{jm+m+1}$), and $z_{j,m}$ is the vector

$$z_{j,m}^T = (1, 2 \cos \omega_{jm}, 2 \cos(2\omega_{jm}), \ldots, 2 \cos(K\omega_{jm}), -2 \ln[2 \sin(\omega_{jm}/2)]) .$$

The maximum likelihood estimator of $\theta$ is obtained by iteratively reweighted least squares using the Newton-Raphson or scoring algorithms outlined in the previous section. In the former case we iterate

$$\tilde{\theta}_{i+1} = \tilde{\theta}_i - [H_m(\tilde{\theta}_i)]^{-1} g_m(\tilde{\theta}_i)$$
until convergence, with starting value obtained by log-periodogram regression,

\[ \hat{\theta}_0 = (\sum_j z_{j,m} z'_{j,m})^{-1} \sum_j z_{j,m} (\ln[2\pi \bar{I}(\omega_{j,m})] - \psi(m)) , \]

and

\[ g_m(\theta) = \frac{\partial}{\partial \theta} L_m(\theta) = -m \sum_j z_{j,m} u_{j,m}, \]

\[ H_m(\theta) = \frac{\partial^2}{\partial \theta \partial \theta'} L_m(\theta) = -m \sum_j W_{j,m} z_{j,m} z'_{j,m}, \]

Notice that \( \psi(m) = E \ln(X_{j,m}) \).

Hence, the estimation of the model under periodogram pooling poses no new difficulties. We need however to establish in what situations pooling leads to an improvement in the estimation of the spectrum.

In general, \( \text{MSE}(\hat{\theta}) \) and quantities like

\[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ \ln[2\pi f(\omega)] - z(\omega)\hat{\theta} \right]^2 d\omega \]  

are very complicated functions of \( m, n \) and \( f(\omega) \). Note that for an EXP model (9) equals \( \text{MSE}(\hat{c}_0) + 2 \sum_k \text{MSE}(\hat{c}_k) \), i.e. the Euclidean distance between the true and the fitted spectrum is a linear combination of the mean square estimation error of the cepstral coefficients.

To illustrate this point we replicate Bloomfield’s simulation study, according to which autoregressive series of length \( n = 128 \) are generated according to \( y_t = \phi y_{t-1} + \xi_t, \xi_t \sim \text{NID}(0,1) \), with \( \phi = 0.5 \). We estimate the EXP model by maximum likelihood.

Table II reports the the trace of the \( \text{MSE}(\hat{\theta}) \) matrix versus \( m \) for EXP(\( K \)) models with \( K \) ranging from 0 to 6. It can be seen that the MSE is not a monotonic function of \( m \). In particular, for a given \( K \), it is a minimum for \( m = 1 \) and we observe two local minima at \( m = 3 \) and \( m = 7 \). There is an explanation for this puzzling behaviour: in fact, for \( m = 3, 7 \), \( M = n^*/m \) exactly, where \( n^* = [(n-1)/2] = 63 \) (or, equivalently, \( n^* = mM \)). In this case the MSE is equal or very similar to that for the case with no pooling \( (m = 1) \); the frequencies \( \omega_{j,m}, j = 0 \) are such that the \( z_{j,m} \) are a quasi-orthogonal set.

In general, given a sample size \( n \), the mean square estimation error will be very similar to the case \( m = 1 \) for all those \( m \) values for which \( n^* \) factorizes as \( n^* = mM \) with integer \( M \). The similarity depends on how fast the true spectrum \( f(\omega) \) varies. For a WN process the MSE will be exactly the same across those \( m \) values.

If we consider the case \( n = 1000 \), then \( n^* = 499 \) is a prime number. There is no \( m \) that leads to the same MSE as \( m = 1 \), not even in the WN case. Figure IX displays the \( \text{MSE}(\hat{c}_0) \) as a function of \( m \), when

\[ ^1 \text{There are notable differences with Bloomfield experiment: first, he somewhat strangely considered hanning, rather than pooling, the periodogram; secondly, he fitted the EXP model by log-periodogram regression rather than maximum likelihood.} \]
Table 1: Simulated AR(1) process with $n = 128$ and $\phi = 0.5$. $100 \times \text{tr} \ [\text{MSE}(\hat{\theta})]$ as a function of $m$ and $K$.

<table>
<thead>
<tr>
<th>$K$</th>
<th>$m = 1$</th>
<th>$m = 2$</th>
<th>$m = 3$</th>
<th>$m = 4$</th>
<th>$m = 5$</th>
<th>$m = 6$</th>
<th>$m = 7$</th>
<th>$m = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>36.26</td>
<td>36.84</td>
<td>36.26</td>
<td>38.13</td>
<td>38.13</td>
<td>38.13</td>
<td>36.26</td>
<td>41.14</td>
</tr>
<tr>
<td>1</td>
<td>4.26</td>
<td>4.32</td>
<td>4.26</td>
<td>4.46</td>
<td>4.46</td>
<td>4.27</td>
<td>4.95</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3.54</td>
<td>3.63</td>
<td>3.54</td>
<td>3.85</td>
<td>3.86</td>
<td>3.55</td>
<td>4.72</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4.28</td>
<td>4.42</td>
<td>4.28</td>
<td>4.79</td>
<td>4.81</td>
<td>4.81</td>
<td>4.30</td>
<td>6.81</td>
</tr>
<tr>
<td>4</td>
<td>5.20</td>
<td>5.38</td>
<td>5.20</td>
<td>5.98</td>
<td>6.01</td>
<td>6.05</td>
<td>5.23</td>
<td>10.57</td>
</tr>
<tr>
<td>5</td>
<td>6.18</td>
<td>6.42</td>
<td>6.18</td>
<td>7.36</td>
<td>7.44</td>
<td>7.47</td>
<td>6.20</td>
<td>20.89</td>
</tr>
<tr>
<td>6</td>
<td>7.17</td>
<td>7.48</td>
<td>7.18</td>
<td>8.90</td>
<td>8.97</td>
<td>9.15</td>
<td>7.21</td>
<td>72.39</td>
</tr>
</tbody>
</table>

$K = 1$ and $K = 4$ (left panels). In the former case the EXP model is grossly misspecified, whereas $K = 4$ is the order that optimizes the bias variance trade-off; notice, in fact, the coefficients $c_k = 0.5^k/k$ are less than 0.01 for $k > 4$.

There are several features that are worth noticing. First and foremost, that the MSE is not a smooth function of $m$; secondly, if the focus of the estimation is the prediction error variance and the EXP model is seriously misspecified, as in the case $K = 1$, then periodogram pooling has some role: the first panel shows in fact that $\text{MSE}(\hat{c}_0)$ is minimized for $m = 25$, whereas the right top panel shows that the $\text{MSE}(\hat{\theta})$ is flat for $m < 10$. The bottom panels, referring to the EXP(4) model, illustrate, however, that when the misspecification is less, periodogram pooling leads to an increase in the MSE.

In FEXP modelling, our experience is that the case for periodogram pooling is even smaller, due to the fact that the spectrum increases very rapidly in the vicinity of the zero frequency, so that the assumption of a constant spectrum in each block of $m$ frequencies is not tenable, with the consequence that increasing $m$ for a fixed $n$ induces large biases. There is still room for choosing $m > 1$ in the presence of model misspecification. For instance, if $y_t \sim FN$ with $d = -0.2$, $n = 1000$ and an EXP(0) is estimated (as if the process were WN), the MSE of the p.e.v. estimator with $m = 20$ is 88% that arising for $m = 1$.

In conclusion, there appears to be little scope for pooling, except in the case of serious misspecification. Nevertheless, pooling is still beneficial in the non Gaussian case (Hurvich, Moulines and Soulier, 2002). Moreover, when the sample is very large and $n^*$ is not prime, setting $m$ equal to a small factor of $n^*$ can render model estimation and selection feasible. The precision is about the same if $f(\omega)$ is slowly varying.

In the long memory case, the possibility of defining $m(\omega)$ as a function of $\langle \omega \rangle$ (e.g. setting it to a small value at the low frequency and a higher one at the high frequencies) has yet to be explored.

Finally, as we shall see later, periodogram pooling is required for estimating generalized linear spectral model with Box-Cox link.
Figure 6: AR(1) process with $n = 1000$ and $\phi = 0.5$. MSE of $\tilde{c}_0$ and $\text{tr} \{\text{MSE} (\tilde{\theta})\}$ (logarithms) as a function of $m$.

7 Maximum $L_q$-Likelihood Estimation

Ferrari and Yang (2010) have introduced the maximum $L_q$-likelihood estimator as the maximizer of a power transform of the density of the observations. The aim is that of emphasising different features of the density and the result is that of a weighted likelihood estimator where the weights are given by powers of the density. This method can be readily adapted to cepstral estimation.

Let $I_j = I(\omega_j)$, $f_j = f(\omega_j)$, $p_j = \frac{1}{f_j} \exp(-I_j/f_j)$ (the exponential density of $I_j$). The $L_q$ transform of $p_j$ is

$$L_q(p_j) = \frac{p_j^{1-q} - 1}{1 - q}.$$  

This can be viewed as the Box-Cox transform of the density $p_j$. The real parameter $q$ measures the degree of distortion. If $q = 1$, $L_q(p_j) = \ln p_j$.

The $L_q$ likelihood function is

$$\ell_q(\theta) = \sum_{j=1}^{n^*} L_q(p_j)$$
and the \( L_q \)-MLE is defined as 
\[
\hat{\theta}_q = \arg\max \ell_q(\theta)
\]
The \( L_q \)-likelihood equations are
\[
\frac{\partial \ell_q(\theta)}{\partial \theta} = n^* \sum_{j=1} g_j w_j, \quad g_j = \frac{\partial}{\partial \theta} \ln p_j = z_j u_j, w_j = p_j^{1-q}
\]
\[
\frac{\partial^2 \ell_q(\theta)}{\partial \theta \partial \theta'} = \sum_{j=1} w_j \left[ H_j + (1 - q) g_j g_j' \right], \quad H_j = \frac{\partial}{\partial \theta \partial \theta'} \ln p_j = W_j z_j z_j'
\]
The case \( q = 1 \) yields the usual Whittle likelihood and the gradient and Hessian are as given above (\( w_j = 1 \)).

If we ignore the second term in the Hessian, the \( i \)-th update of the Newton-Raphson is a weighted update. The weights depend on the power transformation of the density.

The \( ML_q \)E arises as the solution to a weighted likelihood equation and as such it has the interpretation as an extended likelihood (XT-likelihood) discussed in Xia and Tong (2011). A necessary and sufficient condition for asymptotic normality of \( \hat{\theta}_q \) is that the distortion parameter \( q \) as a function of the sample size converge to one when \( n \) tends to infinity, with the same asymptotic efficiency as the MLE. In the case of the exponential distribution, the asymptotic mean square error as a function of \( q \) is available in close form (see Ferrari and Yang, 2010, p.764) and can therefore be maximised with respect to \( q \) in order to get an optimality criterion to select the parameter \( q \).

Figure 7 illustrates the results of fitting an EXP(K) model by \( ML_q \)E. The top panel reproduces the fitted spectrum obtained from the standard MLE estimator (the EXP(2) model being selected by both AIC and BIC). The weights attached to the different frequencies are uniform. The central panel considers the case \( q = 1.5 \): larger weights are attached to large periodogram ordinates which are clustered in the low frequency portion. The model selected by BIC has the order unchanged and does not capture a cycle. However, the AIC selects an high order and the fitted spectrum has several spectral modes. When \( q \) is raised further to 1.75, the weights become more concentrated at the high periodogram ordinates with low density and the fit captures the cyclical features of the periodogram. BIC selects an EXP(5) model which accounts for spectral peak corresponding to a cycle with a period of 10 quarters.

\( ML_q \)E seem to provide an effective strategy for fitting the EXP model to cyclical series. The way we have used it (\( q > 1 \)) aims at reducing the bias in spectral estimation at the expenses of an enlarged variance. Notice that if \( q \) is less than zero, the trade-off would be reversed and the spectral estimates would be less sensitive to the cyclical features. In our application we have a strong a-priori for values of \( q \) greater than 1, but it may be felt that the selection of this parameter is an inferential issue. A possible strategy is using frequency domain cross-validation, which amount to choosing \( q \) so as to minimize the sum of the squared cross-validatory Pearson’s residuals:
\[
\frac{2\pi I(\omega_j)}{\exp(z_j' \hat{\theta}_{\{\setminus j\}})} - 1
\]
where \( \hat{\theta}_{\{\setminus j\}} \) is the \( ML_q \)E not using the \( j \)-th periodogram ordinate.
It should perhaps be finally remarked that the fit obtained by ML$_q$E is still a global fit: only the estimation of the EXP model parameters uses the information of the sample spectrum selectively. In fact, it may be still thought that the spectrum implied by the EXP(12) model is overfitting the data.

The EXP and FEXP models lend themselves to local fitting: for an assigned frequency $\omega$, the spectrum at $\omega$ can be estimated by maximising the local likelihood

$$\ell(\omega; \theta) = - \sum_{j=1}^{n^*} \left[ \ln f(\omega_j) + \frac{I(\omega_j)}{f(\omega_j)} K \left( \frac{\omega - \omega_j}{\delta} \right) \right]$$

where $K(\cdot)$ is a kernel (e.g. Epanechnikov) and $\delta \leq \pi$ is the bandwidth parameter.

The parameters $\theta$ can be estimated by a weighted scoring algorithm, similar to the one dealt with in section 4. Local likelihood estimation is a popular estimation method: see Cleveland (17) and Fahrmeir and Tutz (2001). For spectral estimation see Fan and Gijbels (1996, section 6.4.2.), who however focus on the local linear model $\ln 2\pi f(\omega) = \beta_0 + \beta_1 \omega$, rather than on a local trigonometric model.

Another possibility is to make $K$ depend on $\omega$. This parameter, in fact, regulates the flexibility of the spectral fit (a larger $K$ is suitable when the spectrum is rapidly changing, e.g. in the proximity of a spectral peak).
8 Generalized Spectral Models with Box-Cox Link

Another reason why the EXP or FEXP model is not sensitive to the cyclical properties of the spectrum is the use of logarithmic link for the exponential or gamma model for the periodogram ordinates. It is interesting to observe that the canonical link for Gamma distributed data is the inverse link.

Any continuous monotonic transform of the spectral density function can be expanded as a Fourier series. A transformation of the spectral density function that encompasses the logarithmic case is the Box-Cox transform. This yields a generalized regression model with a Box-Cox link, which encompasses the EXP and FEXP models.

Let us consider the Fourier expansion of a power transform of the spectral density function,

\[ [2\pi f(\omega)]^\lambda = c_{\lambda,0} + \sum_{k=1}^{\infty} c_{\lambda,k} \cos(\omega k) \]

with coefficients given by the relation

\[ c_{\lambda,k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} [2\pi f(\omega)]^\lambda \exp(i\omega k) d\omega. \]

When \( \lambda \to 0 \), the log-transform is considered, i.e.

\[ c_{0,k} = c_k. \]

For \( \lambda = 1 \), \( c_{1,k} = \gamma_k \), the autocovariance function of the process is obtained. For \( \lambda = -1 \), \( c_{-1,k} = \gamma_k i \), the inverse autocovariance. The intercept still plays a role since, in the cases \( \lambda = -1, 0, 1 \), \( c_{-1,0} = \gamma_0 \), \( c_{0,0} = \sigma^2 \), and \( c_{1,0} = \gamma_0 \).

The log-likelihood associated with the observed \( \{\sum_{k=1}^{m} 2\pi I(\omega_{jm+k}) \}, j = 0, \ldots, M \} \) is

\[ L_{m,\lambda}(\theta) = \kappa(m) - m \sum_{j=0}^{M} \left[ \lambda^{-1} \ln(z_{j,m}^\prime \theta_\lambda) + \frac{2\pi I(\omega_{jm})}{(z_{j,m}^\prime \theta_\lambda)^{1/\lambda}} \right] \]

where \( z_{j,m}^\prime \theta_\lambda = [2\pi f(\omega)]^\lambda \).

Maximum likelihood estimation of \( \theta \) can be carried out by iteratively reweighted least squares using the method of scoring.

The main difficulty with this approach is enforcing \( z_{j,m}^\prime \theta_\lambda > 0 \). This problem is typical, for instance of the exponential or gamma model with inverse link. In our case several strategies may be useful for overcoming this problem: i. periodogram pooling, which reduces the influence of periodogram ordinates close to zero; ii. weighting the periodogram, so as to exclude in the estimation those frequencies at which the positivity constraint is violated; iii. rescaling the series by a multiplicative factor.
Figure 8: BJ Series A: Spectrum, Log-Spectrum and Inverse spectrum for $\lambda = -1$ and $m = 3$. The sample spectra are defined as $(2m\pi\hat{I}(\omega))^\lambda\Gamma(m)/\Gamma(m + \lambda)$, where the multiplicative factor ensures that the expected value is $[2\pi f(\omega)]^\lambda$.

8.1 Illustrations

In the U.S. GDP case, $\lambda = 0$ was considered in section 5. We need to increase the value of $\lambda$ to be able to capture as cyclical feature. The order selected by both AIC and BIC tend to increase with $\lambda$: for $\lambda = -1$ both orders are equal to 1. On the contrary, setting $\lambda = 2$ both criteria select the order $K = 5$ and the fitted spectrum displays a cyclical feature.

Our second example deals with the estimation of the inverse spectrum for BJ series A. Setting the transformation parameter $\lambda = -1$ yields a gamma generalized linear model with an inverse link. Notice that we need to take at least $m \geq 3$ (i.e. pooling is necessary in the case $\lambda = -1$). The model selected by the two criteria has $K = 1$. Figure 8 displays the spectrum, log spectrum and inverse spectrum fitted by MLE and and compares it with the the spectrum and log spectrum fitted using the logarithmic link.

One important characteristic of the specification with inverse link is that it enables the estimation of the interpolation error variance (IEV) using the reciprocal of the intercept $c_{-1,0}$. The estimated IEV is 0.078, in our case, which amounts to 49% of the unconditional variance. As a result, the index of linear determinism proposed by Battaglia and Banshali (1987), is equal to 0.51, whereas the p.e.v. is 64% of the variance.
8.2 The Treatment of Missing Observations and the Sensitivity of the Spectral Estimates

The generalized gamma spectral model can be used to provide a simple iterative algorithm for the treatment of missing observations. The algorithm alternates two steps, an imputation step and an estimation step:

1. Imputation step: letting $\rho_{ik}$ denote the inverse correlation function,

$$\hat{y}_{t; \tau} = -\sum_{k=1}^{K} \rho_{ik} y_{t+k}$$

Using the cepstral coefficients, the inverse correlations are easily obtained from the AR weights $\pi(B)$. Notice also that the $\rho_{ik}$ coefficients are immediately available from the generalized gamma model with inverse link.

2. Estimation step: estimate the spectral density and the inverse spectrum replacing the missing observations by their imputed values $\hat{y}_{t; \tau}$.

The same strategy can be used to assess the influence of a particular observation on the estimates of the p.e.v. and other features, and more generally on the spectral estimates at a particular frequency.

The above algorithm is similar to the EM algorithm. To implement the latter, however, we need to evaluate the expectation of $I(\omega_j)$ given the available sample and the estimation step refers to the expected complete data likelihood, where the expectation is taken with respect to the distribution of the missing data given the available data, which is assumed to be Gaussian. The implementation of the expectation step of the EM algorithm can be difficult in the presence of an unknown mean, and perhaps its evaluation by Monte Carlo simulation is the most straightforward solution.

9 Separating long memory from short memory

In this section we argue that too much emphasis has been misplaced on the issue of estimating the long memory parameter. We assume that our interest lies in estimating accurately the spectrum or a transformation thereof. We argue, by looking at the sampling properties of the FEXP parameter estimators and by empirical examples, that there is some exchangeability between the cepstral coefficients and the long memory parameter and that separating long memory from short is problematic, to say the least.

Consider the ARFIMA(1, 0.4, 0) process for $\phi = 0.4$. The log spectrum is made up of a long memory component and the cepstral coefficients associated to the two are $c_k^* = 0.4/k$ and $c_k = 0.4^k/k$, respectively. Those of the ARFIMA process are obtained from their sum. We simulated 10,000 series of length 1000 from such a process and for each replication we estimated a FEXP($K$) model, for $0 \leq K \leq 5$. The estimation results are reported below:
In this case, the value of $K$ that minimizes the mean square estimation error for $d$ concides with that the minimizes the log-spectral distance $\sum_{j=1}^{n^*} \left( \ln(2\pi f(\omega_j)) - z_j^0 \tilde{\theta} \right)^2$ or the sum of the squares of the Pearson’s residuals (not reported); notice that this value is very small, $K = 2$, i.e. we need to estimate two coefficients, and in fact $c_3 = 0.02$ is already negligible. The last two rows report the percent frequency by which the FEXP order $K$ is selected as optimal according to the BIC and AIC criteria. It seems that the AIC does a better job in selecting the correct order.

When the short run component is more substantial, as in the case $\phi = 0.9$ the FEXP order that is optimal for $d$ is much larger than that minimizing the distance between the true log-spectrum and the estimated one. Figure 9 reports the results of a simulation experiment referring to 10,000 replications of an ARFIMA(1,0.4,0) process with AR parameter $\phi = 0.9$ of length $n = 1000$. The first plot is the distribution of $\tilde{d}$ when FEXP models of increasing orders are fitted. Obviously, for $K$ small the estimates of $d$ are badly biased, but the variance is small. The bias variance trade-off is optimized for $K = 13$ (see the right upper plot, which displays MSE($\tilde{d}_K$) as a function of the order $K$). However, if our intent is to minimize the estimation error with respect to the true spectral density, the optimal order reduces to $K = 6$ (see the left bottom plot). Thus, even though $c_3^k$ and $c_k$ are individually estimated with larger MSE, their sum is estimated very accurately. Figure 10 illustrates this point for a simulated series: here the selected specification is $K = 6$ and although the long memory component is overestimated and the short run under-represented, the estimated log-spectrum is very close to the true one (right plot). Again we find that between the two information criteria AIC is doing a better job at picking the correct order, while BIC leads to excessively parsimonious specifications (see fig. 9, last plot).

Another instance of the inherent difficulty in separating LM from SM is in the analysis of realized volatility series: we consider for illustration purposes the series IXIC2.rv, referring to Nasdaq 100. When a FEXP($K$) is fitted, $\tilde{d}_K$ is around 0.5 for $K < 25$, then is starts declining until it reaches 0.2. BIC selects $K = 11$, but AIC is a minimum at $K = 35$ and both criteria are very flat with respect to $K$. If the model is selected according to the former $d$ is highly significant and the long memory component of the spectrum is substantial. On the contrary, when the order is selected according to the AIC, $d$ is no longer significant and

<table>
<thead>
<tr>
<th>$K$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias($\tilde{d}$)</td>
<td>0.320</td>
<td>0.089</td>
<td>0.028</td>
<td>0.009</td>
<td>0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td>St. err. ($\tilde{d}$)</td>
<td>0.028</td>
<td>0.043</td>
<td>0.056</td>
<td>0.068</td>
<td>0.079</td>
<td>0.089</td>
</tr>
<tr>
<td>MSE($\tilde{d}$)</td>
<td>10.326</td>
<td>0.976</td>
<td>0.393</td>
<td>0.475</td>
<td>0.626</td>
<td>0.796</td>
</tr>
<tr>
<td>Log-Spectrum distance</td>
<td>44.930</td>
<td>5.979</td>
<td>4.334</td>
<td>5.136</td>
<td>6.132</td>
<td>7.170</td>
</tr>
<tr>
<td>BIC</td>
<td>0.05</td>
<td>81.65</td>
<td>17.29</td>
<td>0.91</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>AIC</td>
<td>0.00</td>
<td>34.52</td>
<td>37.73</td>
<td>12.38</td>
<td>5.41</td>
<td>3.25</td>
</tr>
</tbody>
</table>
Figure 9: Estimation of $d$ and $\sigma^2$ for an ARFIMA($1$, $d$, $0$) process with $d = 0.4$ and $\phi = 0.4$, with $n = 1000$.

Figure 10: Simulated ARFIMA($1$, $d$, $0$) process with $d = 0.4$ and $\phi = 0.9$, with $n = 1000$. 

Figure 11: IXIC2 (Nasdaq 100) Realized volatility series (source: Oxford-Man Institute of Quantitative Finance - Realized Library).

the short memory component provides the most relevant contribution. In terms of the fitted log-spectrum the differences are negligible, though. See figure 11. A noticeable feature is that if we plot the short memory component versus the period in days, we are able to detect a cycle of 5 days (a financial week). If we fit an EXP($K$) model, devoid of the long memory component, the two criteria yield similar orders (32 and 35, respectively) and give a similar fit to the FEXP(35).

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 $c_k \approx d/k$.

As a matter of fact, denoting the log-spectrum of a FEXP($K$) model as $\ln[2\pi f(\omega; c, d)]$, where $c$ is the $K + 1$ vector of cepstral coefficients $c_0, \ldots, c_K$, following Magnus and Vasnev (2007), the change in the log-spectrum when $d$ is estimated, rather than constrained to zero is

$$\ln[2\pi f(\omega; \hat{c}, \hat{d})] - \ln[2\pi f(\omega; \hat{c}, 0)] = S(\omega) \hat{d},$$

where $\ln[2\pi f(\omega; \hat{c}, 0)]$ is the estimated log-spectrum when $d$ is taken to be zero (i.e. the EXP($K$) model is estimated) and $S(\omega)$ is the sensitivity of the log-spectrum to the $d$ parameter,

$$S(\omega) = -2 \left[ \ln|2\sin(\omega/2)| + \sum_{k=1}^{K} \frac{\cos(\omega k)}{k} \right].$$
Interestingly, $S(\omega)$, which for fixed $K$ is high at the low frequencies, vanishes as $K \to \infty$.

Similarly, the LM test for $H_0 : d = 0$ is based on the coefficient of determination of the Pearson’s residuals $\frac{2\pi}{\exp(z_j \theta)} - 1$ on $-S(\omega_j), j = 1, \ldots, n^*$, and again tends to zero as $K$ increases.

10 Bayesian Model Selection for the EXP model

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

$$\ln [2\pi I(\omega_j)] - \psi(1) = c(\omega_j) + \epsilon_j,$$

where $c(\omega) = \ln[2\pi f(\omega)]$ is a natural cubic spline, assuming a normal distribution for the error term $\epsilon_j \sim NID(0, \psi(1))$. The log-spectrum is estimated as

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

(11)

where $\rho$ 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 (1996) modelled $\ln[2\pi f(\omega)]$ by a periodic cubic spline, but approximated the distribution of $\epsilon_j$ by a Gaussian mixture with 5 components. Deterministic cycles (spikes) are detected when a mixture component with high variance is active.

Pawitan and O’Sullivan (PS, 1994) proposed a nonparametric estimator of the log-spectrum $c(\omega)$ based on the maximisation of the penalised likelihood (PL)

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

(12)

where $\rho$ is a penalty parameter and $c^{(r)}(\omega)$ is the $r$-th derivative of $c(\omega)$. For the EXP model $c(\omega) = c_0 + 2 \sum_k \cos(\omega k)$ the the PL 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 $c_k \sim NID(0, \rho k^{-2r}), k > 0$ on the cepstral coefficients.

Rather than imposing a smoothness prior on the model for $c(\omega)$, or on the coefficients $c_k$, we may consider the full basis expansion for the log-spectrum, which entails $n^* + 1$ cepstral coefficients, and select those that are relevant. Hence, the alternative approach that we propose is to perform Bayesian estimation
and model selection of the log-periodogram regression model with centred log-gamma errors,

\[
\ln[2\pi \hat{I}(\omega_j)] - \psi(m) = c_0 + \sum_{k=1}^{M} \delta_k c_k \cos(\omega_j k) - \delta_d 2d \ln|2 \sin(\omega_j/2)| + \epsilon_j, j = 1, \ldots, M, \tag{13}
\]

which includes all the cosine terms of the Fourier expansion. Here we consider periodogram pooling with fixed \(m, M = \left\lceil n^* / m \right\rceil\) and the frequencies \(\omega_{jm+k}\) have been relabeled \(\omega_j\) for notation simplicity. The coefficients \(\delta_k, k = 1, \ldots, M\), and \(\delta_d\) are indicator variables that take two values, 1 if the component is in the model and 0 otherwise.

The distribution of the disturbance term, \(p(\epsilon_j)\), which is that of the log transformation of a Gamma\((m, 1)\) random variable, is approximated by a Gaussian mixture with 10 components, as advocated by Früwirth-Schnatter and Früwirth-Schnatter (2007):

\[
p(\epsilon_j) = \sum_{r=1}^{10} \pi_r g(\epsilon_j, \mu_r, \sigma_r^2)
\]

where \(g(\cdot, \mu, \sigma^2)\) is the Gaussian pdf with mean \(\mu\) and variance \(\sigma^2\). The weights \(\pi_r\) and the location scale parameters are chosen so as to minimise the Kullback-Leibler divergence with the log-Gamma density (see Früwirth-Schnatter and Früwirth-Schnatter, 2006). In the sequel we let \(\Upsilon = (\upsilon_1, \ldots, \upsilon_{10})\) denote the vector collecting the mixture indicators \(\upsilon_r\).

Let also \(D = (\delta_1, \ldots, \delta_M, \delta_d)\) denote the collection of the variable selection indicators, and \(\theta_D\) be the vector of model parameters \(c_0, c_k, k = 1, \ldots, c_M, d\). Our prior assumes an independent structure between each block of variables, i.e.

\[
p(D, \Upsilon, \theta_D) = \prod_k p(\delta_k) \cdot p(\delta_d) \cdot \prod_r \pi_r \cdot p(\theta_D)
\]

Equation (13) nests \(2^{M+1}\) models. The central issue is establishing a prior distribution for the indicator variables. A smoothness prior can be enforced by letting the prior probabilities for \(\delta_k = 1\) decline with \(k\). A possibility is to use the hierarchical prior

\[
\delta_k \sim \text{Bernoulli}(\pi_k), \quad \pi_k \sim \text{Beta}\left(1, \frac{a}{k + b}\right), \quad k = 1, \ldots, M.
\]

For the indicator for the memory parameter we assume \(\delta_d \sim \text{Bernoulli}(0.5)\). For short memory time series we set \(P(\delta_d = 0) = 1\).

As for the vector \(\theta_D\), we assume a Zellner’s prior, \(\theta_D \sim \text{N}(0, g(Z_D'Z_D)^{-1})\), where \(Z_D\) is the matrix of explanatory variables for the model indexed by the vector \(D\).

The above log-periodogram generalised regression model lends itself very nicely to variable selection.

- MCMC estimation is easy. A Gibbs (GS) sampler can be designed exploiting auxiliary mixture sampling based on the Gaussian mixture approximation of the log-Gamma distribution.
• Convergence is rewardingly fast for the EXP model due to the quasi-orthogonality of the explanatory variables. Preliminary evidence is that it is extremely good for the FEXP as well.

• Choice of the prior is crucial. A smoothness prior has to be enforced so that high order cepstral coefficients are less likely to enter the selection.

• Easy to generalise to the log-averaged periodogram model, which turns out to be very useful when \( n^* \) is very large and the number of potential regressors makes model selection unfeasible.

The GS scheme entails the following steps:

1. Initialise with \( \mathcal{D}^{(0)}, \Upsilon^{(0)} \),

2. Draw \( \theta^{(i+1)} \) from the full posterior \( \theta|\mathcal{D}^{(i)}, \Upsilon^{(i)}, y \sim \mathcal{N}(VZ_D\Omega^{-1}y,V) \) where \( V = [Z_D'(g^{-1}I + \Omega^{-1})Z_D]^{-1} \), \( \Omega \) is a diagonal matrix with the variances \( \sigma^2_r \) for the mixture components to which the \( j \)-th observation is assigned, and \( y \) is the vector with elements \( \ln|2\pi I(\omega_j)| - \psi(1) \).

3. For each \( j = 1, \ldots, n^* \) draw the mixture indicators \( \Upsilon^{(i+1)} \) from \( p(\Upsilon|y,\mathcal{D}^{(i)}, \theta^{(i)}) \) which is multinomial with probabilities proportional to

\[
\frac{\pi_r}{\sigma_r} \exp\left(-\frac{1}{2\sigma_r^2}(y_j - z_j^r\theta^{(i)}) - \mu_r\right), r = 1, \ldots, 10.
\]

4. Draw the selection indicators: given \( (\delta^{(i)}_1, \ldots, \delta^{(i)}_M, \delta^{(i)}_d) \)

   (a) generate \( \delta^{(i+1)}_1 \) from \( p(\delta_1|\delta^{(i)}_2, \ldots, \delta^{(i)}_M, \delta^{(i)}_d, \Upsilon^{(i+1)}, \theta^{(i+1)}) \propto p(\delta_1)p(y|\delta_1, \delta^{(i)}_2, \ldots, \delta^{(i)}_M, \Upsilon^{(i+1)}, \theta^{(i+1)}) \), where the second factor (marginal likelihood) is

\[
\exp\left(-0.5(\ln|\Omega| + \ln|g(Z_D'Z_D)^{-1}| - \ln|V|) + y'\Omega^{-1}[\Omega - Z_DVZ_D']\Omega^{-1}y\right)
\]

(here \( Z_D \) is the matrix of explanatory variables corresponding to the selection indicators upon which we are conditioning.

(b) Generate \( \delta^{(i+1)}_2 \) from \( p(\delta_2|\delta^{(i)}_1, \delta^{(i)}_3, \ldots, \delta^{(i)}_M, \delta^{(i)}_d, \Upsilon^{(i+1)}, \theta^{(i+1)}) \propto p(\delta_2)p(y|\delta_2, \delta^{(i+1)}_1, \ldots, \delta^{(i)}_M). \)

(c) Generate \( \delta^{(i+1)}_3, \ldots, \delta^{(i+1)}_M \) in a similar fashion.

(d) Generate \( \delta^{(i+1)}_d \) from \( p(\delta_d|\delta^{(i)}_1, \delta^{(i)}_2, \ldots, \delta^{(i)}_M, \Upsilon^{(i+1)}, \theta^{(i+1)}) \propto p(\delta_d)p(y|\delta^{(i+1)}_1, \ldots, \delta^{(i+1)}_M). \)

We applied the previous sampling scheme to the BJ series A, with 15,000 MCMC replications, after a burn in of 5,000 iterations, and \( m = 1 \). For the prior hyperparameters we set \( a = 1, b = 0 \). The main result is that the posterior probability for the presence of the long memory component is close to 1

28
Figure 12: BJ series A. Posterior distributions of $c_0$ and $d$ estimated using 30,000 MCMC random draws and sample correlograms of the draws.

(p(δ = 1|y) = 0.99). The regressors $\cos(\omega_jk)$ are all selected with probability less that 0.5, the maximum value being obtained for $k = 1$ (p(δ = 1|y) = 0.44).

Hence, Bayesian variable selection suggests the simple specification:

$$\ln[2\pi\bar{I}(\omega_j)] - \psi(1) = c_0 - 2d\ln|2\sin(\omega_j/2)| + \epsilon_j, j = 1, \ldots, n^*,$$

Figure 12 displays the posterior distributions of the model parameters estimated from 30,000 random draws. The posterior mean for $d$ is 0.417 and the posterior standard deviation is 0.0618. Notice that the logpev distribution is downward biased. This is so since the posterior mean of $\epsilon_j$ is not zero (only the prior is). This is corrected by subtracting the estimate of $E(\epsilon|y)$ obtained from the draws (this is what is displayed in the figure).

11 Conclusions

References


