Alma Mater Studiorum · Università di Bologna

Dottorato di Ricerca in Metodologia Statistica per la Ricerca Scientifica
XXVI ciclo

Settore Concorsuale: 13/D1-STATISTICA
Settore Scientifico Disciplinare: SECS-S/01 - STATISTICA

LOCAL TRIGONOMETRIC METHODS FOR TIME SERIES SMOOTHING

Relatore:
Chiar.mo Prof.
ALESSANDRA LUATI

Dott.ssa:
MARIA GENTILE

Coordinatore :
Chiar.mo Prof.
ANGELA MONTANARI

Marzo 2014

Acknoledgement

Developing a research is a stimulating and enriching task, but the path toward scientific discoveries is sprinkled onto uncertainty and hurdles. The enthusiasm I poured into statistical studies would not have found its way out without the support and encouragement that I received from the people I met during my doctoral program. This is why I feel in debt with my tutor, prof. A. Luati, for the faith she put on me and the ability to help me to trace this route. Likewise, I have to thank my family for the patience and the love they had in every moment.

Contents

List of Figures

List of Tables

Introduction

Time series analysis provides a framework for modeling a wide variety of phenomena, among which those in electrical engineering, astronomy, oceanography, ecology, demography, macroeconomy, insurances. Some models are settled in continuous time, though statisticians have paid specific attention to discrete models. A main problem faced in presence of sampled a time series is the detection of the "signal" in conjunction with the removal of the "noise". This terminology, borrowed from electrical engineering with reference to electrical devices, furnishes an insight also in modeling time series in different contexts.

Series derived from natural phenomena are often thought as "stationary", meaning that the generating process should maintain the same properties during the time passing. Stationary time series thus gave rise to a very popular and successful family of models (see Box and Jerkins, 1970)

The hypothesis of stationarity immediately fails in presence of changing phenomena. In particular the decomposition of time series in economics modeling arises from the modern Macroeconomic theories, which use to explain Economic phenomena as made up of two components: one of long-run equilibrium (trend), studied in the growth theories, and one of short-run fluctuations (cycle), typically in the range of 1.5 and 8 years. Under this assumption a discrete univariate time series admits the decomposition:

$$y_t = \tau_t + \kappa_t + \varepsilon_t. \tag{1}$$

We shall refer to τ_t as the trend, to κ_t as the cycle, and we let ε_t be a random error. Further, τ_t, κ_t and ε_t are modeled separately. Equation (1) furnishes a simple version of an Unobserved Component (UC) model, and it constitutes one of the most discussed model used for trend-cycle decomposition.

Even in presence of a long lasting trend it is interesting to capture the cyclical component of a series. Deviation from the trend can heavily affect the evolution of a phenomena and thus have an impact on decisions to be undertaken. This is evident for macroeconomic series data.For economic series, theoretical constructs and empirical analysis have put the attention on the "Business Cycle". In their seminal paper Burns and Mitchell (1946) defined the business cycle as follows:

Business cycles are a type of fluctuation found in the aggregate economic activity of nations that organize their work mainly in business enterprises: a cycle consists of expansion occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle: this sequence of changes is recurrent but not periodic; in duration business cycles vary from more than one year to ten or twelve years; they are not divisible into shorter cycles of similar character with amplitude approximating their own.

A specific quasi-periodical component of an economical time series is addressed as "seasonality", and is easily ascribed to alternating seasons and to the effect of it on human activities.

The distinction among trend, cycle and seasonality can be more blurred, thus the estimation of these components if often conducted jointly.

The aim of this thesis is modeling the cyclical component of time series by means of a local trigonometric model.

Trigonometric functions appear to be a very natural technique to model cycles, and their use can be ascribed first to Ancient Greeks, who described the motion of planets by means of eccentrics, deferentes cycles, epicycles and gave a first application of what would have be called later "Fourier analysis" (see Gallavotti, 2001). Schuster (1897, 1906) and Fisher (1929) later tried to identify hidden periodicities in astronomical time series fitting series by means of trigonometric trend.

Fitting by means of trigonometric functions gives rise to specific problems. First, approximating a non periodic function in a finite interval by means of trigonometric functions generates the so called "Gibbs phenomenon", that is the presence of wiggles at the extremes of the interval, which does not fade by increasing the order of the approximation. Secondly, the rate of convergence trigonometric function is $o(n^{-1/2})$ globally and $o(n^{-2/3})$ locally, and it does not reach the theoretical maximum order of convergence $o(n^{-4/5})$ available in nonparametric regression.

Open problems remain: estimating the order of the model and balancing between localization in the time domain and in the frequency domain.

A variety of alternatives is available in order to model cycles in a time series, among which stochastic harmonic models and pseudocycles generated in a $ARMA(p, q)$ (see Harvey, 1993).

The thesis is structured as follows.

Chapter 1 reviews some basic concepts related to time series and to trigonometric models, including the different possible definitions of trigonometric process, the notion of filter of a series and the main cases of kernel smoothers and LTI filter, and the setting of Fourier analysis, which allows to represent a series both in the "time domain" and in the "frequency domain" .

Chapter 2 examines some time-limited filter commonly employed to extract cycles from a macroeconomic time series. Among the chosen models there are the family of Wiener-Kolmogorov filters, ad hoc filter such as the Hodrick-Prescott filter, the Baxter and King filter, the Christiano and Fitzgerald filter, the family of Butterworth filters, providing a comparison among their theoretical properties. In this chapter it is proposed a generalization of the filter proposed by Christiano and Fitzgerald (2003) suitable to detect cycles associated to specific frequencies in presence of high order integrated processes, and they are performed simulation on IMA(2,1), IMA(2,2) processes.

Chapter 3 is focused on local methods of smoothing. Local models are gaining a major importance in time series analysis because of their ability to better exploit information relative to a fixed instant, possess lower variance and allow a faster detection of turning points of the phenomenon under exam.It is presented the terminology associated to local processes, and relevant examples are furnished. Then, it is proposed a local trigonometric model and worked out its statistical properties, with application to the smoothing of $ARIMA$ processes.

In *chapter 4* it is discussed the choice of the minimizing function arising in $L^2[0, 21\pi]$ and its related rates of convergence. Namely, Mean Squared Error (MSE), Integrated Mean Squared Error (MISE) and the Point-wise Mean Squared Error (PMSE) are examined. Further there are compared some information criteria such as AIC, BIC for selecting the order of the trigonometric model.

In *chapter 5* will give some insight in some open questions, such as the problem of the balance between time localization and frequency localization of the process, the use of alternative local methods such as splines or wavelets, the possibility of minimizing a different error criteria arising from L^p norms.

Chapter 1

Preliminary concepts

In this chapter they are discussed the first concepts needed to model deterministic and stochastic cycles of a time series. We recall that often the distinction between trend and cycle is not sharp, and it make sense to speak of a "trend-cycle" component.

The notation $y(t)$ will be used for a continuous process, that is $t \in I \subseteq \mathbb{R}$, while the notations y_i, y_{t_i} or y_t are used for discrete processes, that is $i = 1, \ldots, n$ or $t = 1, \ldots, n$.

1.1 Trigonometric Processes

We can distinguish two different way of modeling the cyclical component: as a trigonometric trend or as a harmonic processes.

Definition 1. A model with *trigonometric trend* is the following:

$$y_i = \mu + \sum_{k=1}^{\lambda} \left(c_k \cos(\omega_k t_i) + s_k \sin(\omega_k t_i)\right) + \varepsilon_i, \quad i = i, \ldots, n. \tag{1.1}$$

In this model the trend is a deterministic periodic function in sine and cosine terms, and the random component is the noise ε_i. The ε_i are usually assumed to be independent and identically distributed (i.i.d.) with zero mean and finite variance, or satisfying an Autoregressive Moving Average process of orders p and q ($ARMA(p, q)$) , or being a weakly mixing process. The frequencies ω_k are usually the Fourier frequencies, and they depend on the number of observation n: $\omega_k = \frac{2\pi k}{n}$. Different choices are possible for the ω_k.

The same model can be written only in cosine terms, or only in cosine terms:

$$y_i = \mu + \sum_{k=1}^{\lambda} \varrho_k \cos(\omega_k t_i - \phi_k) + \varepsilon_i. \tag{1.2}$$

for $\varrho_k = \sqrt{(c_k^2 + s_k^2)}$, $\cos\phi_k = c_k/\varrho_k$, or

$$y_i = \mu + \sum_{k=1}^{\lambda} \varrho_k \sin(\omega_k t_i + \phi_k) + \varepsilon_i$$

for $\varrho_k = \sqrt{(c_k^2 + s_k^2)}$, $\sin\phi_k = c_k/\varrho_k$.

The same model can be written in complex notation as

$$y_i = \mu + \sum_{k=-\lambda}^{\lambda} A_k e^{-i\omega_k t_i} + \varepsilon_i.$$

for $A_0 = \mu$, $A_k = \frac{c_k - is_k}{2}$, $A_{-k} = \frac{c_k + is_k}{2}$ $(k > 0)$, complex numbers.

Definition 2. A *harmonic process* consists in a sum of sine and cosine terms amplified by a random coefficient:

$$y_i = \mu + \sum_{k=1}^{\lambda} \alpha_k \cos(\omega_k t_i + \phi_k) \tag{1.3}$$

where the α_k are random processes, with zero mean and finite variance.

An alternative way of defining a harmonic process is taking the α_k fixed and assuming that the phases ϕ_k are independent random variables uniformly distributed over $[-\pi, \pi]$. If the α_k are fixed and the ϕ_k are independent rvs $\phi_k \sim U(-\pi, \pi)$, the harmonic model can be rewritten as

$$y_i = \mu + \sum_{k=1}^{\lambda} [\beta_k \cos(\omega_k t_i) + \gamma_k \sin(\omega_k t_i)], \tag{1.4}$$

for β_k, γ_k independent zero mean random variables. Under these assumptions the process y_i is stationary with mean equal to μ and covariance function given by:

$$cov(y_i, y_j) = \sum_{l=1}^{\lambda} \alpha_l^2 \cos(\omega_l (t_j - t_i))/2,$$

see Percival and Walden (1998).

The definition of harmonic process having random coefficients however is more general than the harmonic process with the α_k fixed. The complex form of eq.(1.3) is

$$y_i = \sum_{k=-\lambda}^{\lambda} A_k e^{i\omega_k t_i},$$

for $A_k = \alpha_k e^{i\phi_k}/2$ $(k \geq 0)$, $A_{-k} = \alpha_k e^{-i\phi_k}/2$ $(k \leq 0)$, $A_0 = 0$, $\omega_{-k} = -\omega_k$.

If both the α_k and the ϕ_k are fixed, the model becomes the deterministic part of the trigonometric trend model.

Let $\{X_t | t \in \mathbb{Z}\}$ a discrete zero mean stationary stochastic process,and let $\mathcal{P}_{t-1}X_t$ be the projection of X_t on the space generated by $\{X_s, s < t\}$ (see 1.4). X_t The Wold theorem states that if the one step prediction error $\sigma^2 = E|X_t - \mathcal{P}_{t-1}X_t|^2 > 0$, X_t is the sum of an $MA(\infty)$ process U_t and of a deterministic process V_t, with U_t and V_t uncorrelated.

$$X_t = U_t + V_t = \sum_{j=0}^{+\infty} \psi_j Z_{t-j} + V_t$$

Harmonic processes appear as the deterministic component of a zero-mean stationary process in the Wold decomposition.

A stationary process can be approximated by an harmonic process consisting of a finite number of terms. (Doob, 1953)

1.2 Filters

From an abstract point of view, the cyclical component of a deterministic or stochastic process can be obtained from the process itself by applying a filter to it.

Definition 3. A filter is an operator L that associates a process $y(\cdot)$ to the process $x(\cdot)$.

The filter L is said to be *linear* if $L(\alpha x(\cdot) + \beta y(\cdot)) = \alpha L(x(\cdot)) + \beta L(y(\cdot))$, for every choice of the series x_t and y_t and of the constants α and β.

The filter operates instantaneously if $y_t = L(x_t)$.

L is said to be *time-invariant* if the identity $y_t = L(x_t)$ implies $y_{t+h} = L(x_{t+h})$, for every number h.

The filter L is said to be *bounded* if y_t is bounded whenever x_t is bounded.

The filter L is said to be *realizable* or *causal* if y_t depends only on the past value x_s, $s \leq t$ of the input series $\{x_t\}$.

Definition 4. The *transfer function* $B(\lambda)$ is the response of the filter to the input series $x_t = e^{i\lambda t}$:

$$B(\lambda) = L(e^{i\lambda t}).$$

In a physical context, the index λ is called *pulsation*; if $\lambda = 2\pi f$, f is called *frequency* and it is the inverse is the period T. The process $x_t = e^{2\pi i f t}$ is an eigenfunction for a linear time invariant (LTI) filter, since $L(e^{2\pi i f t}) = G(f) e^{2\pi i f t}$. The eigenvalue $G(f)$ is called *frequency response function.* In general $G(f)$ is a complex function, thus it can be written as $G(f) = |G(f)| e^{i\theta(f)}$. Its modulus $|G(f)|$ is called *gain* of the filter, while its phase $\theta(f)$ is called *phase* of the filter.

A discrete linear filter $y = L(x)$ acquires the form $y_t = \sum_{i=-\infty}^{+\infty} \psi_{t,i} x_i$. Such a filter is time invariant if the weights $\psi_{t,i}$ depend only on $t - i$: $y_t = \sum_{i=-\infty}^{+\infty} \psi_{t-i} x_i$.

A general class of linear transformation of $x(\cdot)$ is obtained by kernel smoothing:

$$y(t) = \int_{-\infty}^{\infty} K(t, t') x(t') dt',$$

for any choice of the function $K(\cdot, \cdot)$, which is called *kernel.* Usually kernel function need to be symmetric and positive function. Such type of filter is also time invariant if and only if $K(t, t') = g(t - t')$. Thus LTI filter obtained by kernel smoothing are convolution products:

$$y(t) = \int_{-\infty}^{\infty} g(t - t') x(t') dt'.$$

The discrete version of the convolution product is

$$y_t = \sum_{i=-\infty}^{\infty} g_{t-i} x_i.$$

If we allow $g(\cdot) = \delta(\cdot)$, for $\delta(\cdot)$ Dirac delta function, then the class of the convolution products coincides with the class of the LTI filters. The convolution product is symmetric:

$$y(t) = \int_{-\infty}^{\infty} g(t') x(t - t') dt'.$$

The function $g(\cdot)$ that appears in the formulas above is called *impulse response function* of the filter, and it is uniquely determined by the filter as the response to the unit impulse, since

$$L(\delta(t)) = \int_{-\infty}^{\infty} g(t') \delta(t - t') dt' = g(t).$$

Its Fourier transform is the transfer function of the filter: $\mathcal{F}(L(\delta(t); f)) = L(e^{2\pi i f t})$.

A LTI is bounded if and only if its impulse response function is summable.

A LTI filter is *symmetric* if it has a symmetric impulse response function; in this case the phase of the filter is one.

Fundamental examples of LTI filters are:

- a high-pass filter defined by having transfer function $G(f) = 0$ for $f < W$ and $G(f) = 1$ for $f > W$;

- a low-pass filter defined by having transfer function $G(f) = 1$ for $f < W$ and $G(f) = 0$ for $f > W$;

- a band-pass filter defined by having transfer function $G(f) = 1$ for $|f| < W$ and $G(f) = 0$ for $|f| > W$.

The number W is called *cut off frequency* .

The impulse response function of a band-pass filter is

$$g(u) = \frac{\sin(2\pi W u)}{\pi u} \quad u \neq 0$$
$$2W \quad u = 0.$$

A fundamental result in filter theory was worked out separately by Kolmogorov (1941) and Wiener (1949):

Theorem 1. *If the input series y_t admits the decomposition $y_t = \eta_t + \varepsilon_t$, where η_t is the signal and ε_t is the noise, then the minimum mean square estimator of the signal is $\hat{\eta}_t = E(\eta_t | x_t, x_{t-1}, \dots)$.*

1.3 Fourier Transform

Analysis of time series can be cast in time domain or in frequency domain. In the following paragraphs we will make precise the meaning of this expression, introducing the Fourier transform defined on different spaces of function.

If x_t, $t \in \mathbb{Z}$ is in l_1, that is $\sum_{t=-\infty}^{+\infty} |x_t| < +\infty$, then x_t is the Fourier series of a periodic function $y(\omega)$, $\omega \in [-\pi, \pi]$. Moreover, if $x_t \in l_2$, then $y(\omega) \in L^2([-\pi, \pi])$, and this correspondence is an isomorphism between l_2 and $L_2([-\pi, \pi])$:

$$y(\omega) = \sum_{t=-\infty}^{+\infty} x_t e^{-it\omega}, \quad \omega \in [-\pi, \pi] \Leftrightarrow x_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} y(\omega) e^{i\omega t} d\omega, \quad t \in \mathbb{Z}.$$

The equivalence of scalar products is written as:

$$\sum_{-\infty}^{+\infty} x_n \bar{x}'_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} y(\omega)\bar{y}'(\omega) d\omega.$$

If $x(t) \in L^1(\mathbb{R})$, that is $\int_{-\infty}^{+\infty} |x(t)| dt < +\infty$, then $x(t)$ admits the Fourier transform $y(\omega) = \int_{-\infty}^{+\infty} x(t) e^{-i\omega t} dt$. If $x(t) \in L^2(\mathbb{R})$, then $y(\omega) \in L^2(\mathbb{R})$, and the following correspondence in an isomorphism between l_2 and $L^2(\mathbb{R})$ (see Rudin, 1974):

$$y(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x(t) e^{-it\omega} dt, \quad \omega \in \mathbb{R} \Leftrightarrow x(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y(\omega) e^{i\omega t} d\omega, \quad t \in \mathbb{Z}.$$

The equivalence of scalar products in this case is written as:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x(t)\bar{x}'(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} y(\omega)\bar{y}'(\omega) d\omega$$

If x_n is a finite discrete sequence of length T, then it admits Fourier expansion y_k:

$$y_k = \frac{1}{T} \sum_{n=0}^{T-1} x_t e^{\frac{-i2\pi nk}{T}}, \quad k = 0, \ldots, T-1 \Leftrightarrow x_n = \sum_{n=0}^{T-1} y_k e^{\frac{i2\pi nk}{T}}, \quad n = 0, \ldots, T-1.$$

The above mentioned identities makes sense in the counting measure. The equivalence of scalar products in this last case is written as:

$$\sum_{n=0}^{T-1} x_n \bar{x}'_n = \sum_{n=0}^{T-1} y_n \bar{y}'_n.$$

These relations constitutes the basis for Fourier analysis of deterministic processes. Often they are referred as relations between "time domain" and "frequency domain", since in physical applications t plays the role of time and ω plays the role of frequency.

Among the many useful relations available for Fourier transforms we recall the following theorem:

Theorem 2. *If $h(t) = f(t)g(t)$, then the Fourier transform of $h(t)$ is given by the convolution product of the Fourier transform of $f(t)$ and the Fourier transform of $g(t)$: $(\mathcal{F}h)(\omega) = (\mathcal{F}f)(\omega) * (\mathcal{F}g)(\omega)$. Viceversa, if $H(t) = F(t) * G(t)$, then the Fourier transform of $H(t)$ is given by the product of the Fourier transform of $F(t)$ and the Fourier transform of $G(t)$ $(\mathcal{F}H)(\omega) = (\mathcal{F}F)(\omega) * (\mathcal{F}G)(\omega)$.*

The proof is found in the appendix. From this theorem follows:

Theorem 3. *If y_t is the process obtained from the process x_t by applying a linear filter, that is $y_t = \sum_{i=-\infty}^{+\infty} g_{t-i}x_t$, then the gain of the filter is $Y(e^{-i\omega}) = G(e^{-i\omega})X(e^{-i\omega})$, where $G(e^{-i\omega}) = \sum_{k=-\infty}^{-\infty} g_k e^{-i\omega t}$.*

Fourier analysis has been extended by Wiener (1949) in a probabilistic context, as we will see in the following paragraph.

1.4 Spectral Analysis of stationary time series

The space of complex random variable X on a measure space $(\Omega, \mathcal{F}, P)$ satisfying $E|X|^2 < +\infty$, endowed with the scalar product $< X, Y >= E(X\bar{Y})$ constitutes a Hilbert space.

If $\{X_t| \in \mathbb{Z}\}$ is a discrete stochastic process,the *lag operator* is defined by means of $L(X_t) = X_{t-1}$. The *projection* of X_t on the space generated by $\mathcal{M} = \{X_s, s < t\}$ is the process $Y \in \mathcal{M}$ having minimum distance from X, and it is denoted by means of $\mathcal{P}_{t-1}X_t$.

Definition 5. A stochastic process is *weakly stationary* if, for each $t \in \mathbb{Z}$ and for each $k \in \mathbb{Z}$, it appens $E(y_t) = \mu$ and $E((y_{t-k} - \mu)(y_t - \mu)) = \gamma_k$ independently of t, or, in an equivalent formulation, if the first two moments of the process does not depend on the time t.

The function γ_k, $k \in \mathbb{Z}$ is called **autocovariance function** (ACF) of the process X_t, and it is a symmetric positive definite function.

Every discrete zero-mean stationary process X_t admits a decomposition into a series of sinusoidal components with uncorrelated random effects, i.e.

$$X_t = \int_{(-\pi,\pi]} e^{it\nu} dZ(\nu)$$

where $Z(\lambda)$ is a suitable right continuous orthogonal increment stochastic process. (Brockwell and Davis, 1994). Correspondingly, the covariance function γ_k of X_t, being a summable sequence, admits the spectral representation (Herglotz theorem):

$$\gamma_k = \int_{(-\pi,\pi]} e^{ik\nu} dF(\nu),$$

where $F(\cdot)$ is a non-decreasing, right continuous bounded function on $[-\pi, \pi]$ with $F(-\pi) = 0$ and $F(\pi) = \gamma(0) = E|X_t|^2$. The function F is called *spectral distribution* of γ and if $F(\lambda) = \int_{-\pi}^{\lambda} f(\nu)d\nu$ then f is called *spectral density* of γ. The orthogonal increment stochastic process Z and the spectral distribution F are linked by the relation $F(\mu) - F(\lambda) = |Z(\mu) - Z(\lambda)|^2$, $-\pi \leq \lambda \leq \mu \leq \pi$, and if T is the isomorphism of $\bar{sp}\{X_t, t \in \mathbb{Z}\}$ onto $L(F)$ then $Z(\lambda) = T^{-1}(\chi_{(-\pi,\lambda]})$ where $\chi_{(-\pi,\lambda]}$ is the indicator function of $(-\pi, \lambda]$.

If the one step prediction error σ^2 is greater than zero then the spectral distribution F_X can be decomposed as $F_X = F_U + F_V$, where F_U and F_V are respectively the spectral distribution of the $MA(\infty)$ process U_t and the spectral distribution of the deterministic V_t in the Wold decomposition of X_t. F_U is absolutely continuous with respect to the Lebesgue measure, and admits spectral density f_U: $F_U(A) = \int_A f_U(u)du$; F_V has no absolutely continuous component, and $F_V(A) = \sum_{\lambda_j \in A} p(\lambda_j)$, where $p(\lambda)$ is the spectral mass concentrated in λ.

A harmonic process without an additive error term has a pure discrete spectrum (Percival and Walden, 1998).

In the following we will deal mainly with stationary processes, and when necessary we will detrend series of data.

1.5 Ergodicity

The process y_t is said to be *ergodic with respect to the second moments* if the autocovariance function calculated with respect to the time converges almost surely to autocovariance function calculated with respect to the ensamble:

$$C_W(\tau) \equiv \lim_{T\to\infty} \frac{1}{T} \int_{-T}^{T} y_{t+\tau} y_t dt = E[y_{t+\tau} y_t] \equiv C(\tau) \quad a.s.$$

A linear process x_t is always ergodic (Grenander and Rosenblatt, 1957). A normal process is ergodic if and only if its spectrum is continuous.

Chapter 2

Extraction of Business cycle in finite samples

In this chapter there are presented some of the most frequently used filters for the extraction of the cyclical component of an economic time series. A starting point is the Wiener-Kolmogorov filter, which arises from the minimization of a quadratic function. This kind of filter requires adjustments to deal with economic series. A first problem is to smooth short series, so it make sense to work out finite sample version of the filter, or, further local version of the filter. A second problem is the treatment of nonstationarity. A modification of the minimization function with the addition of a penalized term is the Hodrick-Prescott filter.

The problem of finding the filter can be settled both in the time domain and in the frequency domain. The Baxter and King filter and the Christiano and Fitzgerald filter rise from the analysis in the frequency domain. This second framework allows to select the frequencies to describe the business cyclical activity.

As an alternative, one can try to model directly the stochastic cycle by modulating a white noise, or a colored noise applying to it trigonometric coefficients (Harvey and Trimbur, 2003).

In the second part of the chapter there are examined the links between the ideal band-pass filter, the Baxter and King filter and the Christiano and Fitzgeralg filter, by means of the analysis in the frequency domain. It is worked out the explicit solution of the smoothing problem following the asset of CF in the case of a $MA(q)$ error.Then it is furnished a generalization of the CF filter appropriate to extract cycle from a high order integrated series without a preparatory differentiation. Simulations are made in the case of $I(2)$process.

2.1 Mainstream in detecting Business Cycle

In econometric time series, the trend component is supposed to be originated by structural causes such as institutional events, demographic and technological changes, new ways of organization, and it has a slow evolution. The business cycle received different definitions, and can be described as a quasi-periodic oscillation characterized by periods of expansions and contractions. Often the estimation of both trend and cycle are conducted simultaneously.

The literature distinguishes between real business cycle and business cycle tout court. In the following paragraph we will discuss some of the most popular filters used to extract the real business cycle, pointing out similarities and differences by means of a theoretical analysis.

2.1.1 Wiener-Kolmogorov filter

Many filters used in physical and econometrical applications are encompassed by the family of Wiener Kolmogorov (WF) filters.

The classical theory of linear filtering was developed independently by Wiener (1941) and Kolmogorov (1941), and it assumes that data generating processes are stationary and that adequately long data series are available. After some necessary adjustments (Bell, 1984), this theory is often applied to the treatment of economic data.

The Wiener-Kolmogorov filter extract the signal of a sequence under the assumption that observations are sum of signal and noise:

$$y_t = s_t + n_t.$$

The estimate of the signal is a linear combination of the data points available:

$$\hat{s}_t = \sum_{j=-p}^{q} \psi_{t,j} y_{t-j}.$$

Under the assumptions that the filtered series $\hat{s}_t$ is a LTI filter should minimize the mean square error:

$$E[(y_t - \hat{s}_t)^2] = min_{\psi_i}.$$

Classical theory assumes that error and noise are independent, or at least uncorrelated. So, denoted by $\gamma^{xx}(z)$ the autocovariance generating function (AGV) $\sum_{k=-\infty}^{+\infty} \gamma_k z^k$ of a process x_t, the AGV of the process y admits the decomposition

$$\gamma^{yy}(z) = \gamma^{ss}(z) + \gamma^{nn}(z).$$

An autocovariance is a positive definite function, so it admits the Cramér-Wold decomposition (see Brockwell and Davis, 1991):

$$\gamma^{yy}(z) = \phi(z^{-1})\phi(z), \quad \gamma^{ss}(z) = \theta(z^{-1})\theta(z), \quad \gamma^{nn}(z) = \theta_n(z^{-1})\theta_n(z).$$

The minimum mean square error criterion yields

$$0 = E[y_{t-j}(s_t - \hat{s}_t)] = E[y_{t-j}s_t] - \sum_{k=-p}^{q} \psi_{t,k} E[y_{t-j}y_{t-k}] = \gamma_j^{ys} - \sum_{k=-p}^{q} \psi_{t,k}\gamma_{j-k}^{yy},$$

where, for each $j \in \mathbb{Z}$, $\gamma_j^{ys} = E[y_t s_{t-j}]$. Multiplying the previous equation by z^j, defined $\gamma^{ys}(z) = \sum_{j=-\infty}^{+\infty} \gamma_j^{ys} z^j$, one obtains the finite sample version of the WF filter:

$$\gamma^{ys}(z)|_{(-p,q)} = [\gamma^{yy}(z)\psi(z)]|_{(-p,q)}.$$

The subscript $(-p, q)$ means that only the coefficient of z^j for $-p \leq j \leq q$ are involved, and that ψ_j=0 for $j \notin [-p, q]$. For a casual Infinite Impulse Response (IIR) filter ($p = 0$, $q = +\infty$) the WK filter becomes (Whittle, 1983)

$$\psi(z) = \frac{1}{\phi(z)} \left[\frac{\gamma^{ss}(z)}{\phi(z^{-1})} \right]_+ .$$

If y_t is available for $-\infty < j < +\infty$, the WF filter is simply given by

$$\gamma^{ss}(z) = \gamma^{yy}(z)\psi(z),$$

that is

$$\psi(z) = \frac{\gamma^{ss}(z)}{\gamma^{yy}(z)} = \frac{\gamma^{ss}(z)}{\gamma^{ss}(z) + \gamma^{nn}(z)}.$$

Let $By_t = Ly_t = y_{t-1}$ the backward operator , and $Fy_t = L^{-1}y_t = y_{t+1}$ the forward operator. If the observed series admits a signal plus noise decomposition:

$$x_t = s_t + n_t,$$

and both signal and noise are generated by ARMA processes (see Kaiser and Maravall, 2005):

$$\phi_s(B)s_t = \theta_s(B)a_{st},$$

$$\phi_n(B)n_t = \theta_n(B)a_{nt},$$

with a_{st} and a_{nt} mutually independent white-noise processes whit zero mean, and variance σ_s^2 and σ_n^2 respectively, polynomials $\phi_s(B)$ and $\phi_n(B)$ coprime, i.e. having no common factors, $\theta_s(B)$ and $\theta_n(B)$ share no unit root in common, then also x_t follows an ARMA process:

$$\phi(B)x_t = \theta(B)a_t,$$

with a_t white-noise process, $\theta(B)$ invertible and $\phi(B)$ given by $\phi(B) = \phi_s(B)\phi_n(B)$, and a_t satisfying the equation:

$$\theta(B)a_t = \phi_n(B)\theta_s(B)a_{st} + \phi_s(B)\theta_n(B)a_{nt}.$$

The Wiener-Kolmogorov filter designed to extract the signal s_t in this model is

$$\hat{s}_t = \frac{AGF(s_t)}{AGF(x_t)}x_t = \frac{\sigma_s^2 \frac{\theta_s(B)\theta_s(F)}{\phi_s(B)\phi_s(F)}}{\sigma_a^2 \frac{\theta(B)\theta(F)}{\phi(B)\phi(F)}}x_t = k_s \frac{\theta_s(B)\phi_n(B)}{\theta(B)} \frac{\theta_s(F)\phi_n(F)}{\theta(F)}x_t, \tag{2.1}$$

or

$$\hat{s}_t = k_s \frac{\theta_s(B)}{\phi_s(B)} \frac{\theta_s(F)\phi_n(F)}{\theta(F)}a_t,$$

with $k_s = \sigma_s^2/\sigma_a^2$. The same filter can be construed as the $AGF$of the process z_t satisfying

$$\theta(B)z_t = \theta_s(B)\phi_n(B)b_t,$$

where b_t is white noise whit variance $k_s = \sigma_s^2/\sigma_a^2$.

If s_t and n_t are orthogonal, the spectral density of x_t admits the decomposition

$$g(\omega) = g_s(\omega) + g_n(\omega).$$

The gain of the WK filter is the Fourier transform of the ratio of the two gains:

$$G(\omega) = g_s(\omega)/g(\omega)$$

and the gain of the Minimum Mean Square Error (MMSE) estimator $\hat{s}_s$ is given by:

$$g_{\hat{s}}(\omega) = \left[\frac{g_s(\omega)}{g(\omega)}\right]^2 g(\omega) = \frac{g_s(\omega)}{g(\omega)} g_s(\omega) = G(\omega) g_s(\omega).$$

Since $G(\omega) \leq 1$, we have $g_{\hat{s}}(\omega) \leq g(\omega)$, hence the MMSE filter underestimates the variance of the theoretical component. The WK filter is well definite even if the ϕ-polynomials contain unit roots, and thus can be extended in a straightforward manner to the nonstationary case. In fact in the latter case, the ϕ_s-polynomial can be factorized as $\phi_s(B) = \varphi(B) D_s(B)$, where D_s contains all the unit roots and φ_s is stable. Thus, applying D_s to equation (2.1) and replacing $D_s x_t$ by $[\theta(B)/(\varphi_s(B))\phi_n(B)]a_t$, it is obtained

$$\hat{s}_t = k_s \frac{\theta_s(B)}{\varphi_s(B)} \frac{\theta_s(F)\phi_n(F)}{\theta(F)} a_t,$$

which provides the model that generated the stationary transform of the estimator $\hat{s}_t$.

Wiener-Kolmogorov filter and Kalman filter yield the same results, although Kalman filter is computationally advantageous and Wiener-Kolmogorov filter allows to show more easily theoretical properties of the filter.

2.1.2 The Hodrick - Prescott filter

The Hodrick Prescott (HP) filter, introduced by Hodrick and Prescott (1997), constitutes a standard method for removing trend movement in the business cycle literature. The HP filter for the trend component is a highpass filter obtained as the solution of

$$\min_{\tau_t} \sum_{t=1}^{T} [(x_t - \tau_t)^2 + \lambda((\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}))^2] \tag{2.2}$$

where the residuals $z_t = x_t - \tau_t$ represent the business cycle, while the parameter λ penalizes the second differences of the x_t, and must be chosen by the researcher. This criterium is the discrete version of (3.1): the summation takes the place of the integral and the second derivative of the trend is substituted by $\Delta^2 \tau_t = \tau_{t+1} - 2\tau_t + \tau_{t-1}$.

The infinite sample version of the Hodrick-Prescott filter defines the cyclic component of a time series y_t as

$$y_t^c = \frac{\lambda(1-L)^2(1-L^{-1})^2}{1+\lambda(1-L)^2(1-L^{-1})^2} y_t$$

where λ is a parameter that penalizes the variation in the growth component. It removes non-stationary components that are integrated of order four or less. It is symmetric, so there is no phase shift. It is a two-sided moving average filter of infinite order.

The Hodrick-Prescot filter is an ad-hoc filter frequently used by national central bank to detect and to predict the business cycle. Some important limitation of the HP filter are imprecise end-point estimation, large revision in recent estimators, spurious result, noise contamination of the cyclical signal, moreover the choice of λ is not supported by an established theory (a common choice is $\lambda = 1600$ for GDP quarterly data). A frequently used method for the extraction of the business cycle from macroeconomic time series is applying the Hodrick-Prescott filter to X11- seasonally adjusted time series.

An alternative to asymmetric filters, when the last k observations are missing, is to substitute the future observations with their optimal forecasts, obtained by means on an $ARIMA$ estimators.

Another way of obtaining the HP filter for the cyclical component is to search for a weighted averages of the original data:

$$y_t^c = \sum_{h=1}^{T} d_{ht} y_h$$

with $\sum_{h=1}^{T} d_{ht} = 0$ for each t. More precisely (King and Rebelo, 1993)

$$y_t^c = \frac{\theta_1\theta_2}{\lambda}\Big(\sum_{j=0}^{+\infty}(A_1\theta_1^j + A_2\theta_2^j)y_{t-j} + \sum_{j=0}^{+\infty}(A_1\theta_1^j + A_2\theta_2^j)y_{t+j}\Big),$$

with $\theta_1, \theta_2, A_1, A_2$ depending on λ.

The Hodrick-Prescott filter belongs to the family of the Butterworth filters, which are characterized by a Gain function of the form:

$$G(\omega) = \Big[1 + \Big(\frac{\sin(\omega/2)}{\sin(\omega_0/2)}^{2d}\Big)\Big]^{-1}, \quad 0 \leq \omega \leq \pi.$$

Thanks to the trigonometric identity

$$4\sin^2(\omega/2) = (1 - e^{-i\omega})(1 - e^{i\omega}),$$

substituted $e^{-i\omega}$ by the backward operator B and $e^{i\omega}$ by the forward operator F, we obtain the time-domain representation of the Butterworth filter:

$$\nu(B, F) = \frac{1}{1 + \lambda[(1-B)(1-F)]^d} = \frac{\frac{1}{\lambda[(1-B)(1-F)]^d}}{1 + \frac{1}{\lambda[(1-B)(1-F)]^d}},$$

with $\lambda = [4\sin(\omega_0/2)^2]^{-d}$.

This identity shows that a Butterworth filter is the Wiener-Kolmogorov filter used to estimate the signal s_t in the model $x_t = s_t + n_t$ when the signal s_t is a $IMA(d, 0)$ process:

$$\nabla^d s_t = a_{st}. \tag{2.3}$$

The equation (2.3) admits various generalizations, depending on the nature of the phenomenon under examination, and can be substituted by an ARIMA model; so accompanied by the unobserved component model it give raise to a reduce form equation. Applying the WK filter for the signal to the reduce form equation, it is obtained a more general HP filter, called HP-ARIMA filter. (Kaiser and Maravall, 1999).

The smoothing of an economic time series has a prominent role in detecting turning points. A possible definition of turning point is the first of at least two successive periods of negative/positive growth. The ability of detecting a turning point by means of an established filter can be tested by counting, in a set of series, both the mean number of turning points that are dated on the original series and missed by the filtered one, and the mean number of turning points detected on the filtered series but not present in the original one, maintaining separated "peaks" and "throughs".

2.1.3 The ideal band-pass filter

In the following we assume that process generating the data x_t has the decomposition $x_t = y_t + \tilde{x}_t$, where y_t has power only in the frequencies belonging to the interval $I = \{(a, b) \cup (-b, -a)\} \in (-\pi, \pi)$, and $\tilde{x}_t$ having power only in the complement of this interval in $(-\pi, \pi)$. This happens in particular if the spectral density f_X and f_Y are linked by $f_Y(\omega) = f_X(\omega)\chi_I(\omega)$, where χ_I is ideal band pass of I (that is the characteristic function of the interval I). This property allows to represent the process y_t by applying to x_t a LTI filter: $Y_t = \sum_{j=-\infty}^{+\infty} B_j X_{t-j}$, Since $f_Y(\omega) = |B(e^{-i\omega})|^2 f_X(\omega)$, where $B(e^{-i\omega}) = \sum_{j=-\infty}^{+\infty} B_j e^{-ij\omega}$, the weights B_j can be computed as Fourier coefficients of $B(e^{-i\omega}) = \chi_I(\omega)$. This yields:

$$\begin{aligned} B_j &= \frac{\sin(jb) - \sin(ja)}{\pi j}, \quad j \geq 1, \\ B_0 &= \frac{b-a}{2}, \quad a = \frac{2\pi}{T_u}, \quad b = \frac{2\pi}{T_l}. \end{aligned} \tag{2.4}$$

The symmetric linear filter isolating a period of oscillation between T_l and T_u ($2 \leq$

$T_l < T_u < \infty$) minimizing the mean squared error criterion $E[(y_t - \hat{y}_t)^2|x]$, where $x = [x_1, \ldots, x_T]$ is the observed sample, is

$$\hat{y}_t = \sum_{k=-\infty}^{+\infty} B_{t-k} x_k$$

This is the ideal band-pass filter. Since it requires infinite, past and future observations, and infinite weights, it is unfeasible. Obviously, it is not causal.

2.1.4 The Baxter and King filter

Baxter and King (1999) developed an approximate band-pass filter which isolates business-cycle fluctuations in macroeconomic time series. The filter was designed to isolate fluctuations that persist for periods of two through eight years, and it also renders stationary a series that is integrated of order one or two or that contains a deterministic trend.

The Baxter-King (BK) filter exhibits several desirable properties: it is a moving average consisting of infinite terms, that extracts a specified interval of frequencies, it does not introduce phase shift, it is optimal with respect to a specific loss function, it renders stationary time series integrated of order one or two or presenting a quadratic trend. As a consequence, the BK filter can be applied to the rough data without pre-filtering. The underlying model implies that very slow moving can be interpreted as a trend, and very high frequency components represent an irregularity in the phenomenon. Thus the problem is to specifies which frequencies can be considered involved in the business cycle.

The filtered series is

$$y_t^* = \sum a_k y_{t-k}.$$

The BK filter has the property $\sum_{k=-K}^{K} a_k = 0$, and it is symmetric. Simple algebra shows that these two properties imply that 1 and -1 are roots of the lag polynomial:

$$a(L) = \sum_{k=-K}^{K} a_k L^k = (1-L)(1-L^{-1})\psi(L).$$

The ideal low-pass filter $\beta(\omega)$ which passes only frequencies $-\underline{\omega} \leq \omega \leq \underline{\omega}$ for a suitable cut off frequency $\underline{\omega}$ has the time-domain representation $b(L) = \sum_{h=-\infty}^{+\infty} b_h L^h$, where $b_0 = \underline{\omega}/\pi$, $b_h = \sin(h\underline{\omega})/h\pi$. Since it consists of infinite terms, it makes sense to approximate it with finite moving average filter $\alpha(L)$. If $\alpha_K(\omega)$ is the Fourier transform of discrete filter minimizing a quadratic loss function, in which every frequency has the same weight:

$$Q = \int_{-\pi}^{\pi} |\beta(\omega) - \alpha_K(\omega)|^2 d\omega, \tag{2.5}$$

then $\alpha_K(\omega)$ is obtained by simply truncating the ideal filter's weights a_k at lag K.

If b_h are the weights of the ideal low-pass filter, then the weights of the ideal high-pass filter are $1 - b_0$ at $h = 0$, and $-b_h$ at $h \neq 0$, and the optimal $K-$approximation of the HP_∞ filter is obtained by truncating the ideal HP_∞ filter.

If $\bar{\beta}(\omega)$ and $\underline{\beta}(\omega)$ are the ideal symmetric low-pass filter relative to $-\bar{\omega} \leq |\omega| \leq \bar{\omega}$ and to $-\underline{\omega} \leq |\omega| \leq \underline{\omega}$ respectively, the ideal band-pass relative to the interval of frequencies $\underline{\omega} \leq |\omega| \leq \bar{\omega}$ is obviously given by $\bar{\beta}(\omega) - \underline{\beta}(\omega)$, and it has weights $\bar{b}_h - \underline{b}_h$.

Baxter and King (1999) deduce the optimal approximating low-pass filter minimizing the quadratic form (2.5) under the constraints $a_j = 0$ for $|j| > K$, $\sum_{h=-K}^{K} a_h = 1$ and $a_h = a_{-h}$ (so that $1 - \sum_{h=-K}^{K} a_h = 0$), which is determined by

$$a_h = b_h + \theta, \qquad \theta = (1 - \sum_{h=-K}^{K} b_h)/(2K+1). \tag{2.6}$$

Working in the frequency domain, researchers usually calculate the discrete Fourier transform (DFT), computing the periodic components associated to a finite number of harmonic frequencies, then they abruptly drop out the frequencies that lie outside the band of interest, finally they calculate the inverse Fourier transform to get the time-domain filtered series. The main risks of using this method are, firstly, the need of detrending the series of the observation before applying the DFT, in order to remove unit roots, and, secondly, the dependence of the weights and of the filtered series on the sample length T, since the procedure is not recursive.

Baxter and King (1999) provide a detailed comparison between the approximated band-pass filter and the Hodrick-Prescott filter, recognizing that the second one is a good approximation of the BK filter for quarterly Gross National Product (GNP) data.

2.1.5 First differencing

If we extract the cyclical component of a time series by a first difference filter: $y_t^c = (1 - L)y_t$, we obtain a filter that is not symmetric and introduces a time shift between variables, more over, the filter are reweights strongly toward higher frequencies. This is easily seen from $|G(f)| = \sqrt{2 - 2\cos(2\pi f)}$, $\arctan(\theta(f)) = \frac{\sin(2\pi f)}{1-\cos(2\pi f)}$.

In general, the application of a filter to a series of observation could produce some distortion and generate spurious cycles (Yule-Slutsky effect) . For difference filter and summation filter the reason is the following. If the first difference operator $1 - L$ is applied d times to the series y_t, and then the summation filter $1 + L$ is applied s times to the resulting series, the effect of the difference filter is to attenuate the low frequencies, while the effect of the summation filter is to attenuate the high frequencies. Thus the overall effect is the transfer function shows a peak, that could be misinterpreted as the presence of a cycle.

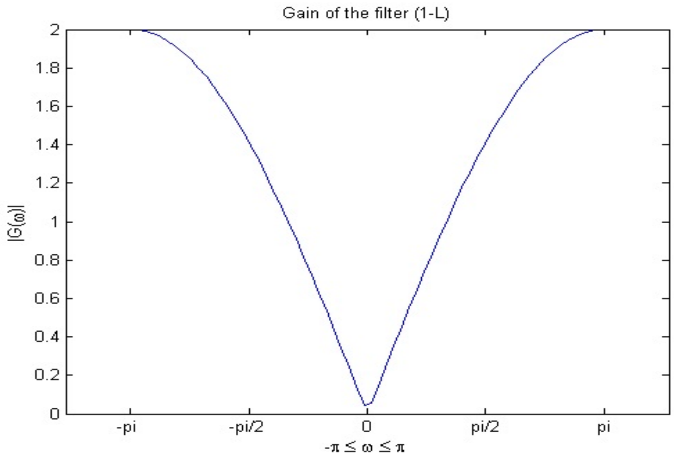

Figure 2.1: Gain of a first difference filter

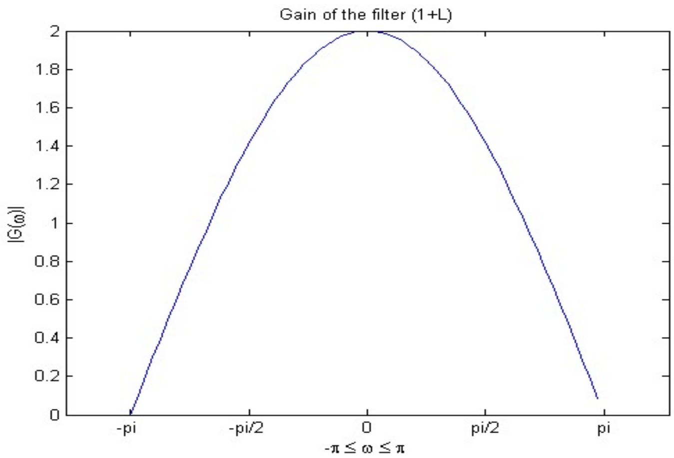

Figure 2.2: Gain of a first summation filter

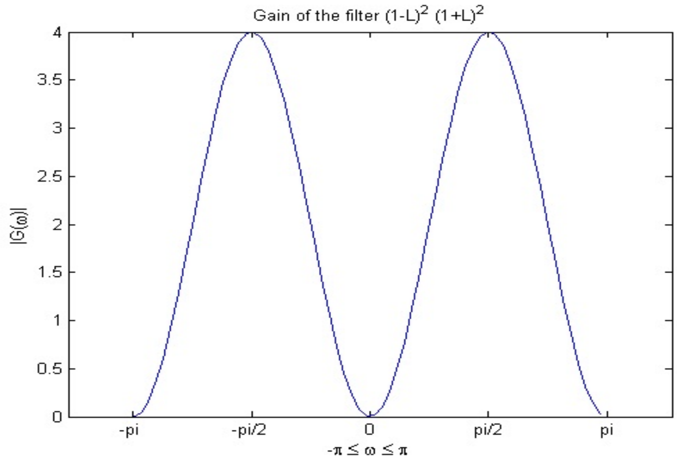

Figure 2.3: Gain of the filter $(1-L)^2(1+L)^2$

2.1.6 Christiano and Fitzgerald filter

Christiano and Fitzgerald (2003) (CF) developed optimal finite-sample approximation of the pass band filter. They used as weighting scheme the spectral density function $f_x(\omega)$, to obtain a filter as the solution of the minimization problem:

$$\min_{\hat{B}_j^{p,f}} \int_{-\pi}^{\pi} |B(e^{-i\omega}) - \hat{B}^{p,f}(e^{-i\omega})|^2 f_x(\omega) d\omega. \tag{2.7}$$

The rationale under this choice is to give a higher weight to more pronounced frequency. The CF filter differs from BK filter because the weights solution of (2.7) will be attached to X_t and not to the generating process u_t.

Simulations show that such a filter is more accurate in the selected range of frequencies.

If the estimate $\hat{x}_t$ of x_t is calculated by means of the observations $x_{-f}, \ldots, x_p$, symmetry of the filter can be obtained by choosing $f \equiv T - t = t - 1 \equiv p$. In general, the CF filter is not symmetric.The minimization problem depends on t, T different filters are obtained for each data, and hence the filter is not stationary.

The so called Random Walk (RW) filter is obtained in the minimization problem by putting $f_x(\omega)$ pseudo spectral density of $x(t)$ Random Walk, (or x_t ARIMA(p,1,q)).

$$f_x(\omega) = \frac{g(\omega)}{(1-e^{-i\omega})(1-e^{i\omega})},$$

where $(1-e^{-i\omega})(1-e^{i\omega}) = 2(1-\cos(\omega))$. Under this assumption, the process X_t does not belong anymore to $L^2([-\pi,\pi])$.

If ε_t is a $MA(q)$ process,

$$g(\omega) =\theta(e^{-i\omega})\theta(e^{i\omega}) = c_0 + c_1(e^{-i\omega} - e^{i\omega}) + \cdots + c_q(e^{-i\omega} + e^{-i\omega})$$
$$= c_0 + 2\sum_{r=1}^{q} c_r \cos(\omega r)$$

Such a filter is much more accurate in a neighborhood of $\omega = 0$ than for higher values of ω, even if $f_x(\omega)$ does not exists for $\omega = 0$. The estimation of $B_t^{p,f}$ involves the spectral density of x_t, that is unknown, and must be estimated from the data. In the simulation study of CF, the RW filter dominates both the Baxter and King filter, and the Trigonometric Regression filter.

The differences are most pronounced for filter approximations designed to extract frequencies lower than the business cycle. In the approach of CF the condition $\hat{B}^{p,f}(1) = 0$ is not imposed as a constraint, but arises from the hypothesis that the data contain a unit root.

The statistic used to compare the Hodrick-Prescott filter to the Random Walk filter is

$$R_t = \Big[\frac{Var_t(\hat{y}_t - y_t)}{Var(y_t)}\Big]^{1/2}.$$

This statistic represents the squared root of the residual variance when $\hat{y}_t$ is calculated by means of the observations available at time t and the total variance, thus a large value of R_t indicates a poor filter approximation.

The proper criterium to choose p and f is not clear, and it corresponds to the choice of the width of the filter, this complication in Christiano and Fitzgerald (2003) is sidestepped by the Random Walk filter, that uses all the data all the time.

2.1.7 Models for the stochastic cycle

Stochastic cycles are often used to model a business cycle (see, Harvey, 1993, Harvey and Trimbur 2003 among others). The reason is easily understood by examining the spectrum of the stationary process. For instance, for an AR(2) process, the spectral density is

$$f(\lambda) = \Big(\frac{\sigma^2}{2\pi}\Big)\Big(\frac{1}{1+\phi_1^2+\phi_2^2-2\phi_1(1-\phi_2)cos(\lambda)-2cos(2\lambda)}\Big).$$

If the root of the characteristic equation $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$ are complex, the

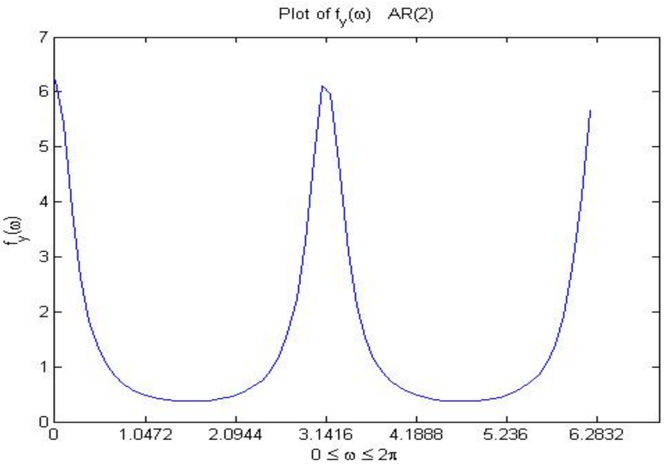

Figure 2.4: Spectral density Ma(2) complex roots

autocorrelation function led to a damped cyclical pattern, and this was interpreted as an indication of some kind of cyclical behavior in the series. A plot shows a peak that indicates a tendency toward a cycle at frequency λ_{max}, and this is an indicator of a pseudo-cyclical behavior. Harvey (1993) shows how a stochastic cycle may be formulated in term of an ARMA(2,1) process. If ψ_t is a deterministic sinusoidal trend: $\psi_t = \alpha \cos(\omega t) + \beta \sin(\omega t)$, a simple model for the cycle is $y_t = \psi_t + \varepsilon_t$. Introducing the complex conjugate process ψ_t^*, the same model can be putted in the form

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \begin{bmatrix} cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix}, \quad t = 1, \ldots, T.$$

with the initial conditions $\psi_0 = \alpha$, $\psi_0^* = \beta$. A first modification is given by adding two white noise disturbance κ_t e κ_t^*:

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \begin{bmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \quad t = 1, \ldots, T.$$

κ_t e κ_t^* are assumed to be uncorrelated and to have the same variance for the identifiability of the model. Further, a damping factor $\rho \in [0, 1]$ is introduced to give the model more flexibility:

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \quad t = 1, \ldots, T.$$

The reduced form for $(\psi_t, \psi_t^*)'$ shows that this process is a vector AR(1) process:

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \begin{bmatrix} 1-\cos(\omega)L & -\rho\sin(\omega)L \\ \rho\sin(\omega)L & 1-\rho\cos(\omega)L \end{bmatrix}^{-1} \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix},$$

and substituting the value of ψ_t in the definition of y_t gives

$$y_t = \frac{(1-\rho\cos(\omega)L)\kappa_t + (\rho\sin(\omega)L)\kappa_t^*}{1-2\rho\cos(\omega)L+\rho^2L^2} + \varepsilon_t, \quad t = 1, \dots, T.$$

that is

$$\begin{aligned} &y_t - 2\rho\cos(\omega)y_{t-1} + \rho^2 y_{t-2} = \\ &\quad \kappa_t - \rho\cos(\omega)\kappa_{t-1} + \rho\sin(\omega)\kappa_{t-1}^* + \varepsilon_t - 2\rho\cos(\omega)\varepsilon_{t-1} + \rho^2\varepsilon_{t-2}. \end{aligned}$$

Thus y_t is an ARMA(2,2) process, while ψ_t is an ARMA(2,1) process. The root of the AR polynomial are $m_1, m_2 = \rho^{-1}\exp(\pm i\omega)$, and they are complex conjugate for $0 < \omega < \pi$. The process is stationary for $0 \le \rho \le 1$. The analysis also shows that not every AR(2) process gives rise to a pseudo-cyclical behavior. For $\omega = 0$ or $\omega = \pi$ the process collapses to an AR(1). In these cases the dynamic of ψ_t is given by $\psi_t = \rho\psi_{t-1} + \kappa_t$ for $\omega = 0$ and by $\psi_t = -\rho\psi_{t-1} + \kappa_t$ for $\omega = \pi$.

The spectrum of ψ_t is

$$g_\psi(e^{-i\lambda}) = \frac{1+\rho^2-2\rho\cos(\omega)\cos(\lambda)}{1+\rho^4+4\rho^2\cos^2(\lambda)-4\rho(1+\rho^2)\cos(\omega)\cos(\lambda)+2\rho^2\cos(2\lambda)}\sigma_\kappa^2,$$

and its plot shows a peak for $\rho < 1$. The autocovariance function (ACF) of ψ_t is $\rho(\tau) = \varrho^\tau\cos(\omega\tau)$.

The model proposed by Harvey for the stochastic cycle can be putted in a vectorial MA form as

$$\Psi_t = \rho^t O^t \Psi_0 + \sum_{s=1}^{t} \rho^{t-s} O^{t-s} \Xi_s,$$

where $\Psi_t = [\psi_t, \psi_t^*]'$, $\Xi_t = [\kappa_t, \kappa_t^*]$, and O is the orthogonal matrix

$$\begin{pmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{pmatrix},$$

and O^t becomes

$$\begin{pmatrix} \cos\lambda t & \sin\lambda t \\ -\sin\lambda t & \cos\lambda t \end{pmatrix}.$$

The real part of the process can be interpreted as a harmonic process with a number of term depending on the same period length t:

$$\begin{aligned}\psi_t =&\rho^t[\cos(\lambda t)\psi_0+\sin(\lambda t)\psi_0^*]+\\ &+\sum_{s=1}^{t}\Big[\rho^{t-s}\kappa_{t-s}\cos\Big(\frac{(t-s)\lambda}{t}t\Big)+\rho^{t-s}\kappa_{t-s}^*\sin\Big(\frac{(t-s)\lambda}{t}t\Big)\Big].\end{aligned}$$

Harvey and Streibel (1998) give a different definition of stochastic cycle, distinguishing indeterministic cycle a cycle modeled as a $MA(\infty)$ process with a peak in its spectrum such as Beveridge and Nelson decomposition and deterministic cycle a cycle modeled as a harmonic process, whose spectral distribution function exhibits a sudden jump.

A modification of the stochastic cycle proposed by Harvey is (Harvey and Trimbur, 2003)

$$\begin{bmatrix}\psi_{1,t}\\ \psi_{1,t}^*\end{bmatrix}=\rho\begin{bmatrix}\cos(\omega) & \sin(\omega)\\ -\sin(\omega) & \cos(\omega)\end{bmatrix}\begin{bmatrix}\psi_{1,t-1}\\ \psi_{1,t-1}^*\end{bmatrix}+\begin{bmatrix}\kappa_t\\ 0\end{bmatrix},\quad t=1,\dots,T.$$

that yields $\psi_{1,t}=c(L)\kappa_t$, for

$$c(L)=\frac{1-\rho\cos(\omega)L}{1-2\cos(\omega)L+\rho^2L^2}.$$

Further, the i-th order stochastic cycle is defined as

$$\begin{bmatrix}\psi_{i,t}\\ \psi_{i,t}^*\end{bmatrix}=\rho\begin{bmatrix}\cos(\omega) & \sin(\omega)\\ -\sin(\omega) & \cos(\omega)\end{bmatrix}\begin{bmatrix}\psi_{i,t-1}\\ \psi_{i,t-1}^*\end{bmatrix}+\begin{bmatrix}\psi_{i-1,t}\\ 0\end{bmatrix},\quad t=1,\dots,T. \tag{2.8}$$

To model (2.8) it corresponds the Wiener-Kolmogorov filter

$$GB_n^{bp}(L)=\frac{q_\kappa c(L)^n c(L^{-1})}{q_\kappa c(L)^n c(L^{-1})+1},$$

$q_\kappa=\sigma_\kappa^2/\sigma_\varepsilon^2$. For $\rho=1$, it is obtained the band-pass Butterworth filter, corresponding to the gain function

$$B_n^{bp}(\lambda,\lambda_c)=\Big[1+\frac{1}{q}\Big(\frac{4(\cos\lambda-\cos\lambda_c)^2}{1+\cos^2\lambda_c-2\cos\lambda_c\cos\lambda}\Big)^n\Big]^{-1}.$$

For $\rho=1$ the spectrum is undefined at λ_c while the gain is still defined.

2.1.8 Seasonality

The seasonal adjustment of time series is the removal of a special cyclical component that is ascribable to climatic and institutional events repeated more or less regularly every year, and that is "nearly" predictable. Seasonal component seems to be "easily" recognized, and a common approach is to remove this component from rough or pre-treated data before searching for trend and cycle components.

The simplest model for the seasonal component (Dagum...) is a regression with dummy variable:

$$Y_t = S_t + \varepsilon_t, t = 1, \ldots, T,$$

$$S_t = \sum_{j=1}^{s} \gamma_j d_{jt} \quad \textit{subject to} \quad \sum_{j=1}^{s} \gamma_j = 0.$$

d is 4 for quarterly data, 12 for monthly data, $\{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2)$. $S_t = S_{t-s}$. This model is deterministic. A stochastic alternative is $S_t = S_{t-s} + \omega_t$ for all $t > s$ where $\{\omega\} \sim WN(0, \sigma_\omega^2)$ and $E(\omega_t \varepsilon_t) = 0$ and $\sum_{j=0}^{s-1} S_{t-j} = \omega_t$, $E(\omega_t) = 0$. The stochastic model can be written as $(1 - B^s)S_t = \omega_t$, and since $1 - B^s = (1 - B)(1 + B + \cdots + B^{s-1})$ the factor $1 - B$ gives rise to a stochastic trend, while the factor $S(B) = 1 + B + \cdots + B^{s-1}$ can be properly attributed to the seasonal component. Thus a model for seasonality is $S(B)S_t = \omega_t$, or $S(B)S_t = \eta_s(B)b_t$, with the right side being a moving average.

A model often used is estimating seasonality is X11ARIMA developed by Dagum (1978).

Hannan,Terrell and Tuckwell (1970) used spectral analysis to model the seasonal component of an economic time series, developing a technique for dealing with a changing seasonal pattern.

The authors compare a trigonometric model for the cycle

$$s_n = \sum_j \{\alpha_j \cos n\lambda_j + \beta_j \sin n\lambda_j\}, \quad \lambda_j = \frac{2\pi j}{12}.$$

with a harmonic model

$$s(n) = \sum_j \{\alpha_j(n) \cos n\lambda_j + \beta_j(n) \sin n\lambda_j\}, \quad \lambda_j = \frac{2\pi j}{12}.$$

α_j, β_i can be $AR(1)$ processes or more generally an $ARMA$ process.

2.2 Comments and Generalization of the presented models

2.2.1 Weighting frequencies with the density function in finite approximation: finite and infinite version

It is unrealistic to dispose of infinite observations, so it makes sense to build a finite version of approximate band pass filter:

$$y_t^* = B_0 x_t + B_1 x_{t+1} + \cdots + B_{T-1-t} x_{T-1} + \tilde{B}_{T-t} x_T + B_1 x_{t-1} + \cdots + B_{t-2} x_2 + \tilde{B}_{t-1} x_1,$$

where the B_k are defined as in the ideal band pass filter (2.4), and the $\tilde{B}_{T-t}$, $\tilde{B}_{t-1}$ linear functions of the B_j's. If only a finite set of observations $[x_1, \ldots, x_T]$ are available, the solution of the projection problem $\hat{y}_t = \sum_{j=-T+t}^{t-1} \tilde{B}_j^{p,f} x_{t-j}$ minimizing the mean squared error in general is not a symmetric filter.

The approximate band-pass filter proposed by Baxter and King does not take into account the properties of the random process generating the observed data x_t. A natural way of exploiting them is to weight the frequencies with the spectral density, if the process admits it. If X_t follows an $ARMA(p,q)$ process, that is $\phi(B)X_t = \theta(B)u_t$, for $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$, $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ and $u_t \sim WN(0, \sigma^2)$, $\phi(z)$ and $\theta(z)$ having roots outside the complex unit circle and not sharing roots, then the spectral density of X_t is

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2}, \quad -\pi \le \omega \le \pi.$$

The minimization problem can be solved in the frequency domain. In fact, exploiting the isomorphism between $L^2(F)$ (F spectral distribution function) and the probability space $L^2(\Omega)$ given by $I(e^{it}) = X_t$,

$$\begin{aligned}
E[(y_t - y_t^*)^2] &= ||y_t - y_t^*||_{L(\Omega)} = ||I(y_t) - I(y_t^*)||_{L([-\pi,\pi])} \\
&= ||Y(e^{-i\omega}) - Y^*(e^{-i\omega})||_{L([-\pi,\pi])} = ||B(e^{-i\omega}) - \hat{B}(e^{-i\omega})||||X(e^{-i\omega})|| \\
&= ||B(e^{-i\omega}) - \hat{B}(e^{-i\omega})||f_X(\omega).
\end{aligned}$$

that is

$$\frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} |B(e^{-i\omega}) - B^*(e^{-i\omega})|^2 \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2} d\omega. \tag{2.9}$$

The virtue of quadratic form (2.9) is that frequencies having higher gain receive a higher weight. This criterium coincides with the one applied by Baxter and King in the trivial case of $X_t \sim WN(0, \sigma^2)$, for which $f_X(\omega) = \frac{\sigma^2}{2\pi}$.

We impose $B_j^* = 0$ for $|j| > K$ since the property of minimality of the linear projection in $L^2([-\pi, \pi])$ allows to state that the best finite approximation of the Fourier series of $B(e^{-i\omega})\theta(e^{-i\omega})/\phi(e^{-i\omega})$ is given by the first K terms of its expansion, whose coefficients are obtained as a convolution product of the Fourier series of $B(e^{i\omega})$ with the Fourier series of $\theta(e^{-i\omega})/\phi(e^{-i\omega})$. Putting $\theta(z)/\phi(z) = \psi(z) = \sum_{k=-\infty}^{+\infty} \psi_k z^k$, we have $(B * \psi)_n = \sum_{k=-\infty}^{+\infty} B_k \psi_{n-k}$.

Thus the best approximation in L^2 of $\chi_I(e^{-i\omega})\theta(e^{-i\omega})/\phi(e^{-i\omega})$ under the constraint $G_j = 0$ for $|j| > K$ (that is minimizing the variance of $y_t - \hat{y}_t$) is

$$B^*(e^{-i\omega})\frac{\theta(e^{-i\omega})}{\phi(-i\omega)} = \sum_{m=-K}^{K} (B * \psi)_m e^{-im\omega}).$$

The weights calculated with this formula apply to u_t:

$$\hat{y}_t = \sum_{m=-K}^{K} (B * \psi)_m u_{t-m}.$$

and since $u_t = \phi(L)/\theta(L)X_t$, and we posses a sample of X_t and not of u_t, we get:

$$\hat{y}_t = \sum_{m=-K}^{K} (B * \psi)_m \frac{\phi(L)}{\theta(L)} \hat{x}_{t-m} = \sum_{m=-K}^{K} (B * \psi)_m \psi^{-1}(L) \hat{x}_{t-m}.$$

This expression contains only a finite number of x_t if x_t is an $MA(q)$ process. This filtered series is what we obtain when first Fourier-transforming the initial series, then cut out frequencies not needed and data distant in time, and then calculate the inverse Fourier transform when the spectral density of X_t is involved.

The non-constrained solution of the minimization problem is

$$\hat{y}_t = \sum_{m=-\infty}^{\infty} (B * \psi)_m \psi^{-1}(L) \hat{x}_{t-m} = ((B * \psi) * \psi^{-1})(L)\hat{x}_t.$$

From algebra of convolution products we have $(B * \psi) * \psi^{-1} = B * (\psi * \psi^{-1}) = B$, and then $\hat{y}_t = B(L)x_t$, that is that is the band pass filter discussed by Baxter and King, not involving the spectral density.

2.2.2 Smoothing of an $ARIMA(p,d,q)$ process in the frequency domain

For a general $ARIMA(p,d,q)$, defined $\tilde{B(\omega)} = \chi_I(\omega)/(1-e^{i\omega})^d$, $b(\omega) = \hat{B}(\omega)/(1-e^{i\omega})^d$, given the spectral density of the $ARMA(p,q)$ part $g(\omega) = \frac{\sigma^2}{2\pi}\psi(e^{i\omega})\psi(e^{-i\omega})$, the coefficients $\hat{B}_j$, $-k \leq j \leq k$ are found by minimizing in the frequency domain the functional

$$\int_{-\pi}^{\pi} |\tilde{B}(\omega) - b(\omega)|^2 g(\omega) d\omega,$$

which yields

$$\int_{-\pi}^{\pi} (\tilde{B}(\omega) - b(\omega)) e^{-il} g(\omega) d\omega = 0, \quad -k \leq l \leq k.$$

In the following paragraphs the solution to this problem will be solved in different cases.

2.2.3 Some calculation for an approximate band-pass filter - MA processes

In the following paragraph it will be calculated explicitly a finite version of the ideal band pass filter with density function as weighting density. Assume that X_t follows a $MA(1)$, process, so that its spectral density is $f_X(\omega) = \theta_0 + \theta_1(e^{-i\omega} + e^{i\omega}) = \theta_0 + 2\theta_1\cos(\omega)$ [1]. We want to determine the best approximate band-pass filter of order $p = 2$, $f = 2$:

$$\hat{y}_t = \hat{B}_{-2}x_{t-2} + \hat{B}_{-1}x_{t-1} + \hat{B}_0 x_t + \hat{B}_{t+1}x_{t+1} + \hat{B}_{t+2}x_{t+2}.$$

The coefficient $\hat{B}_{-2}, \ldots, \hat{B}_2$ are obtained by minimizing the functional:

$$F(\hat{B}_{-2}, \ldots, \hat{B}_2) = \int_{-\pi}^{\pi} |\chi_I(\omega) - \sum_{j=-2}^{2} \hat{B}_j e^{ij\omega}|^2 [\theta_0 + \theta_1(e^{i\omega} + e^{-i\omega})] d\omega,$$

where $I = (-b,-a) \cup (a,b)$, $0 < a < b < \pi$.

The conditions

$$\frac{\partial F}{\partial \hat{B}_k} = 0, \quad k = -2, \ldots, 2,$$

[1] Here the notation is slightly different from other books. The spectral density of a $MA(1)$ process is usually written as

$$f_X(\omega) = \frac{\sigma^2}{2\pi}|1 + \theta e^{-i\omega}|^2 = \frac{\sigma^2}{2\pi}(1 + 2\theta\cos(\omega) + \theta^2).$$

It is $\theta_0 = \frac{\sigma^2}{2\pi}(1+\theta^2)$, $\theta_1 = \frac{\sigma^2}{2\pi}\theta$.

yield

$$\int_{-\pi}^{\pi} \chi_I(\omega)e^{ik\omega}[\theta_0+\theta_1(e^{i\omega}+e^{-i\omega})]d\omega =$$
$$\int_{-\pi}^{\pi} \sum_{j=-2}^{2} \hat{B}_j e^{ij\omega}e^{ik\omega}[\theta_0+\theta_1(e^{i\omega}+e^{-i\omega})]d\omega, \quad k=-2,\ldots,2;$$

$$\int_a^b [e^{ik\omega}+e^{-ik\omega}][\theta_0+\theta_1(e^{i\omega}+e^{-i\omega})]d\omega =$$
$$\sum_{j=-2}^{2} \hat{B}_j \int_{-\pi}^{\pi} e^{i(j+k)\omega}[\theta_0+\theta_1(e^{i\omega}+e^{-i\omega})], \quad k=-2,\ldots,2;$$

$$2\int_a^b \{\theta_0\cos(k\omega)+\theta_1\cos((k+1)\omega)+\theta_1\cos((k-1)\omega)\}d\omega =$$
$$\sum_{j=-2}^{2} \hat{B}_j \int_{-\pi}^{\pi} [\theta_0 e^{i(j+k)\omega}+\theta_1 e^{i(k+j+1)\omega}+\theta_1 e^{i(k+j-1)\omega}]d\omega, \quad k=-2,\ldots,2;$$

$$2\frac{\theta_0}{k}\big[\sin(kb)-\sin(ka)\big]+2\frac{\theta_1}{k+1}\big[\sin((k+1)b)-\sin((k+1)a)\big]$$
$$+\,2\frac{\theta_1}{k-1}\big[\sin((k-1)b)-\sin((k-1)a)\big] =$$
$$\sum_{j=-2}^{2} \hat{B}_j \Big[\frac{\theta_0}{i(j+k)}e^{i(j+k)\omega}+\frac{\theta_1}{i(j+k+1)}e^{(j+k+1)\omega}+\frac{\theta_1}{i(j+k-1)}e^{i(j+k-1)\omega}\Big]_{\omega=-\pi}^{\omega=\pi},$$
$$k=-2,\ldots,2;$$

In the last equation one must replace, for $l=0$,

$$\frac{\sin(lb)-\sin(la)}{l} \quad \text{with} \quad b-a,$$

and, in the same manner,

$$\frac{e^{il\pi}-e^{-il\pi}}{il} \quad \text{with} \quad 2\pi.$$

For $l \neq 0$, $\frac{e^{il\pi}-e^{-il\pi}}{il}=0$.

Since $k+j=0$ if and only if $j=-k$, $k+j+1$ if and only if $j=-(k+1)$, $k+j-1=0$

if and only if $j = -(k-1)$, the second term of the last equality is reduced to

$$2\pi\{\theta_0 \hat{B}_{-k} + \theta_1 \hat{B}_{-(k+1)} + \theta_1 \hat{B}_{-(k-1)}\}.$$

Thus, the $\hat{B}_j$ are the solution of the system of equations:

$$\begin{aligned}
&\theta_0 \hat{B}_{-k} + \theta_1 \hat{B}_{-(k+1)} + \theta_1 \hat{B}_{-(k-1)} = \\
&\theta_0 \frac{\sin(kb) - \sin(ka)}{\pi k} + \theta_1 \frac{\sin((k+1)b) - \sin((k+1)a)}{\pi(k+1)} \\
&+ \theta_1 \frac{\sin((k-1)b) - \sin((k-1)a)}{\pi(k-1)}, \quad k = -2, \ldots, 2.
\end{aligned}$$

where $\hat{B}_l = 0$ for $l \notin \{-2, \ldots, 2\}$.

Defining the tridiagonal matrix

$$\Theta_1 = \begin{pmatrix} \theta_0 & \theta_1 & 0 & 0 & 0 \\ \theta_1 & \theta_0 & \theta_1 & 0 & 0 \\ 0 & \theta_1 & \theta_0 & \theta_1 & 0 \\ 0 & 0 & \theta_1 & \theta_0 & \theta_1 \\ 0 & 0 & 0 & \theta_1 & \theta_0 \end{pmatrix}$$

the vector θ as $[0, \theta_1, \theta_0, \theta_1, 0]'$ the vector $\hat{B} = [\hat{B}_{-2}, \ldots, \hat{B}_2]'$ and B as the sequence $\left\{\frac{\sin(kb)-\sin(ka)}{\pi k}\right\}_{k \in \mathbf{Z}}$ (that is the Fourier transform of the exact band-pass filter), the translated of B $L^l B$ as $\left\{\frac{\sin((k-l)b)-\sin(k-l)a)}{\pi(k-l)}\right\}_{k \in \mathbf{Z}}$, $\theta^* B$ as the vector of the convolutions $\theta^* L^l B$, $l = -2, \ldots, 2$ this system of equation is written in matrix notation as:

$$\Theta_1 \hat{B} = \theta^* B.$$

Finally, since $det(\Theta_1) \neq 0$ (Θ_1 is a diagonally dominant matrix) the solution is $\hat{B} = \Theta_1^{-1} \theta^* B$.

The right side is also written as $\Theta_1^* B|_3$, where

$$\Theta_1^* = \begin{pmatrix} \theta_1 & \theta_0 & \theta_1 & 0 & 0 & 0 & 0 \\ 0 & \theta_1 & \theta_0 & \theta_1 & 0 & 0 & 0 \\ 0 & 0 & \theta_1 & \theta_0 & \theta_1 & 0 & 0 \\ 0 & 0 & 0 & \theta_1 & \theta_0 & \theta_1 & 0 \\ 0 & 0 & 0 & 0 & \theta_1 & \theta_0 & \theta_1 \end{pmatrix}$$

and $B|_3 = \{\frac{\sin(-3b)-\sin(-3a)}{-3\pi}, \ldots, \frac{\sin(3b)-\sin(3a)}{3\pi}\}$.

Theorem 4. *For a general $MA(q)$ process, the $2k+1$ unknown $\hat{B}_k$ are determined by $\hat{B} = \Theta_1^{-1}\Theta_1^* B|_{k+q}$, where$\Theta_1$ is a Toeplitz band matrix of dimension $(2k+1) \times (2k+1)$ having $\{\Theta_1\}_{i,j} = \theta_{|i-j|}$ if $|i-j| \leq q$, $\{\Theta_1\}_{i,j} = 0$ otherwise, and Θ_1^* is a $(2k+1) \times (2k+2q+1)$ band matrix having $\{\Theta_1^*\}_{i,j} = \theta_{|j-i-q|}$.*

These weights are symmetric. They do not sum to zero.

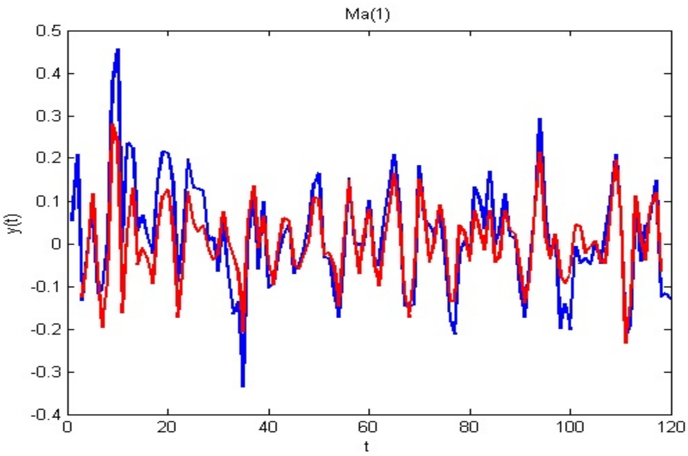

Figure 2.5: Smoothed MA(1) process

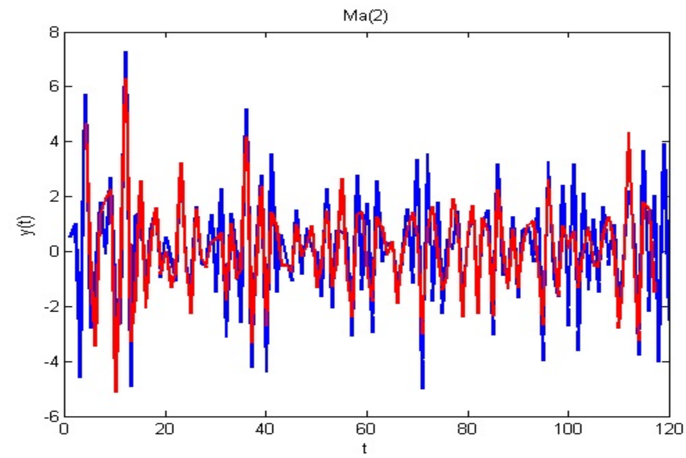

Figure 2.6: Smoothed MA(2) process

2.2.4 A Christiano - Fitzgerald filter for I(d) processes

The approach of Christiano and Fitzgerald can be generalized in order to filter time series which present more than one unit root. Infact if X_t follows an $ARMA(p,d,q)$ process, it is enough to set in the minimization criterion

$$f_x(\omega) = \frac{\sigma^2}{2\pi} \frac{\theta(e^{i\omega})\theta(e^{-i\omega})}{\phi(e^{i\omega})\phi(e^{-i\omega})} \frac{1}{(1-e^{-i\omega})^d(1-e^{i\omega})^d}.$$

Also in this case, if a solution exists, it must be $\frac{\hat{B}^{f,p}(z)}{(1-z)^d} \neq \infty$ for $z=1$, in order to make the integral in (2.7) to converge, and this implies

$$b(z) = \frac{\hat{B}^{f,p}(z)}{(1-z)^d} = b_{p-d}z^{p-d} + b_{p-d-1}z^{p-d-1} + \cdots + b_0 + \cdots + b_{-f+1}z^{-f+1}z^{-f}.$$

If $\hat{B}^{p,f}$ is the vector of the coefficients of the Laurent polynomial $\hat{B^{p,f}}(z)$ and b is the vector of the coefficients of the Laurent polynomial $b(z)$, the link between $\hat{B}^{p,f}$ and b is expressed by: $Q^d\hat{B}^{p,f} = b$, where Q is a $(p+f+1-d)\times(p+f+1)$ matrix, $Q_d = [Q_1^d, \mathbf{0}]$, where $(-1)^dQ_1^d$ is a $(p+f+1-d)\times(p+f+1-d)$ low-triangular matrix whose first column of Q_1^d is the d-st diagonal of the Pascal triangle, and the n-st column is obtained by shifting the $n-1$-st column, and $\mathbf{0}$ is a zero $(p+f+1-d)\times d$ matrix.

$$(-1)^dQ_1^d = \begin{pmatrix} \binom{d-1}{d-1} & 0 & 0 & \cdots & 0 \\ \binom{d}{d-1} & \binom{d-1}{d-1} & 0 & \cdots & 0 \\ \binom{d+1}{d-1} & \binom{d}{d-1} & \binom{d-1}{d-1} & \cdots & 0 \\ \vdots & & & \cdots & 0 \\ \binom{2d-2}{d-1} & \binom{2d-3}{d-1} & \binom{2d-4}{d-1} & \cdots & \binom{d-1}{d-1} \end{pmatrix}.$$

For example, if $d=5$, $p+f+d-1$=6, then

$$-Q_1^5 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 5 & 1 & 0 & 0 & 0 & 0 \\ 15 & 5 & 1 & 0 & 0 & 0 \\ 35 & 15 & 5 & 1 & 0 & 0 \\ 70 & 35 & 15 & 5 & 1 & 0 \\ 126 & 70 & 35 & 15 & 5 & 1 \end{pmatrix}.$$

Calculation proposed in the paper are easily generalized: $d' = A\hat{B}^{f,p}$,

$$\int_{-\pi}^{\pi} \tilde{B}(e^{-i\omega})g(\omega)e^{i\omega j}d\omega = 2\pi F_j Q_d \hat{B}^{p,f},$$

where

$$\tilde{B}^{f,p}(e^{-i\omega}) = B(e^{-i\omega})/(1-e^{-i\omega})^d = \chi_I(\omega)/(1-e^{-i\omega})^d;$$

F_J is the row vector

$$F_j = [0, 0, ..., c, 0, \ldots, 0],$$

where the first $p-q-d-j$ and the last $j-q+f$ positions are zero, for a $MA(q)$ stationary component, and

$$c = [c_q, c_{q-1}, \ldots, c_0, \ldots, c_{q-1}, c_q]$$

is the vector of the autocovariances;

Under the hypothesis that $\hat{B}(z)$ do not possesses a zero for $z=1$ up to $d-1$ order, then for $h = 0, \ldots, d-1$, the identity $\hat{B}_h^{p,f}(z) = \hat{B}^{p,f}(z)/(1-z)^h$ defines a polynomial for which $\hat{B}_h^{p,f}(1) = 0$. These identities give rise to the last d equations of the system that allows to determine the $\hat{B}_j^{p,f}$:

$$d' = \begin{bmatrix} \int_{-\pi}^{\pi} \tilde{B}(e^{-i\omega})g(\omega)e^{i\omega(p-1)}d\omega \\ \vdots \\ \int_{-\pi}^{\pi} \tilde{B}(e^{-i\omega})g(\omega)e^{i\omega(-f+1)}d\omega \\ \int_{-\pi}^{\pi} \tilde{B}(e^{-i\omega})g(\omega)e^{i\omega(-f)}d\omega \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad A = \begin{bmatrix} F_{p-1}Q_d \\ \vdots \\ F_{-f+1}Q_d \\ F_{-f}Q_d \\ 1, \binom{n-1}{2}, \ldots, \binom{n-1}{d} \\ \cdots \\ 1, 3, \ldots, \binom{n-d+1}{2} \\ 1, 2, \ldots, n-d \\ 1, 1, \ldots 1 \end{bmatrix}.$$

It may be useful to derive an Integrated Random Walk filter, which correspond to $d=2$, and can be compared to the Hodrick-Prescott filter. In this case the matrix Q_1^2, for $p+f-1=6$ is

$$Q_1^2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 \\ 3 & 2 & 1 & 0 & 0 & 0 \\ 4 & 3 & 2 & 1 & 0 & 0 \\ 5 & 4 & 3 & 2 & 1 & 0 \\ 6 & 5 & 3 & 3 & 2 & 1 \end{pmatrix}.$$

Also the recursive calculation suggested for the elements of the vector d' still holds:

$$R(j) = \int_{-\pi}^{\pi} \tilde{B}(e^{-\omega}) g(\omega) e^{i\omega j} d\omega,$$

we have

$$R(0) = \int_a^b \frac{2\sum_{k=0}^{d} \binom{d}{k}(-1)^k \cos(\omega k)}{2^d \sum_{k=0}^{d} \binom{d}{k}(-1)^k \cos(\omega)^k} g(\omega) d\omega.$$

$$R(j) - R(j+1) = \int_a^b \Big[\frac{e^{-i\omega j}}{(1-e^{-i\omega})^{d-1}} + \frac{e^{i\omega j}}{(1-e^{i\omega})^{d-1}}\Big] g(\omega) d\omega.$$

These integrals are evaluated numerically.

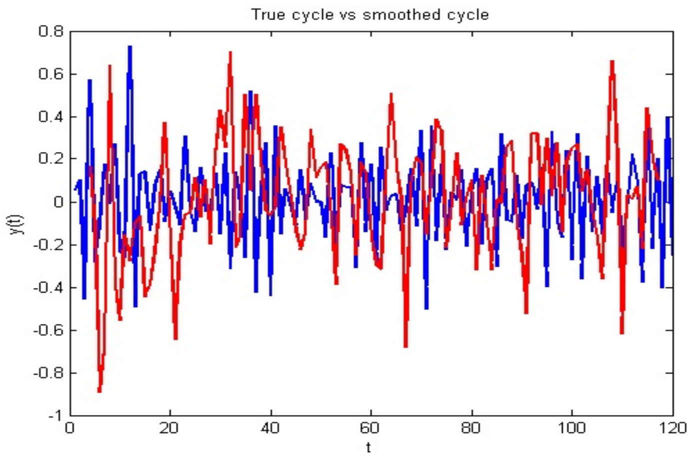

Figure 2.7: IMA(2,2) process smoothed by generalized CF filter. The figure shows only the cyclical component of the MA(2)

In the simulations conducted, the generalized CF filter is applied to the original simulated series. The figure compares the cyclical MA part of the process to the smoothed series. Source codes are available on request.

Chapter 3

A Local Trigonometric Model for smoothing cycles

In this chapter we will introduce the main parameters which characterize local estimators. There are presented models discussed in literature such as the local polynomial regression and the polynomial spline. Then it is proposed a local trigonometric smoother, with its statistical properties, and it is applied to the smoothing of a simulated pseudo-cyclical process. Some insights is given for the choice of the parameters of interest and for the testing of hypothesis. Connections between trigonometric filter and ideal band pass filter are highlighted.

3.1 General properties of local fitting

In chapter 2 we pointed out that in an economic context it is impossible to dispose of an as long series observations as infinite filters would require. Moreover, often they are needed quick estimation for a parameter. Further, a reasonable conjecture is that observation closer to a specific point y_t would help to predict the same y_t better than distant observations, so a "local" estimator would furnishes more precise estimates.

An estimator is said to be *local* if it predicts y_t only by means of observations taken in a neighborhood of y_t.

Several methods are known in literature to perform local fitting of time series. Among these we point out kernel estimators, local polynomial fitting, wavelets, splines, orthogonal series (see Fan and Gijbels, 1996). For a review relative to local polynomial regression and polynomial splines see Proietti and Luati (2007).

Kernel estimators allow asymptotic bias corrections, whereas local regression provides finite sample solutions to the same problems.

3.1.1 Main parameters of a local model

In a local fitting they must be specified

- the bandwidth,
- the degree of the local trigonometric polynomial,
- the weight function,
- the fitting criterion.

The choice of the model depends on

- the variance reducing factor.
- the influence function.
- the degree of freedom.

The local regression estimate is said to be *linear* if for each t there exists a weight diagram vector $l(t) = \{l_i(t)\}_{i=1}^n$ such that the estimate can be written as $\hat{\mu}(t) = \sum_{i=1}^n l_i(t)y_i$. The *variance reducing factor* $||l(t)||^2$ measures the reduction in variance due to the local regression. Under mild condition one can show that (Loader, 1999)

$$\frac{1}{n} \leq ||l(x)||^2 \leq l_i(t_i) \leq 1.$$

The extreme cases $1/n$ and 1 correspond, respectively, to $\hat{\mu}(t)$ being the sample average and interpolating the data. The is the $n \times n$ matrix which maps the data into the fitted values:

$$\begin{pmatrix} \hat{\mu}(t_1) \\ \vdots \\ \hat{\mu}_n \end{pmatrix} = LY.$$

It has rows $(l_1(t_i), \ldots, l_n(t_i))$, $i = 1, \ldots, n$.

The *influence* or *leverage* values are the diagonal elements $infl(t_1) = l_i(t_i)$ of the matrix L. These measure the sensitivity of the fitted curve $\hat{\mu}(t_i)$ to the individual data points.

The *degrees of freedom* of a local fit provide a generalization of the number of parameters of a parametric model. They can be defined as

$$\nu_1 = \sum_{i=1}^{n} infl(t_i) = tr(L);$$

or as

$$\nu_2 = \sum_{i=1}^{n} ||l_i(t_i)||^2 = tr(L^T L).$$

The two definitions of degrees of freedom coincide if L is a symmetric and idempotent matrix, but in general $1 \leq \nu_2 \leq \nu_1 \leq n$.

Choosing the length of a graduation rule, or bandwidth, involves a compromise between systematic error (bias) and random error (variance)

In a discrete setting a filter is given by any one of the rows of the matrix $W = L$ (Proietti and Luati, 2007). Thus it is possible to investigate the effect of the filter induced on a particular sequence $y_i = \cos(\omega t)$, where ω is the frequency in radians. Applying standard trigonometric identities, the filtered series is

$$\begin{aligned}
\sum_j w_j y_{t-j} &= \sum_j w_j \cos(\omega(t-j)) \\
&= \sum_j w_j \cos(\omega t)\cos(\omega j) + \sum_j w_j \sin(\omega t)\sin(\omega j) \\
&= \alpha(\omega)\cos(\omega t) + \alpha^*(\omega)\sin(\omega t) \\
&= G(\omega)\cos(\omega t - \theta(\omega))
\end{aligned}$$

where $\alpha(\omega) = \sum_j w_j \cos(\omega j)$, $\alpha^*(\omega) = \sum_j w_j \sin(\omega j)$.

The function

$$G(\omega) = \sqrt{\alpha^2(\omega) + \alpha^{*2}(\omega)}$$

is the gain of the filter and measures how the amplitude of the periodic components that make up a signal are modified by the filter. If the gain is 1 at a particular frequency, this implies that the periodic component defined at that frequency is preserved; if the gain is less than 1 for some frequency, that frequency is compressed.

The function

$$\theta(\omega) = \arctan\left[\frac{\alpha^*(\omega)}{\alpha(\omega)}\right]$$

is the phase function and measures the displacement of the periodic function along the time axis. For symmetric filters the phase function is zero, since $\sum_j w_j \sin(\omega j) = 0$.

If the trigonometric trend presents more frequencies in general each frequency has different gain and phase shift, in fact

$$y_t = \sum_{i=1}^{n} \{a_i \cos(\omega_i t) + b_i \sin(\omega_i t)\} = \sum_{i=1}^{n} r_i \cos(\omega_i t - \theta_i),$$

with $r_i = \sqrt{a_i^2 + b_i^2}$ and $\theta_i = \arctan{(b_i/a_i)}$ when smoothed gives rise to

$$\sum_j w_j y_{t-j} = \sum_i r_i \Big\{ \Big[\sum_j -w_j \sin(\omega_i j) \Big] \cos(\omega_i t - \theta_i) + \Big[\sum_j w_j \cos(\omega_i j) \Big] \sin(\omega_i t - \theta_i) \Big\},$$

so that the frequency ω_i receives a gain and a phase shift given respectively by

$$G(\omega_i) = \sqrt{\alpha(\omega_i)^2 + \alpha^*(\omega_i)^2}, \quad \theta(\omega_i) = \arctan \frac{\alpha^*(\omega_i)}{\alpha(\omega_i)}.$$

3.1.2 Estimation of the goodness of fit

In the following section there will be presented some useful statistics to test the goodness of fit of a local model.

$$Var(\hat{m}_t) = E[\hat{m}_t - E(\hat{m}_t)]^2 = E\Big[\sum_j w_j (y_{t-j} - \mu_{t-j}) \Big]^2 =$$
$$\sigma^2 \sum_j W_j^2 = \sigma^2 e_1' (X'KX)^{-1} X'K^2 X (X'KX)^{-1} e_1.$$

The term $\sum_j W_j^2 = \sigma^2$ is addressed as *variance inflation factor*, and it represents the proportionate increase in the variance of a filtered white noise sequence after the smoothing.

If $\hat{m}_{t \backslash t}$ is the two-sided estimate of the signal at time t that doesn't use y_t then

$$\hat{m}_{t \backslash t} = e_1' (X'KX - \kappa_0 e_1 e_1')^{-1} (X'Ky - \kappa_0 y_t e_1) = \frac{1}{1 - w_0} \hat{m}_t - \frac{w_0}{1 - w_0} y_t,$$

and the leave-one-out residual leave-one-out residual is

$$y_t - \hat{m}_{t \backslash t} = \frac{1}{1 - w_0} (y_t - \hat{m}_t).$$

The cross-validation score is the sum of the squared deletion residuals:

$$CV = \sum_{t=1}^{n} (y_t - \hat{m}_{t\backslash t})^2 = \sum_t \frac{(y_t - \hat{m}_t)^2}{(1 - w_{0t})^2},$$

where one writes w_{0t} since the filter weights are different at the extremes of the sample.

The estimation of σ^2 can be done by using the residuals from the local polynomial fit: $y_t - \hat{m}_t = y_t - \sum_j w_{jt} y_{t-j}$. Under the hypothesis of a polynomial trend of degree p and thanks to the polynomial preservation property of the filter, the expectation of the residual sum of squares (RSS) is:

$$E(RSS) = E\Big[\sum_{t=1}^{n} \Big(y_t - \sum_j w_{jt} y_{t-j}\Big)^2\Big] = \sigma^2 \Big[n - 2\sum_{t=1}^{n} w_{0t} + \sum_{t=1}^{n} \Big(\sum_j w_{jt}^2\Big)\Big].$$

This suggests using the following estimator for the error variance:

$$\hat{\sigma}^2 = \frac{RSS}{n - 2\sum_{t=1}^{n} w_{0t} + \sum_{t=1}^{n} \Big(\sum_j w_{jt}^2\Big)}.$$

An approximate 95% confidence interval for μ_t is

$$\hat{m}_t \pm 2\Big(\hat{\sigma}^2 \sum_j w_{jt}^2\Big)^{\frac{1}{2}}.$$

3.2 Some examples of local model

Very popular kernel smoothers are the Henderson smoother, the Macaulay smoother, the Epanechnikov smoother.

3.2.1 Local polynomial regression

In general it is assumed an additive model as

$$y_t = \mu_t + \epsilon_t, \quad t = 1, \ldots, n,$$

where μ_t is the trend component, also termed the signal, and ϵ_t is the noise, or irregular component. μ_t can be either stochastic or deterministic. In the case of a deterministic trend it is often assumed to be p times differentiable in t. If $2h+1$ equispaced observations y_{t+j}, $|j| \leq h$ are available in a neighbor of t, then μ_{t+j} is approximated by its truncated

Taylor polynomial of order p:

$$m_{t+j} = \beta_0 + \beta_1 j + \cdots + \beta_p j^p, \quad j = 0, \pm 1, \ldots, \pm h.$$

The problem of finding the vector of coefficients β is known as local polynomial regression. p is the degree of the approximation and h is the bandwidth. The local polynomial model is

$$y_{t+h} = \sum_{k=0}^{p} \beta_k j^k + \epsilon_{t+j}, \quad j = 0, \pm 1, \ldots, \pm h,$$

in matrix notation $y = X\beta + \epsilon$, $\epsilon \sim N(0, \sigma^2 I)$. The vector β is chosen as the vector minimizing a weighted squares summation:

$$S(\hat{\beta}_0, \ldots, \hat{\beta}_p) = \sum_{j=-h}^{h} \kappa_j (y_{t+j} - \hat{\beta}_0 - \hat{\beta}_1 j - \cdots - \hat{\beta}_p j^p)^2$$

for a suitable choice of kernel function κ_j, $j = 0, \pm 1, \ldots, \pm h$ such that $\kappa_j \geq 0$ and $\kappa_j = \kappa_{-j}$. Such a kernel function is time-invariant. In matrix notation the solution is $\hat{\beta} = (X'KX)^{-1}X'Ky$. In particular, if $e_1 = [1, 0, \ldots, 0] \in \mathbb{R}^{p+1}$, $\hat{m}_t = \hat{\beta}_0$ is given by

$$\hat{m}_t = e_1'\hat{\beta} = e_1'(X'KX)^{-1}X'Ky = w'y = \sum_{j=-h}^{h} w_j y_{t-j}.$$

The vector $w = e_1'(X'KX)^{-1}X'K$ is a filter. It is symmetric since κ is symmetric. The condition $X'w = e_1$ is equivalent to

$$\sum_{j=-h}^{h} w_j = 1, \quad \sum_{j=-h}^{h} j^l w_j = 0, \quad l = 1, \ldots, p.$$

These conditions imply that the filter w preserves a polynomial of degree p, that is it reproduces it exactly, and in this case $\hat{m}_t = \beta_0 = y_t$. The central element of the vector w, w_0, represents the leverage, as defined above, that is the contribution of y_t on the estimate of the signal at time t.

Luati and Proietti (2010) establishes the conditions under which the generalized least squares of the regression parameters is equivalent to the weighted least squares estimator. The equivalence conditions allows to derive the optimal kernel associated with a particular covariance structure of the measurement error.

3.2.2 Polynomial splines

An alternative way of overcoming the limitations of a global polynomial model is represented by a polynomial spline. Given the set of points $t_1 < \cdots < t_i < \cdots < t_k$, a polynomial spline function of degree p with k knots $t_1, \ldots, t_k$ is a polynomial of degree p in each interval $[t_i, t_{i+1}[$, with $p-2$ continuous derivatives whereas the $(p-1)$st derivative can have jumps at the knots. It can be represented as:

$$\mu(t) = \beta_0 + \beta(t - t_1) + \cdots + \beta_p(t - t_1)^p + \sum_{i=1}^{k} \eta_i (t - t_i)_+^p,$$

where

$$(t - t_i)_+^p = \begin{cases} (t - t_i)^p, & t \geq t_i, \\ 0, & t < t_i. \end{cases}$$

It has been pointed out that the piecewise nature of the spline "reflects the occurrence of structural changes". The knots t_i are the timing of a structural break. The coefficients η_i regulate the size of the break, and can be considered fixed or random; in the second case the function $(t - t_1)_+^p$ describes the impulse response function, that is the impact of the future values of the trend.

For η_i random, the spline model can be formulated as a linear mixed model. Denoting $y = [y(t_1), \ldots, y(t_n)]'$, $\eta = [\eta_1, \ldots, \eta_n]'$, $\epsilon = [\epsilon(t_1), \ldots, \epsilon(t_n)]'$, $\mu = X\beta + Z\eta$, the spline model is

$$y = \mu + \epsilon = X\beta + Z\eta + \epsilon,$$

where the t-th row of X is $[1, (t - 1), \ldots, (t - 1)^p]$, and Z is a known matrix whose i-th column contains the impulse-response signature of the shock η_i, $(t - t_i)_+^p$.

The spline model encompasses several type of models, such as the local level model ($p = 0$), the local linear trend model ($p = 1$), that is an integrated random walk, and the cubic spline. The cubic splines displays too much flexibility for economic time series, that is paid for with excess variability, especially at the beginning and and at the end of the sample period. Out of the sample forecast tend to be not very reliable, as they are subject to high revision as new observations become available. This flexibility is usually limited by imposing the so called boundary conditions, which constrain the spline to be linear outside the boundary knots, i.e., the second and third derivative are zero for $t \leq 1$ and $t \geq n$.

A smoothing spline is a natural cubic spline which solves the following penalized least square indexpenalized least square(PLS) problem

$$\min\Big\{(y-\mu)'(y-\mu)+\tau\int[\mu''(t)]^2dt\Big\}. \tag{3.1}$$

Minimizing the PLS objective function is equivalent to maximizing the posterior density $f(\mu|y)$ assuming the prior density $\gamma \sim N(0, \sigma^2_\gamma R^{-1})$, R being a suitable diagonally dominant tridiagonal matrix for the smoothing spline, $\tau = \sigma^2_\epsilon/\sigma^2_\gamma$.

3.3 Local Trigonometric Regression

In chapter 2 we examined and extended the finite sample approximation of the exact band-bass filter by working in the domain of frequency. In particular, we saw that the optimal approximating smoother can be expressed as a finite sum of trigonometric function. In this section we focus on time domain, and we shall build a local trigonometric filter.

Trigonometric regression has been studied by Walker (1971), Hannan (1973), Quinn (1979), Wang (1993) among others.

The minimization of the sum of squares of a local trigonometric regression in time domain is

$$F(c)=\sum_{t=-T/2}^{T/2}\kappa_{j,t}\Big|y_{t+j}-\sum_{k=0}^{\lambda}c_k e^{ixtk}\Big|^2 = min_{c_k}, \tag{3.2}$$

where c_k are assumed to be real. In the remaining part of the paragraph it is written κ_t instead of $\kappa_{j,t}$ for simplicity.

$\frac{\partial F}{\partial c_k}=0$ yields

$$\sum_{t=-T/2}^{T/2}\kappa_t y_{t+j}\cos(xtk)=\sum_{t=-T/2}^{T/2}\sum_{l=0}^{\lambda}\kappa_t c_l\cos(xt(k-l)),\quad k=0;\ldots,\lambda.$$

The simplest case is the uniform kernel: $\kappa_t = 1$, $\forall t$. In this case we have:

$$\sum_{t=-T/2}^{T/2}e^{rit}=\frac{\sin(r\frac{(T+1)}{2})}{\sin(\frac{r}{2})},\quad r\neq k\pi,$$

$$\sum_{t=-T/2}^{T/2} e^{rit} = T+1, \quad r = 2k\pi,$$

$$\sum_{t=-T/2}^{T/2} e^{rit} = -T-1, \quad r = (2k+1)\pi,$$

and thus

$$\sum_{t=-T/2}^{T/2} y_{t+j}\cos(xtk) = \sum_{l=0}^{\lambda} c_l \frac{\sin(x(\frac{T+1}{2})(k-l))}{\sin(x\frac{(k-l)}{2})}, \quad k = 0;\ldots,\lambda.$$

For instance, if $k = 3$, this system of equation is explicitly written as:

$$\begin{aligned}
a(T+1) + b\frac{\sin\frac{x(T+1)}{2}}{\sin(\frac{x}{2})} + c\frac{\sin x(T+1)}{\sin(x)} + d\frac{\sin\frac{3x(T+1)}{2}}{\sin(\frac{3x}{2})} &= \sum_{t=-T/2}^{T/2} y_{j+t} \\
a\frac{\sin\frac{x(T+1)}{2}}{\sin(\frac{x}{2})} + b(T+1) + c\frac{\sin\frac{x(T+1)}{2}}{\sin(\frac{x}{2})} + d\frac{\sin x(T+1)}{\sin(x)} &= \sum_{t=-T/2}^{T/2} \cos(xt) y_{j+t} \\
a\frac{\sin x(T+1)}{\sin(x)} + b\frac{\sin\frac{x(T+1)}{2}}{\sin(\frac{x}{2})} + c(T+1) + d\frac{\sin\frac{x(T+1)}{2}}{\sin(\frac{x}{2})} &= \sum_{t=-T/2}^{T/2} \cos(2xt) y_{j+t} \\
a\frac{\sin\frac{3x(T+1)}{2}}{\sin(\frac{3x}{2})} + b\frac{\sin x(T+1)}{\sin(x)} + c\frac{\sin\frac{x(T+1)}{2}}{\sin(\frac{x}{2})} + d(T+1) &= \sum_{t=-T/2}^{T/2} \cos(3xt) y_{j+t}
\end{aligned}$$

If the instant t belongs to the boundaries of the interval, the summations are modified as

$$\sum_{t=-T/2+m}^{T/2} Re(e^{rit}) = \cos\big(\frac{rm}{2}\big)\frac{\sin(r\frac{(T-m+1)}{2})}{\sin(\frac{r}{2})}, \quad r \neq k\pi,$$

$$\sum_{t=-T/2+m}^{T/2} e^{rit} = T-m+1, \quad r = 2k\pi,$$

$$\sum_{t=-T/2+m}^{T/2} e^{rit} = -T+m-1, \quad r = (2k+1)\pi;$$

and

$$\sum_{t=-T/2}^{T/2-m} Re(e^{rit}) = \cos\big(\frac{-rm}{2}\big)\frac{\sin(r\frac{(T-m+1)}{2})}{\sin(\frac{r}{2})}, \quad r \neq k\pi,$$

$$\sum_{t=-T/2}^{T/2-m} e^{rit} = T - m + 1, \quad r = 2k\pi,$$

$$\sum_{t=-T/2}^{T/2-m} e^{rit} = -T + m - 1, \quad r = (2k+1)\pi.$$

Here x is a wrapping parameter, that must be non zero to guarantee the uniqueness of the solution. The problem of choosing x, k, T here is faced chiefly by means of simulations, and the simultaneous estimation of these parameters is still an open problem.

The smoother is local if it is used only for the prediction of y_j $(t = 0)$. Thus $\hat{y}_j = \sum_{k=0}^{\lambda} \hat{c}_k$. The obtained filter satisfies the *trigonometric reproducing property* up to the order of the system: if the real process y_t is a trigonometric polynomial, the system of equations is trivially satisfied.

More generally, the minimization of the functional $S(t) = \sum_{j=-T/2}^{T/2} \kappa_j |y_{t+j} - \sum_{k=0}^{\lambda} c_k e^{ix(t+j)k}|^2$ with respect to $c_0, \ldots, c_\lambda$, maintaining the κ_j fixed, can be putted in matrix form as follow. Let $t = 0$. Define $\gamma = [c_0, c_1, \ldots, c_\lambda]'$, $y = [y_{-T/2+t}, y_{-T/2+1+t}, \ldots, y_{T/2-1+t}, y_{T/2+t}]'$,

$$J = \begin{pmatrix} 1 & e^{-ixT/2} & e^{-2ixT/2} & \cdots & e^{-x\lambda T/2} \\ 1 & e^{-ix(T-1)/2} & e^{-2ix(T-1)/2} & \cdots & e^{-x\lambda(T-1)/2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{ix(T-1)/2} & e^{2ix(T-1)/2} & \cdots & e^{x\lambda(T-1)/2} \\ 1 & e^{ixT/2} & e^{2ixT/2} & \cdots & e^{x\lambda T/2} \end{pmatrix},$$

$$\Lambda = diag\{\kappa_{-T/2}, \kappa_{-T/2+1}, \ldots, \kappa_{T/2-1}, \kappa_{T/2}\}.$$

Then γ is the solution of $\bar{J}'\Lambda J\gamma = \bar{J}'\Lambda y$, where $\bar{J}$ is the complex conjugate of J, which implies $y = (\bar{J}'\Lambda J)^{-1}\bar{J}'\Lambda y$. Hence $\hat{y}_t = \sum_{k=0}^{\lambda} c_k = \sum_{k=0}^{\lambda} e_k(\bar{J}'\Lambda J)^{-1}\bar{J}'\Lambda y \equiv \psi' y$, for $e_k = [0, \ldots, 0, 1, 0, \ldots, 0]'$, (1 in the k-th position) and the filter ψ transmits without alteration the trigonometric trend of order λ.

The recourse to the use of complex number can be avoided by postulating a trend of the form $y_{j+t} = \sum_{k=-\lambda}^{\lambda} c_k e^{ixtk}$, with $c_{-k} = \bar{c}_k$. Thus the same trend assume the usual form $\sum_{k=0}^{\lambda}(a_k \cos(xkt) + b_k \sin(xtk))$, with $a_k = c_k + c_{-k} = 2\Re(c_k)$, $b_k = c_k - c_{-k} = -2i\Im(c_k)$.

Under this assumption, the system of equation becomes

$$\sum_{k=-\lambda}^{\lambda} c_k \sum_{t=-T/2}^{T/2} \kappa_t e^{ix(k-l)t} = \sum_{t=-T/2}^{T/2} e^{-ixlt} y_{t+j}, \quad l = -\lambda, \ldots, \lambda.$$

Let y_t be a local trigonometric trend written in exponential form:

$$y_{j+t} = \sum_{k=-\lambda}^{\lambda} c_k e^{itk},$$

so that $y_j = \sum_{k=-\lambda}^{\lambda} c_k$. Let ψ_t, $t = -T/2, \ldots, T/2$ a discrete system of weights satisfying

$$\sum_{t=-T/2}^{T/2} \psi_t e^{itk} = 1; \quad \forall k, \tag{3.3}$$

for each choice of c_t. Thus we have: the weights have the trigonometric reproducing property :

$$\hat{y}_{t+j} = \sum_{s=-T/2}^{T/2} \psi_s y_{j+t+s} = \sum_{s=-T/2}^{T/2} \psi_s \sum_{k=-\lambda}^{\lambda} c_k e^{i(t+s)k} =$$
$$= \sum_{k=-\lambda}^{\lambda} c_k e^{itk} \sum_{s=-T/2}^{T/2} \psi_s e^{iks} = \sum_{k=-\lambda}^{\lambda} c_k e^{itk}.$$

Taking $T = 2k$ the system (3.3) can be putted in matrix form: $A\psi = \mathbf{1}$, where $a_{lm} = e^{ixtk} = e^{ix(l-1-T/2)(m-1-\lambda)}$.

if $y_t = \mu_t + \varepsilon_t$, with $\varepsilon_t \sim IID(0, \sigma^2)$, then the variance of the estimator is $\sigma^2 \sum_{t=-T/2}^{T/2} |\psi_t|^2$.

Under the same hypothesis the local trigonometric estimator is unbiased. In fact if $y_j = \eta_j + \varepsilon_j$, the signal η_j is a deterministic trigonometric trend, $\eta_j = \sum_{k=0}^{\lambda} c_k e^{ijk}$, and ε_t is a white noise process, then

$$E[\hat{y}_j] = E[\sum_{t=-T/2}^{T/2} w_t y_{j+t}] = \sum_{t=-T/2}^{T/2} w_t E[\eta_{j+t}] + \sum_{t=-T/2}^{T/2} w_t E[\varepsilon_{j+t}]$$
$$\sum_{t=-T/2}^{T/2} w_t E[\sum_{k=0}^{\lambda} c_k e^{i(t+j)k}] = \sum_{k=0}^{\lambda} c_k e^{ijk} \sum_{t=-T/2}^{T/2} w_t e^{itk} = \eta_j.$$

The figures show the smoothing of a $AR(2)$ process $(L^2 - 0.6L + 0.08)y_t = \varepsilon_t$ where $\varepsilon_t \sim WN(0, \sigma^2)$, $\sigma^2 = 0.01$. It has been build the smoother for different choices of bandwidth and maximum frequency. The parameter x has been taken as small as possible. The presence of higher frequencies could lead to oversmoothing. The gain of the filter is about 1 in a neighborhood of zero frequency. Aliasing is also evident.

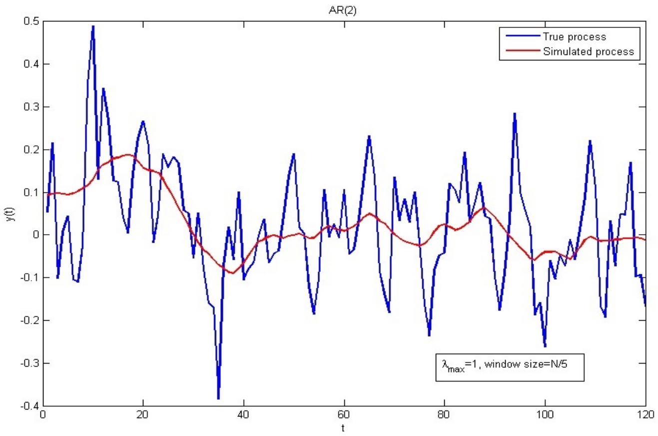

Figure 3.1: Smoothed AR(2) process, $\lambda = 1$, W=N/5

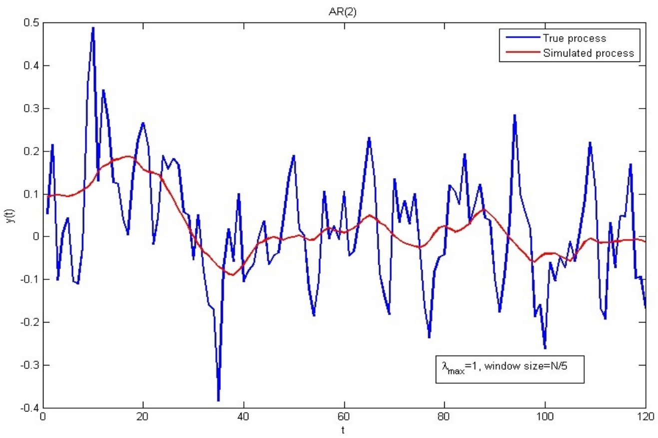

Figure 3.2: Smoothed AR(2) process, $\lambda = 1$, W=N/5

Table 3.1: CV for different λ and windows

	1	2	3	4
N/12	2.0143	1.8414	1.5757	1.7874
N/10	2.0891	1.9727	1.7182	1.7871
N/6	2.0173	2.0637	1.9164	2.1725
N/5	2.0292	2.0852	1.9433	2.1962

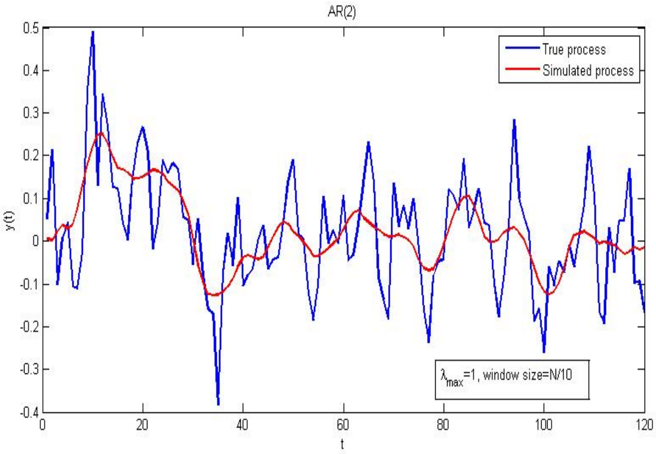

Figure 3.3: Smoothed AR(2) process, $\lambda = 1$, W=N/10

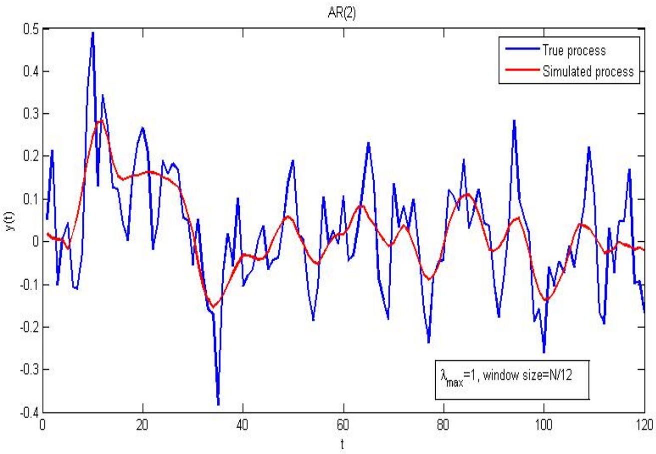

Figure 3.4: Smoothed AR(2) process, $\lambda = 1$, W=N/12

Table 3.2: VIF, t inner point for different λ and windows

	1	2	3	4
N/12	1.013	1.4535	1.5144	2.1628
N/10	1.0126	1.457	1.4881	2.1657
N/6	1.0119	1.4504	1.4806	2.1617
N/5	1.0117	1.453	1.4729	2.1447

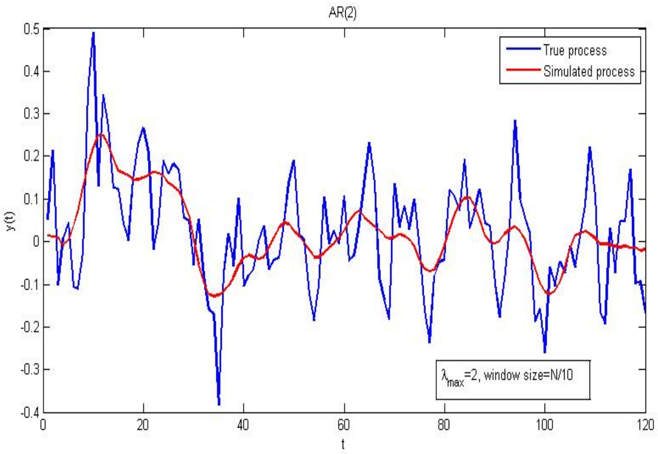

Figure 3.5: Smoothed AR(2) process, $\lambda = 2$, W=N/10

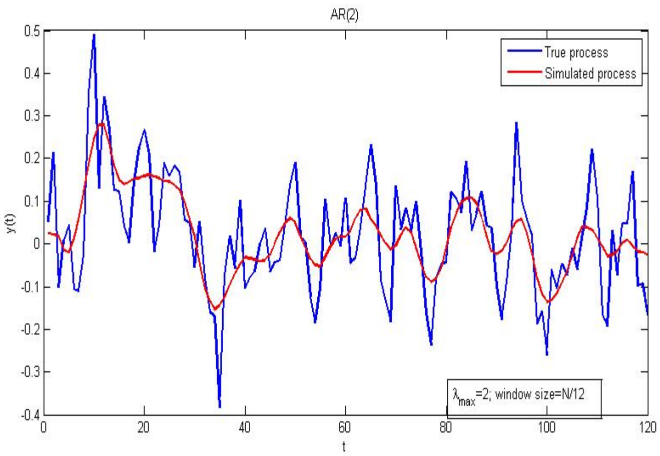

Figure 3.6: Smoothed AR(2) process, $\lambda = 2$, W=N/12

Table 3.3: RSS for different λ and windows

	1	2	3	4
N/12	1.1097	1.0762	0.6395	0.6206
N/10	1.2981	1.2547	0.7783	0.754
N/6	1.6112	1.5914	1.353	1.3374
N/5	1.6976	1.6652	1.487	1.4721

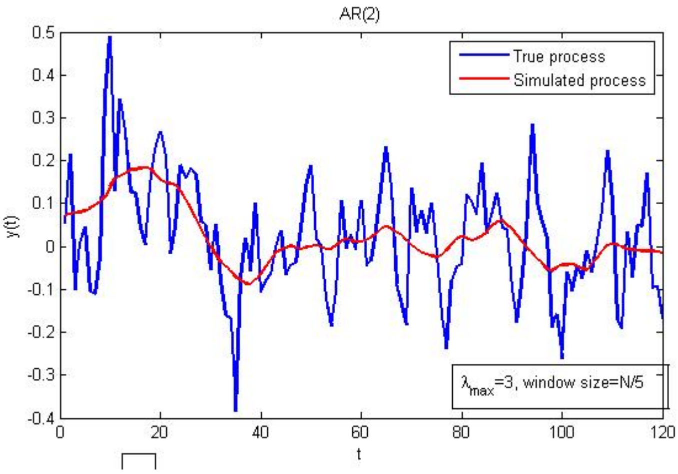

Figure 3.7: Smoothed AR(2) process, $\lambda = 3$, W=N/5

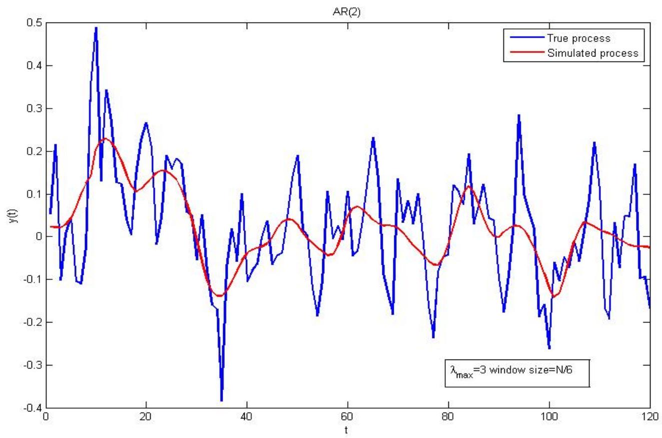

Figure 3.8: Smoothed AR(2) process, $\lambda = 3$, W=N/6

Table 3.4: $\hat{\sigma}^2$ for different λ and windows

	1	2	3	4
N/12	0.0101	0.0113	0.0059	0.0072
N/10	0.0116	0.0127	0.0071	0.0082
N/6	0.0140	0.0148	0.0119	0.0129
N/5	0.0147	0.0152	0.0130	0.0139

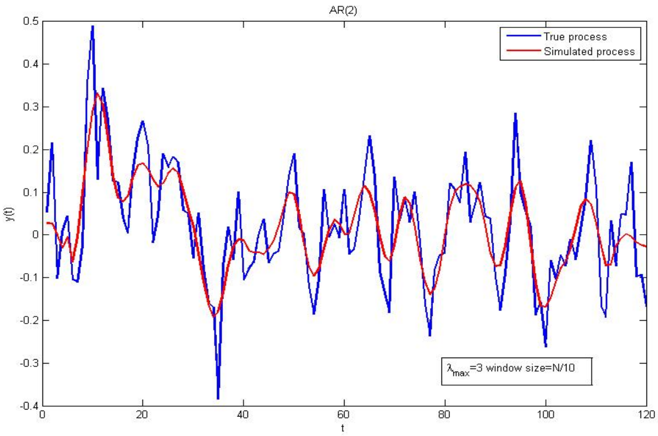

Figure 3.9: Smoothed AR(2) process, $\lambda = 3$, W=N/10

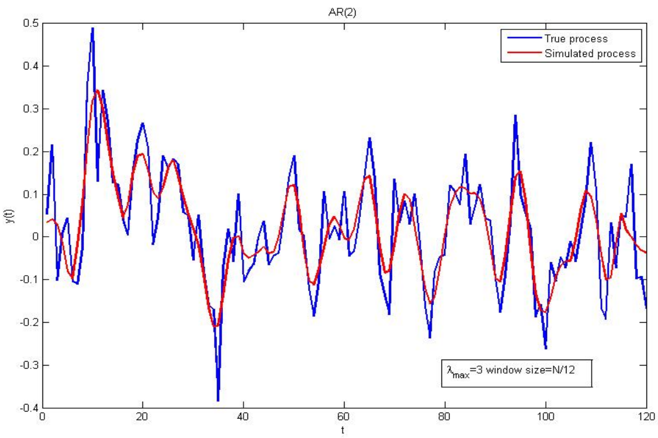

Figure 3.10: Smoothed AR(2) process, $\lambda = 3$, W=N/12

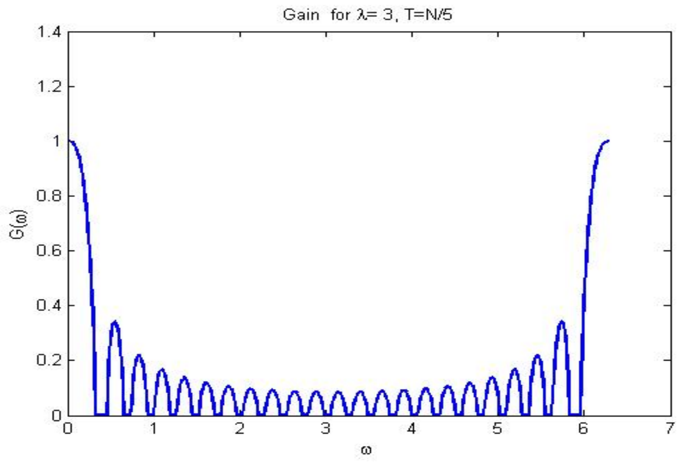

Figure 3.11: Gain of Smoothed AR(2) process, $\lambda = 3$, W=N/5

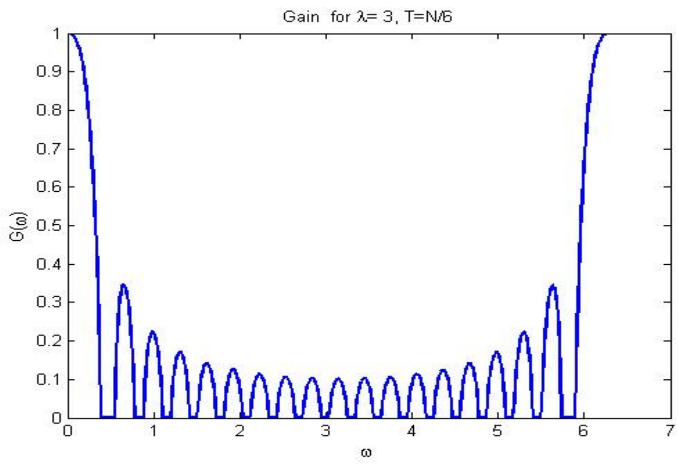

Figure 3.12: Gain of Smoothed AR(2) process, $\lambda = 3$, W=N/6

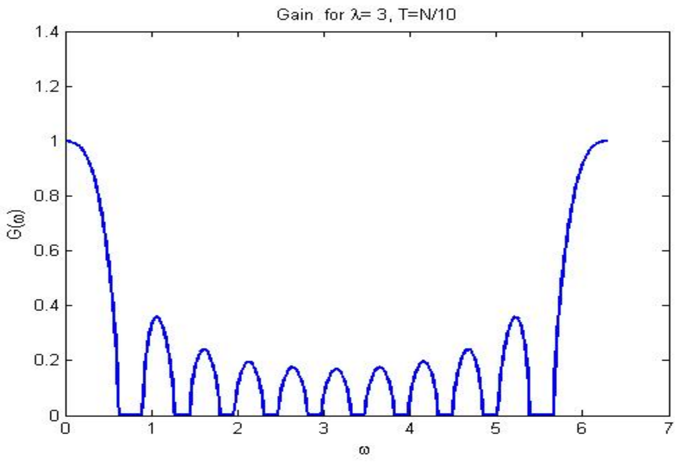

Figure 3.13: Gain of Smoothed AR(2) process, $\lambda = 3$, W=N/10

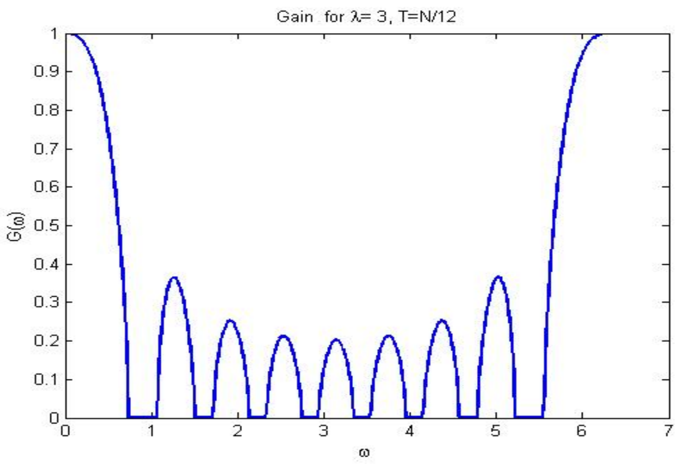

Figure 3.14: Gain of Smoothed AR(2) process, $\lambda = 3$, W=N/12

3.3.1 The choice of the sampling frequency

One problem faced in the previous section (3.3) was the choice of the parameter x which decides the frequency of sampling. Pollock (2012) sheds some light on this point. In fact, if the process under exam shows a precise range of frequencies, i.e. it is band limited, an optimal choice of the parameter x is possible. x can be thought as the maximum sampling frequency of a continuous time underlying process.

Macroeconomic data processes in fact are commonly thought as composed of components that fall within limited frequency bands.

The following relation holds for $x(t) \in L^2(\mathbb{R})$:

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{i\omega t}\xi(\omega)d\omega \leftrightarrow \xi(\omega) = \int_{-\infty}^{+\infty} e^{-i\omega t}x(t)dt.$$

By sampling $x(t)$ at integer time points, a sequence $\{x_t, t = 0, \pm1, \pm2, \dots\}$ is generated, of which the transform $\xi_S(\omega)$ is a 2π-periodic function.

$$x_t = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega t}\xi_S(\omega)d\omega \leftrightarrow \xi_S(\omega) = \sum_{k=-\infty}^{+\infty} x_k e^{-i\omega k}.$$

At the sampling point $x_t = x(t)$ yields

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{i\omega t}\xi(\omega)d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega t}\xi_S(\omega)d\omega,$$

which implies

$$\xi_S(\omega) = \sum_{k=-\infty}^{+\infty} \xi(\omega + 2k\pi).$$

The two functions will coincide if $\xi(\omega) = 0$ for $|\omega| \geq \pi$, otherwise $x_S(\omega)$ will be subject to a process of aliasing, since elements of the continuous function that are at frequencies in excess of π are confound with elements with frequencies less than π. Thus, the so called Nyquist frequency of π radians per period represents the limit of what is directly observable in sampled data.

If the condition is fulfilled $\xi(\omega) = 0$, then it should be possible to reconstitute the continuous function $x(t)$ from its sampled ordinates. This is the statement of the Nyquist-Shannon sampling theorem.

Theorem 1. *If f is a continuous periodic function of period T, which results square-integrable function in $[0, T]$, then it can be reconstructed by means of its sample taken at* *the frequency $f_c = 2/T$.*

If $x(t)$ is a periodic function, then

$$x(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\sum_{k=-\infty}^{+\infty} x_k e^{-i\omega k}\right\} e^{i\omega\pi} =$$

$$\frac{1}{2\pi}\sum_{k=-\infty}^{+\infty} x_k \int_{-\pi}^{\pi} e^{i\omega(t-k)}d\omega = \sum_{k=-\infty}^{+\infty} x_k \frac{\sin(\pi(t-k))}{\pi(t-k)}.$$

The sequence of sinc functions $\phi(t-k) = \frac{\sin(\pi(t-k))}{\pi(t-k)}$, $k \in \mathbb{Z}$ constitute an orthonormal basis for the set of all functions band-limited to the frequency interval $[-\pi, \pi]$. In fact, recall that the sinc function is the Fourier transform of

$$\chi(\omega) = \begin{cases} 1, & \omega \in (-\pi, \pi) \\ 1/2, & \omega = \pm\pi \\ 0, & \omega \notin [-\pi, \pi]. \end{cases}$$

Thus

$$\begin{aligned}
&\int_{\mathbb{R}} \phi(t)\phi(\tau - t)dt = \int_{\mathbb{R}} \phi(t)\Big\{\frac{1}{2\pi}\int_{\mathbb{R}} \chi(\omega)e^{i\omega(\tau-t)}d\omega\Big\}dt \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \chi(\omega)\Big\{\int_{\mathbb{R}} \phi(t)e^{-i\omega t}dt\Big\}e^{i\omega\tau}d\omega \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \chi(\omega)\chi(\omega)e^{i\omega t}d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega\tau}d\omega = \delta(t).
\end{aligned}$$

A continuous function $y(t)$ that is limited by the frequency value $\omega_c < \pi$ can be similarly expressed as

$$y(t) = \sum_{k=-\infty}^{+\infty} y_k \frac{\sin(\omega_c(t-k))}{\omega_c(t-k)}.$$

The sequence of functions $\phi_c(t-k) = \frac{\sin(\omega_c(t-k))}{\omega_c(t-k)}$, $k \in \mathbb{Z}$ constitute an orthonormal basis for the set of all functions band-limited to the frequency interval $[-\omega_c, \omega_c]$.

If moreover $\{h_t, t \in \mathbb{Z}\}$ is the sequence of ordinates sampled from the function $\phi_c(t)$ at unit intervals of time, then according to the Shannon-Nyquist theorem

$$\phi_c(t) = \sum_{k=-\infty}^{+\infty} h_k \phi(t-k).$$

As a consequence, the same $y(t)$ can be re-expressed as

$$y(t) = \sum_{j=-\infty}^{+\infty} y_j \Big\{ \sum_{k=-\infty}^{+\infty} h_k \phi(t-j-k) \Big\}.$$

If $x(t)$ is defined in $[0, T]$, then

$$x(t) = \sum_{k=-\infty}^{+\infty} \xi_k e^{i\omega_k t} \leftrightarrow \xi_k = \frac{1}{T} \int_0^T x(t) e^{-i\omega_k t} dt, \quad \omega_k = \frac{2\pi k}{T}.$$

By truncating the series expansion of $x(t)$ it is obtained a trigonometric polynomial

$$x(t) = \sum_{k=0}^{T-1} \xi_k e^{i\omega_k t} \leftrightarrow \xi_k = \frac{1}{T} \sum_{k=0}^{T-1} x_t e^{-i\omega_k t}.$$

Thus

$$x(t) = \sum_{j=0}^{T-1} \Big\{ \frac{1}{T} \sum_{k=0}^{T-1} x_t e^{-i\omega_k t} \Big\} e^{i\omega_j t} = \frac{1}{T} \sum_{j=0}^{T-1} x_k \sum_{j=0}^{T-1} e^{i\omega_j (t-k)}$$
$$= \sum_{j=0}^{T-1} x_k \frac{\sin(\omega_1 (t-k) T/2)}{\sin(\omega_1 (t-k)/2)} e^{i\omega_1 (t-k)(T-1)/2}.$$

where the Dirichlet kernel $\frac{\sin(\omega_1(t-k)T/2)}{\sin(\omega_1(t-k)/2)}$ constitutes the discrete version of the sinc function.

The forcing function $\varepsilon(t)$, i.e. the error term, of the underlying continuous process can be estimated by means of the sampled error term as follows:

$$\varepsilon(t) = \sum_{k=-\infty}^{+\infty} \varepsilon_k \phi(t-k)$$

Similarly, the covariance function of the continuous process and the covariance of the sampled process are linked by the relation:

$$\gamma_\varepsilon(\tau) = E[\varepsilon(t)\varepsilon(t+\tau)] = \sum_{k=-\infty}^{+\infty} E[\varepsilon_k \varepsilon_t] \phi(t+\tau-k) = \sigma_\varepsilon^2 \phi(\tau).$$

A discrete ARMA process is obtained by sampling the continuous analogous ARMA process.

$$\gamma(\tau) = \sum_{k=-\infty}^{+\infty} \gamma_i \phi(\tau - i).$$

The prediction of a time limited stochastic process can be enhanced by resampling the original data sequence to a rate $r = \omega_c$(Pollock). The frequency ω_c can be evaluated by visual inspection of the spectral density of the process, or, by inspection of the periodogram (see 4.3.1) when the spectral density is not available.

The consequence of applying an unrestricted estimator to data that are strictly band -limited will be to create an estimated autoregressive polynomial in which the complex roots approach the perimeter of the unit circle of the complex plane, exhibiting an artificial nonstationarity.

3.3.2 Test of hypothesis for trigonometric models

Rosenblatt and Grenander (1957) give some insight relative to the comparison of two different trigonometric regressions. Suppose that we want to compare the two statistical models

$$H_j = y_t = x_t + m_t^{(j)}, \quad j = 0, 1,$$

where $x_t \sim N(0, \sigma^2)$ with spectral distribution $F(\omega)$. Assume that the two vector $m^{(0)} = (m_1^{(0)}, \ldots, m_n^{(0)})$ and $m^{(1)} = (m_1^{(1)}, \ldots, m_n^{(1)})$ have real components. After having observed $y_1, \ldots, y_n$ we want to test H_0 against H_1. Assume that the covariance matrix R of the disturbance x_t is not singular. Under H_j, $j = 0, 1$, the vector x_t has pdf

$$f_j(y_1, y_2, \ldots, y_n) = \frac{1}{(2\pi)^{n/2}} \exp\left\{\frac{1}{2}(y - m^{(j)})' R^{-1} (y - m^{(j)})\right\}, \quad j = 0, 1.$$

The most powerful test of H_0 against H_1 has the critical region

$$W = \{y' R^{-1} (m^{(1)} - m^{(0)}) > K\}$$

Assume that the $m^{(j)}$ admits Fourier expansion

$$m_t^{(j)} = \int_{-\pi}^{\pi} e^{it\omega} d\varphi_j(\omega), \quad j = 0, 1$$

$\varphi_j(\omega)$ bounded variation function. Let the expected value be

$$E_j y' R^{-1} (m^{(1)} - m^{(0)}) = \mu_j, \quad j = 0, 1,$$

and the variance

$$D^2[y'R^{-1}(m^{(1)} - m^{(0)})] = v,$$

which does not depend on j. Then there exists a consistent test of H_0 against H_1 if and only if

$$\tau \equiv \frac{(\mu_1 - \mu_0)^2}{v} \to^n \infty.$$

or equivalently, if and only if

$$\int_{-\pi}^{\pi} \frac{|d(\varphi_1(\omega) - \varphi_0(\omega))|^2}{dF(\omega)} = +\infty.$$

(Hellinger integral). In other words, the spectral densuty of the random error x_t have to be smaller than the squared Fourier transform of Δm to assure the consistency of the test. The presence of a discontinuity in the spectrum of x_t makes it harder to discriminating between H_0 and $H_1$and to find a consistent test.

If the disturbance x_t are normally distributed and the covariance matrix R of $x_1, \ldots, x_n$ are fixed and known, then $\varphi' R^{-1} y$ is a minimal sufficient statistic for the class of distribution of m . The linear estimate

$$c = (\varphi' R^{-1} \varphi)^{-1} \varphi' R^{-1} y$$

is an unbiased estimate of $\gamma(\omega) \equiv \varphi_1(\omega) - \varphi_0(\omega)$.

Suppose $c_n = \sum_{t=1}^n a_t^{(n)} y_t$ is a sequence of consistent estimates of $\gamma(\omega)$ in mean square sense.

$$c_n = \frac{\sum_{t=1}^n \bar{\varphi}_t y_t}{\sum_{t=1}^n |\bar{\varphi}_t|^2}.$$

c_n is an asymptotic unbiased estimates if and only if

$$\lim_{n \to +\infty} \sum_{t=1}^n |\varphi_t|^2 = +\infty.$$

In the trigonometric regression

$$y_t = \gamma_1 e^{it\lambda_1} + \gamma_2 e^{it\lambda_2} + \cdots + \gamma_p e^{it\lambda_p} + x_t$$

if the frequencies λ_1, λ_2, …, λ_p are distinct, the least squares estimates of γ_1, γ_2, …, γ_p are asymptotically efficient.

3.3.3 Trigonometric filter as an approximation of the ideal band-pass filter

The global Trigonometric filter $\hat{y}_t = B_t^*(L)x_t$ arises from a Trigonometric Regression (1.3) without mean μ by rearranging the terms of the summations:

$$B_t^*(L)x_t = \sum_{l=t-T}^{t-1} \Big\{\frac{2}{T}\sum_{j\in J}\cos(\omega_j l)\Big\}x_{t-j}, \quad \frac{T}{2} \notin J,$$

$$B_t^*(L)x_t = \sum_{l=t-T}^{t-1} \{\frac{2}{T}\sum_{j\in J, j\neq \frac{T}{2}} \cos(\omega_j l) + \frac{1}{T}\cos(\pi(t-l))\cos(\pi t)\}x_{t-j}\}, \quad \frac{T}{2} \in J,$$

$$t = 1,\ldots,T, \quad \omega_j = \frac{2\pi}{T}j, \quad J \subseteq \{1,\ldots,T\}.$$

has a unit roots since $B_t^*(1) = 0$, and it has at least two unit roots if it is symmetric. If compared with the Random Walk filter, it has al lower correlation function in all the frequencies, has a higher variance, especially in the lowest frequencies , and it shows a substantial departure from covariance-stationarity (see CF, 2003).

If the frequencies selected are the first λ, that is $J = \{1, 2, \ldots, \lambda\}$, the weights of the trigonometric filter are:

$$B_{lt}^* = \frac{2}{T}\sum_{j=1}^{\lambda}\cos(\omega_j l) = \frac{2}{T}\sum_{j=1}^{\lambda}\cos(\omega_l j) = \frac{2}{T}\Big[\frac{\sin(\lambda+1/2)\omega_l}{2\sin(\omega_l/2)} - \frac{1}{2}\Big] =$$
$$= \frac{\sin(\lambda+1/2)\omega_l - \sin(\omega_l/2)}{T\sin(\omega_l/2)} = \frac{\sin[(2\pi\lambda/T + \pi/T)l] - \sin(\pi l/T)}{T\sin(\pi l/T)}.$$

The filter is linear and time invariant. The weight B_{lt}^* is zero if $\lambda l/T$ is an integer.

If $b = (2\lambda+1)\pi/T$ and $a = \pi/T$, since $T\sin(\pi l/T) = \pi l + o((l/T)^2)$, and if B_l is the l-th weight of the ideal band-pass filter for the band $[a, b]$, one obtains $B_l^* = B_l + o((l/T)^2)$. If $\lambda(T)/T \to c$ for $T \to \infty$, this weight converges to the weight B_l of an ideal band-pass with $a = 0$ and $b = 2\pi c$.

If the frequencies selected are those between λ_1 and λ_2, then

$$B_{lt}^* = \frac{2}{T}\sum_{j=\lambda_1+1}^{\lambda_2}\cos(\omega_j l) = \frac{2}{T}\sum_{j=\lambda_1+1}^{\lambda_2}\cos(\omega_l j) = \frac{\sin(\lambda_2+1/2)\omega_l - \sin(\lambda_1+1/2)\omega_l}{T\sin(\omega_l/2)}$$
$$= \frac{\sin[(2\pi\lambda_2/T + \pi/T)l] - \sin[(2\pi\lambda_1/T + \pi/T)l]}{T\sin(\pi l/T)}.$$

and the weight approximates the B_l of an ideal band-pass with $a = (2\lambda_1 + 1)\pi/T$ and $b = (2\lambda_2 + 1)\pi/T$. Moreover, if $\lambda_1(T)/T \to c_1$ and $\lambda_2(T)/T \to c_2$ for $T \to \infty$, then the weight B_l^* converges to the weight B_l of an ideal band-pass with $a = 2\pi c_1$ and $b = 2\pi c_2$.

Chapter 4

Convergence properties of the proposed models

In this chapter it will be discussed the rate of convergence of global and local trigonometric estimators. Optimality criteria such as Mean Squared Error, Integrated Mean Squared Error and Pointwise Mean Squared Error are proposed, and there are given the condition on the order of the model that have to be satisfied in order to guarantee the consistency of the estimators.

Then it is discussed the problem of choosing the order of the model by means of data driven methods. Selection criteria such as the Generalized Cross Validation, AIC-like, BIC and Mallows C_p are examined, showing their asymptotic equivalence and comparing their performances.

4.1 General convergence properties of local models

Polynomial regression can be shown to produce an estimator of μ that attains the theoretical optimal rate of convergence for mean squared error in certain sets (Rafajlowicz, 1987; Cox, 1988).

The rate of convergency of a trigonometric estimator can be not so satisfactory. In fact the mean squared error convergence rates for trigonometric series estimators are as slow as $n^{-\frac{1}{2}}$ globally, or $n^{-\frac{2}{3}}$ locally for a twice differentiable, non periodic function (Hall, 1981, 1983) rather than the optimal $n^{-\frac{4}{5}}$ rate obtained both by kernel and cubic smoothing spline (Walter and Blum (1979)). It is due to the fact that a trigonometric polynomial is always periodical whereas the unknown function μ can be aperiodical.

It can be shown that the boundary behavior of a trigonometric series estimator dominates its squared error.

The result is that when a data driven method based on a mean squared error estimate such as cross-validation is used to choose the number of trigonometric functions in the regression, the result is often as estimator involving too many terms that undersmooths and exhibits anomalous wiggles.

Eubank and Speckman (1990) try to improve the rate of convergence of the trigonometric estimator by adding to it a polynomial term of order d playing the role of a deterministic trend. For a similar model the generalized cross-validation for selecting λ is defined as

$$GCV(\lambda) = nRSS(\lambda)/(n - 2\lambda - d - 1)^2$$

and the unbiased risk criteria is defined as

$$\hat{R_\lambda} = n^{-1}RSS(\lambda) + 2\sigma^2(2\lambda + d + 1)/n \tag{4.1}$$

where

$$RSS(\lambda) = \sum_{i=1}^{n} (y_i - m_\lambda(t_i))^2.$$

The t statistic can be employed to aid the detection of terms in the estimator which do not contribute to the overall fit.

Define the mean squared prediction error of μ_λ as

$$R_n(\lambda) = n^{-1} \sum_{i=1}^{n} E\{\mu(t_i) - \mu_\lambda(t_i)\}^2$$

and assume that the t_i are distributed as a sample from a distribution function W and continuous positive density w on $[0, 2\pi]$. If $0 \le t_1 \le \cdots \le t_n$, let W_n be the corresponding empirical distribution function $W_n(t) = k/n$ for $t_k \le t \le t_{k+1}$ $(k = 1, \cdots, n)$ and let $\delta_n = sup_t|W(t) - W(t_n)|$.

It can be proved that if μ has $d-1$ absolutely continuous derivatives with $\mu^{(d)}$ square integrable, then

$$R_n(\lambda) = O(\lambda^{-2d}) + \sigma^2(2\lambda + d + 1)/n + O(\delta_n \lambda^{-2d+1}).$$

Thus, taking $\lambda \propto n^{\frac{1}{2d+1}}$ we obtain $n^{-\frac{2d}{2d+1}}$ as a rate of decay for $R_n(\lambda)$. Stone (1982) and Speckman (1985) have shown that $n^{-\frac{2d}{2d+1}}$ is the best uniform rate for linear estimator over function with the same smoothness properties as μ.

The orthogonality of the trigonometric functions allows one to avoid problems of collinearity , so the proposed method is also a practical alternative to the use of orthogonal polynomials. If a polynomial trend is added, the orthogonality is lost. The solution could be again detrending the series of the observations, and then fitting via trigonometric regression.

4.2 Convergence of trigonometric estimators in L^2

Some properties of the global Trigonometric Regression Estimation has been highlighted by Popinski (1999).

Consider the model

$$y_i = f(t_{i,n}) + \eta_i, \quad i = 1, \dots, n,$$

where the function $f(t) = \tau(t) + c(t)$ is the deterministic sum of trend and cycle, and $\eta(t)$ is the random component. For simplicity, we take the equidistant observations $t_{i,n} = \frac{2\pi(i-1)}{n}$.

It is well known that each function $f \in L^2[0, 2\pi]$ has the representation

$$f(t) = \sum_{k=0}^{\infty} c_k e_k(t),$$

for

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(s) e_k(s) ds$$

being the Fourier transform of f, where the functions:

$$e_0(x) = 1, \quad e_{2l-1}(x) = \sqrt{2} \sin(lx), \quad e_{2l} = \sqrt{2} \cos(lx), \quad l = 1, 2, \cdots ,$$

constitute a complete normalized orthogonal system in the space $L^2[0, 2\pi]$.

In the Fourier decomposition of the function f, the highest components play the role of the cycle, while the lowest components play the role of the trend. Both the components are deterministic.

If one minimizes the function

$$\sum_{i=1}^{n} \left[y_i - \sum_{k=1}^{\lambda} C_k \cos(kt_{i,n}) + S_k \sin(kt_{i,n}) \right]^2 = \sum_{i=1}^{n} \left[y_i - \sum_{k=1}^{2\lambda} c_k e_k(t_{i,n}) \right]^2$$

one derives the estimates [1]

$$\widehat{c}_{k,n} = \frac{1}{n}\sum_{i=1}^{n} y_i e_k(t_{i,n}), \quad k = 1, 2, \cdots, n,$$

that are the mean squared estimates of the Fourier coefficient c_k.Notice that the $\hat{c}_{k,n}$ are the discrete Fourier transform of f, that is $\hat{c}_{k,n} = J_n(\omega_k)$.

If the regression function f is continuous, the estimators $\widehat{c_k}$ of the Fourier coefficients c_k are asymptotically unbiased and consistent in the mean-square sense.

In the definition of an optimality criterium, it should be taken into account both the error committed by truncating the infinite sum representing f, both the random error η, and this requirement can be fulfilled in different ways. Following Popinski (1999) we explore three different criteria for estimating the prediction error: the mean-square prediction error

$$R_n(\lambda) = \frac{1}{n}\sum_{i=1}^{n} E(f(x_{in}) - \hat{f}_{\lambda(n)})^2 \equiv \frac{1}{n}\sum_{i=1}^{n}\int_{\Omega}(y_i - \hat{f}_{\lambda(n)} - \eta_i(\omega))^2 d\omega,$$

the integrated mean-square error

$$E||f - \hat{f}_{\lambda(n)}||^2 \equiv \int_{\Omega}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}(y_i - \hat{f}_{\lambda(n)}(x) - \eta_i(\omega))^2 dx\right\}d\omega,$$

and the pointwise mean-square error

$$E(f(x) - \hat{f}_{\lambda(n)})^2 \equiv \int_{\Omega}(y_i - \hat{f}_{\lambda(n)}(x_i) - \eta_i)^2 d\omega$$

of the estimator

$$\hat{f}_{\lambda(n)} = \sum_{k=0}^{\lambda(n)} \hat{c}_k e_k(x), \quad \text{for} \quad f \in C[0, 2\pi].$$

The choice of the function f in the model arises from the theory under consideration (economic, biological, astronomical...) such approach is not completely data driven, since in a complete data driven approach it would be enough to minimize a functional of the random error $\eta(t)$, but this procedure can be useful if we have a strong believe in the form of the function f, and f is difficult to be computed. Another approach could be choosing f with a global trigonometric regression, and afterwards, determining the $\hat{f}_\lambda(n)$ by a local fitting. Moreover, minimizing such a composite criterium allows to work with

[1] Notice that if the observations are equispaced, then $\frac{1}{n}\sum_{i=1}^{n} e_k(t_{i,n})e_l(t_{i,n}) = \delta_{k,l}$.

a series not filtered in advanced.

The paper proves that if f is a continuous 2π-periodic and not a trigonometric polynomial, the estimator $\hat{f}_{\lambda(n)}$ is consistent in the sense of the mean-square prediction error if and only if the sequence of numbers $\lambda(n)$, $n = 1, 2, \cdots$ satisfies

$$lim_{n\to\infty}\lambda(n) = \infty, \quad lim_{n\to\infty}\frac{\lambda(n)}{n} = 0.$$

As a corollary it follows that if the function f satisfies the Lipschitz condition with exponent $0 < \alpha \leq 1$ and if the sequence of natural numbers $\lambda(n)$ satisfies $\lambda(n) \sim n^{\frac{1}{1+2\alpha}}$, then

$$\frac{1}{n}\sum_{i=1}^{n} E(f(x_{in}) - \hat{f}_{\lambda(n)}(x_{in}))^2 = O\big(n^{-\frac{2\alpha}{1+2\alpha}}\big).$$

This result complements the one obtained by Eubank and Speckman (1991) for a more general fixed point design under more restrictive conditions on the smoothness of the regression function.

In the case of the integrated meas-square error $E||f - f_\lambda\hat{(}n)||$ Popinski (1999) finds that, for f absolutely continuous function, if

$$lim_{n\to\infty}\lambda(n) = \infty, \quad lim_{n\to\infty}\frac{\lambda(n)^{3/2}}{n} = 0$$

then the estimator $\hat{f}_{\lambda(n)}$ of the absolutely continuous function f is consistent in the sense of the integrated mean-square error.

With respect to the rate of convergence, if f is absolutely continuous, and $\lambda(n) \sim n^{1/2}$, then $E||f - \hat{f}_{\lambda(n)}||_2^2 = O(n^{-1/2})$. In general with the same argument it can be shown that if f is absolutely continuous, and $\lambda(n) \sim n^\alpha$ for some positive $\alpha \leq \frac{2}{3}$, then $E||f - \hat{f}_{\lambda(n)}||_2^2 = O(n^{-\alpha})$ for $0 < \alpha \leq 1/2$ and $E||f - \hat{f}_{\lambda(n)}||_2^2 = O(n^{3\alpha-2})$ for $1/2 < \alpha \leq 2/3$.

Finally, for point-wise mean-square error, the paper demonstrates that under the more restrictive hypothesis

$$lim_{n\to\infty}\lambda(n) = \infty, \quad lim_{n\to\infty}\frac{\lambda(n)^2}{n} = 0,$$

if f is an absolutely continuous function, then for any $\delta \geq 0$ the estimator $\hat{f}_{\lambda(n)}$ is uniformly consistent in the sense of the pointwise mean-square error in the interval $(\delta, 2\pi - \delta)$. If moreover the function f is 2π-periodic, then the pointwise mean-square estimator $\hat{f}_{\lambda(n)}$ converges uniformly on $[0, 2\pi]$.

As a criterion of order selection, Popinski suggests the minimization of the Mallows's C_p:

$$C(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}_\lambda(t_{i,n}))^2 + \frac{2\lambda\hat{\sigma}^2}{n} \tag{4.2}$$

where $\hat{\sigma}$ is a consistent estimator of σ_η^2. This criterion does not require the knowledge of the best deterministic trend f, and thus is completely data-driven.

The Mallows's C_p furnishes a consistent estimate of the sum of squared error of the regression obtained retaining only the first n regressors.

If $\hat{\lambda}(n)$ is the minimizer of $C(\lambda)$, Posinski shows that for $\mu(t)$ absolutely continuous and not a trigonometric polynomial of finite order, if there exits a sequence of positive number ε_k such that $(k+1)\varepsilon_k$ is not increasing, $|c_k| \leq \varepsilon_k$, $\sum_{k=0}^{\infty}\varepsilon_k < \infty$, if $\mu_4 = \sup E\eta_i^4 < \infty$ and $\hat{\sigma}^2 \to \sigma_\eta^2$ as $n \to \infty$, then

$$\int_{-\pi}^{\pi}(f - \hat{f}_{\hat{\lambda}(n)})^2 = O_p(n^{-1/2}).$$

If moreover the function $f = \mu$ is 2π-periodic and satisfies the Lipschitz condition for $0 < \alpha \leq 1$, then the discrete version of the loss function $d_n(\lambda)$ satisfies

$$d_n(\lambda) \equiv \frac{1}{n}\sum_{i=1}^{n}(f(t_{i,n}) - \hat{f}_\lambda(t_{i,n}))^2 = O_p(n^{-\frac{2\alpha}{1+2\alpha}}).$$

4.3 Estimation of the order of the model based on the Periodogram

4.3.1 Definition of Periodogram

A basic tool to do inferences on the frequency-domain properties of a time series is the periodogram, which helps in the detection of specific frequency contained in a time series and in the choosing of the order of the model.

Definition 6. The *periodogram* $I_y(\omega)$, $\omega \in [-\pi, \pi]$ is defined as:

$$I_y(\omega_j) = \frac{1}{T}\Big|\sum_{t=1}^{T} y(t)e^{-i\omega_j t}\Big|^2, \quad \text{for} \quad \omega_j = \frac{2\pi j}{T}$$

$$I_y(\omega) = I_y(\omega_j) \quad \text{if} \quad \omega \in]\omega_j - \pi/T, \omega_j] \cap [0, \pi]$$

$$I_y(\omega) = I_y(-\omega) \quad \text{if} \quad \omega \in [-\pi, 0].$$

where $y = (y_1, \ldots, y_T)$ is a sample of the (complex) time series Y_t.

Taken the finite Fourier transform

$$J_T(\omega) = \frac{1}{T}\sum_{t=0}^{T-1} y(t)e^{-i\omega t}$$

$I_T(\omega_k)$ and $J_T(\omega_k)$ are linked by $I_T(\omega_k) = TJ_T(\omega_k)^2$.

The periodogram decomposes $|y|^2$ into a sum of components associated with the Fourier frequencies ω_j:

$$|x|^2 = \sum_{j \in F_T} I_T(\omega_j), \quad F_T = \{-[(T-1)/2], \ldots, [T/2]\}.$$

It is easily seen that the periodogram is closely related to the sample autocovariance function $\hat{\gamma}(k)$, $|k| < T$: if ω_j is any non-zero Fourier frequency, and the sample autocovariance function is defined as

$$\hat{\gamma}(k) = \frac{1}{T}\sum_{t=1}^{T-k} (y(t+k) - m)(\bar{y}(t) - \bar{m}), \quad k \geq 0,$$

$\hat{y}(t) = \bar{\hat{y}}(-k)$, for $k < 0$, where

$$m = \frac{1}{T}\sum_{t=1}^{T} y(t),$$

then

$$I_T(\omega_j) = \sum_{|k|<T} \hat{\gamma}(k)e^{-ik\omega_j}.$$

Recall the expression of the spectral density of a stationary process :

$$f(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty} \gamma(k)e^{-ik\omega}$$

with $\sum_{k\in\mathbb{Z}} |\gamma(k)| < +\infty$.

If the time series consists in a periodic trend plus a Gaussian error the periodogram $I_T(\omega_j)$ at any set of frequencies $0 < \omega_1 < \cdots < \omega_m < \pi$ are asymptotically independent exponential random variables with mean $2\pi f(\omega_j)$ and variance $(2\pi)^2 f^2(\omega) + O(T^{(-1/2)}$ for $\omega \in (0, \pi))$, where $f(\omega)$ is the spectral density of $y(t)$. Consequently, the periodogram

I_T is not a consistent estimator of $2\pi f$.

The presence of a trigonometric trend can be tested with a test for hidden periodicity based on the periodogram by means of test such as Fisher's test (Fisher, 1929). The visual inspection of the periodogram, or of a suitable smoothed version of it helps in detecting hidden periodicities, since these correspond to lines in the spectrum, and hence to local maxima of $f(\omega)$.

Since for large T the periodogram ordinates are approximatively uncorrelated with variance changing only slightly over small frequency intervals, consistent estimator can be constructed by averaging the periodogram ordinates in a small neighboring of ω. If

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{|k|<m_T} W_T(k) I_T(\omega_{j+k}),$$

under suitable hypothesis for the weight function W_T and the integer m_T $\hat{f}(\omega)$ converges in mean square to f uniformly on $[-\pi, \pi]$.

4.3.2 A comparison among estimators presented in literature

The estimation of the periodogram is needed in some information criteria used to select the order of a global or local Trigonometric Regression. A possible strategy when smoothing a series by means of a LTR is to decide in advance a low order for the trigonometric polynomial estimating the signal in each point t. An alternative strategy is to estimate the order of the regression, and after constructing the LRT. This method is not exonerated by criticism, since it alters the level of statistic tests; moreover the convergence to the "true" model is not uniform(Pötscher and Leeb, 2005).

Quinn (1989) suggests an AIC-like estimator for the number of terms in a trigonometric regression. In details, let the model consist only in cosine terms

$$y(t) = \mu + \sum_{j=1}^{\lambda} \rho_j \cos(\omega_j t + \phi_j) + \varepsilon_t \tag{4.3}$$

let the ϕ_j and ρ_j be real numbers, with $\rho_1 \geq \rho_2 \cdots \geq \rho_\lambda > 0$, and let the ω_j be the Fourier frequencies $2\pi j/T$, $1 \leq j \leq [(T-1)/2]$, and further assume (a) that ε_t is stationary and ergodic, with $E(\varepsilon_t^2) = \sigma^2 < \infty$ and $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_t$ denotes the σ-field generated by $\{\varepsilon_s, s \leq t\}$, or the more restrictive condition (b) ε_t i.i.d. sequence with $E(\varepsilon_t^6) < \infty$, and

$$\sup_{|s|>s_0>0} |\psi(s)| = \beta(s_0) < 1$$

where $\psi(s)$ is the characteristic function of ε_t.

If we fit this model by the least square method with $\omega_1, \ldots, \omega_\lambda$ known, the residual sum after fitting is

$$\sum_{t=0}^{T-1} (y(t) - \bar{y})^2 - \sum_{j=1}^{\lambda} I_y(\omega_j)$$

where

$$\bar{y} = \frac{1}{T} \sum_{t=0}^{T-1} y(t)$$

and

$$I_y(\omega) = \frac{2}{T} \Big| \sum_{t=0}^{T-1} y(t) e^{i\omega t} \Big|^2.$$

If, otherwise, the ω_j are unknown, the residual mean square is

$$\hat{\sigma}_\lambda^2 = \frac{1}{T} \{ \sum_{t=0}^{T-1} (y(t) - \bar{y})^2 - S_\lambda \}$$

where S_λ is the sum of the largest λ ordinates of the periodogram $I_y(\omega)$, for ω Fourier frequencies. Thus Quinn (1989) constructs the information criterion

$$\phi_g(\lambda) = T \log(\hat{\sigma}_\lambda^2) + 2\lambda g(T), \quad (\lambda = 0, 1, \ldots)$$

where $g(T)$ is a suitable function satisfying $T^{-1} g(T) \to 0$ as $T \to \infty$ The number of sinusoids is estimated by the first local minimum of $\phi_g(\lambda)$, in other words $\hat{\lambda}$ is the first integer for which $\phi_g(\lambda) < \phi_g(\lambda + 1)$. This $\hat{\lambda}$ is a consistent estimator of λ under suitable conditions on $g(T)$.

The proof of the consistency of the estimator $\hat{\lambda}$ exploits the boundedness of the periodogram as stated in An et al (1983), and uses the fact that

$$I_y(\omega) = \begin{cases} I_\varepsilon(\omega) & \text{if} \quad \omega \neq \omega_j, \quad j = 1, \ldots, \lambda, \\ |I_\varepsilon(\omega)^{\frac{1}{2}} + (\frac{T}{2})^{1/2} \rho_j e^{-i\phi_j}|^2 & \text{if} \quad \omega = \omega_j, \quad j = 1, \ldots, \lambda. \end{cases}$$

It follows that $T^{-1} I_y(\omega_j)$ converges a.s. to $\rho_j^2/2$, for $j = 1, \ldots, \lambda$, and $\hat{\sigma}_\lambda^2$ converges a.s. to $\sigma^2 + \sum_{j=\lambda+1}^{\lambda_0} \rho_j^2/2$ if $\lambda < \lambda_0$ and to σ^2 for $\lambda = \lambda_0$. Thus for $\lambda < \lambda_0$, $T^{-1}\{\phi_g(\lambda) - \phi_g(\lambda_0)\}$ converges a.s to $\log(1 - \sum_{\lambda+1}^{\lambda_0} \rho_j^2/(2\sigma^2))$, which is strictly positive and decreasing with λ, while $T^{-1}\{\phi_g(\lambda_0 + 1) - \phi_g(\lambda_0)\}$ is proved to be a.s. greater than zero for $T \to \infty$, and finally $\hat{\lambda}$ converges to λ_0.

Wang (1993) discusses the same model (4.3), removing the hypothesis that the ω_j are the Fourier frequencies. In fact one can take the frequencies $\omega_1, \ldots, \omega_\lambda$ as those frequencies maximizing the periodogram on $(-\pi, \pi]$ Taken the finite Fourier transform $J_T(\omega)$, the mean squared error of the fitted model of order λ can be written as

$$\sigma_\lambda^2(T) = \frac{1}{T} \sum_{t=0}^{T-1} \left| y(t) - \sum_{j=1}^{\lambda} J_T(\hat{\omega_j}) e^{i\hat{\omega_j}t} \right|^2.$$

Then the order of the regression λ is estimated as the minimum of the Best Information Criterion $BIC_{T,b}(\lambda)$ defined as

$$BIC_{T,b}(\lambda) = T \log \sigma_\lambda^2 + b_T \lambda,$$

where b_T is a sequence of number satisfying $T^{-1} b_T \to 0$ for $T \to \infty$. The $\hat{\lambda}$ minimizing such $BIC_{T,b}(\lambda)$ obviously coincides with the estimator of Quinn (1989) if the ω_j are the Fourier frequencies.

If ε_t has continuous spectral density, and if it is ergodic and under suitable assumption on $\max_\omega I_\varepsilon(\omega)/(\log T \max f_\varepsilon(\omega))$, $\hat{\lambda}$ is a consistent estimator of λ.

The problems of estimating the frequencies and of estimating the order of the regression are often treated together.

The frequencies of the trigonometric model can be found by Maximum likelyhood estimation or maximizing the periodogram.

Hannan (1973) examines a harmonic model

$$y(t) = \rho_j \cos(\omega_j t + \phi_j) + \varepsilon_t$$

and proves that, under the hypothesis of Gaussian error, the estimator $\hat{\omega}$ for the only frequency ω_0 is such that $T^{3/2}(\hat{\omega} - \omega_0)$ is asymptotically normal with variance $4\pi f(\omega_0)/\rho_0^2$, for $f(\omega)$ spectral density.

The results can be extended to a trigonometric regression with non-zero mean and λ sinusoids, by first estimating μ by $\bar{y}$ and then estimating $0 < \omega_1 < \omega_2 < \cdots < \omega_\lambda$ by locating the first ω relative maxima of the periodogram.

Kavalieris and Hannan (1994) have studied the model (1.1) under the hypothesis of random errors following an ARMA process. Namely, the model is

$$y(t) = \mu + \sum_{j=1}^{r} \alpha_j \cos(\lambda_j t) + \beta_j \sin(\lambda_j t) + u(t), \quad 0 < \lambda_j < \pi,$$

where μ, α_j, β_j, λ_j are constant to be estimated, the λ_j are unrelated and $u(t)$ is a stationary process with absolutely continuous spectrum and continuous spectral density. The authors wish to estimate the order of the model r, comparing three criteria, each of them obtained as an AIC criterion, the first introduced by Hannan (1993)

$$\phi_H(r,m) = \log\{\hat{\sigma}_r^2(m)\} + 5r\frac{\log T}{T} + \frac{\log m}{2m},$$

with

$$\log\{\hat{\sigma}_r^2\} = \frac{2M}{T}\sum_{k=1}^{M}\log\left[\frac{1}{m}\sum_{j=(k-1)m+1}^{km} I_r(\omega_j)\right]$$

and the periodogram

$$I_r(\omega_j)\frac{1}{T}\Big|\sum_{t=1}^{T}\hat{u}_r(t)e^{it\omega_j}\Big|^2, \quad \omega_j = \frac{2\pi j}{T}$$

and $\hat{u}_r(t)$ is the residual from the regression of $y(t) - \bar{y}$ on $\cos(\hat{\lambda}_j t)$, $\sin(\hat{\lambda_j t})$, $j = 1,\ldots,r$. $M = \lfloor (T-1)/2m \rfloor$ and m needs to be estimated. Notice that this regression furnishes also the estimates of α_j and β_j, while the $\hat{\lambda}_j$ can be found sequentially by locating the maximum of $I_{j-1}(\omega)$ over $0 < \omega < \pi$. The penalty terms takes into account the cost of encoding the parameters of the r fitted sinusoid. Since the r estimators $\hat{\alpha}_j$ and $\hat{\beta}_j$ have standard deviation $O(T^{-1/2})$, and the r $\hat{\lambda}_j$ has standard deviation $O(T^{-3/2})$, the code length for $\hat{\alpha}_j$, $\hat{\beta}_j$, $\hat{\lambda}_j$ are respectively $(1/2)\log T$,$(1/2)\log T$,$(3/2)\log T$, giving a total code length of $(5/2)\log T$ for each sinusoid. The term $\log m/2m$ is derived from the code length of the smoothed periodogram.

The second criterion has been proposed by Wang (1993), and can be rewritten as:

$$\phi_W(r,c) = \log\frac{1}{T}\sum \hat{u}_r(t)^2 + cr\frac{\log T}{T},$$

where the choice of c depends on $\frac{\max 2\pi f(\omega)}{\int f(\omega)d\omega}$. Hannan (1993) offers no proof of consistency of any procedure based on $\phi_H(r,n)$. The third criterium is

$$\phi(r,h) = \log\tilde{\sigma}_r^2(h) + (5r+h)\frac{\log T}{T}$$

where $\tilde{\sigma}_r^2(h)$ is an estimate of the prediction variance of $u(t)$, and h must be estimated.

$\phi_W(r,c)$ is shown to be consistent only if

$$c > \frac{2\max_\omega 2\pi f(\omega)}{\int f(\omega)d\omega}.$$

A simulation study with a real model having $r = 2$ shows that $\phi(r,h)$ behaves better, especially in presence of a "large" sample.

Under the hypothesis of $u(t)$ being an $AR(\infty)$ process, say $u(t) + \sum_{j=1}^{\infty} \kappa_j u(t-j) = \varepsilon(t)$, with $\sum |\kappa_j| < \infty$, $\kappa(z) \neq 0$ for $|z| \leq 1$, if $\varepsilon(t)$ is a stationary sequence of martingale difference, where

$$E\{\varepsilon(t)|\mathcal{F}_{t-1}\} = 0, \quad E\{\varepsilon(t)^2|\mathcal{F}_{-\infty}\} = \sigma^2, \quad E\{\varepsilon(t)^4\} < +\infty,$$

$\mathcal{F}_t$ being the filtration generated by $\varepsilon(t)$, $s \leq t$, if $\tilde{\sigma}^2(h)$ is the estimated variance from the $AR(h)$ model for the correct trigonometric regression and $\tilde{\sigma}_0^2(h)$ is the estimated variance from the $AR(h)$ model for the model $y(t) = u(t)$, then, uniformly for $h \leq H$,

$$\tilde{\sigma}_0^2(h) = \tilde{\sigma}^2(h) + \sum_{k=1}^{r_0} \frac{\rho_k^2}{2} \frac{|\kappa_h(e^{i\lambda_k})|^2 2\pi f(\lambda_k)}{h} + o(h^{-1}).$$

Under the same hypothesis, if $\hat{r}$ is the first local minimum of $\phi(r, \hat{h}_r)$, then $\hat{r} \to r_0$ almost surely.

4.4 Asymptotic equivalence of the Information criteria proposed

If the lowest Fourier frequencies have the largest ordinates, a $\phi_g(\lambda)$ information criterion can be built from the Mallows's C_p discussed above. In fact $\hat{\sigma}_\lambda^2 = \frac{1}{T}\sum_{i=0}^{T-1}(y_i - \hat{f}_\lambda(t_{i,n}))^2$, and $C(\lambda) = \hat{\sigma}_\lambda^2(1 + \frac{\lambda}{T}\frac{\hat{\sigma}_\eta^2}{\hat{\sigma}_\lambda^2})$,

$$\log C(\lambda) = \log \hat{\sigma}_\lambda^2 + \log\big(1 + \frac{\lambda}{T}\frac{\hat{\sigma}_\eta^2}{\hat{\sigma}_\lambda^2}\big) = \log \hat{\sigma}_\lambda^2 + \frac{\lambda}{T}\frac{\hat{\sigma}_\eta^2}{\hat{\sigma}_\lambda^2} + o\big(\frac{\lambda}{T}\big)$$

for $|\frac{\lambda}{T}\frac{\hat{\sigma}_\eta^2}{\hat{\sigma}_\lambda^2}| \leq 1$, and thus

$$\phi_g(\lambda) = T\log C(\lambda) + \lambda\frac{\hat{\sigma}_\eta^2}{\hat{\sigma}_\lambda^2}$$

where $g(T) = \frac{\hat{\sigma}_\eta^2}{\hat{\sigma}_\lambda^2}$, and obviously $g(T)/T = (1/T)\frac{\hat{\sigma}_\eta^2}{\hat{\sigma}_\lambda^2} \to 0$ for $T \to \infty$.

In the same fashion,if $\lambda/T \to 0$ as $T \to \infty$, the unbiased risk criterion (4.1) leads to a $\phi_g(\lambda)$ information criterion by putting $\phi_g(\lambda) = T\log(\hat{R}_\lambda) + o(T^{-1})$, for $g(T) = 2\lambda + d + 1$.

Chapter 5

Concluding remarks and further developments

In this thesis we have developed and discussed local trigonometric models for the smoothing of time series, in particular for estimating business cycle. We have seen that many models can be viewed as finite version of Wiener Kolmogorov filter, and we have proposed a solution for smoothing $ARIMA(p, d, q)$ series without prefiltering.

A local trigonometric model as been proposed, and it has been applied to the smoothing of pseudocyclical time series.

We have evaluated the rates of convergence of a trigonometric model under suitable criteria, and we have given criteria to best decide the order of the model.

We have seen that a problem associated with trigonometric estimators is the slowness of the convergence, in particular in presence of a non periodic underlying signal. This problem has been mitigated by means of local estimates arising from the minimization of a L^2 functional. Other approaches are possible.

One major problem faced when fitting a time series by means of trigonometric function is the Gibbs phenomenon, that implies a systematic distortion of the estimators at the end of the sample. This phenomenon can be mitigated thanks to a different minimizating function: in fact the a L_1-approximation of $f(x)$ shows smaller wiggles than usual L_2 approximations. A different choice of the orthonormal basis function, such as some families of wavelets functions allows to avoid Gibbs phenomenon under suitable conditions. Some hints of this possibilities are given in this section.

5.1 Gibbs phenomenon in splines

Richard and Foster (1991) discussed the problem of a Gibbs phenomenon in spline approximation. They calculate an overshoot for linear splines of 0.134. They also showed that for a spline of higher degree

$$\lim_{k\to\infty} S^{[k]}(x) = \tau(x) \quad \text{uniformly in x,}$$

for $\tau(x) = \frac{2}{\pi}Si(x) = \frac{2}{\pi}\frac{\sin(\pi x)}{\pi x}$.

Richard (1991) also showed that the value of the overshoot is independent of the location of the discontinuity ξ if ξ is irrational and depends of the overshoot if ξ is rational.

Foster and Richard were able to prove that an L_p-approximation would result in a Gibbs phenomenon even when piecewise- continuous function are used instead of the trigonometric basis of the Fourier series, and then the overshoot is much larger than the classical one of Fourier series.

Moskona, Petrushev and Saff proved that the Gibbs overshoot and undershoot occurs also for best L_p approximation, $p \geq 1$. They also noticed that the amount of overshooting is a decreasing function of p. M.P. and S. define a Gibbs function for L_1:

$$G(x) = -\frac{\sin \pi x}{\pi}\int_0^1 u^{|x|-1}\frac{1-u}{1+u}du, \quad t \neq 0, \quad G(0) = 0,$$

and the number γ as

$$\gamma = \max_{|t|\geq 1} |G(x)| \approx 0.06578389,$$

that is 1/2.7 times the Gibbs constant for L_2 approximation, and $B_n(x) = B_{n,p,f}(x)$ be the best L_p-approximation of $f(x)$ in the class T_n of the trigonometric polynomial of degree at most n. The authors showed that if f is a 2π-periodic function with only one jump at 0:

$$\lim_{n\to\infty} \max_{x\in(0,\pi]} [B_{n,1,f}(x) - f(x)] = \frac{f(0_+) - f(0_-)}{2}\gamma,$$

$$\lim_{n\to\infty} \min_{x\in[-\pi,0)} [B_{n,1,f}(x) - f(x)] = -\frac{f(0_+) - f(0_-)}{2}\gamma,$$

$$\lim_{n\to\infty} \max_{x\in(0,\pi]} [B_{n,1,f}(x) - f(x)] = \lim_{n\to\infty} \max_{x\in[\frac{\pi}{n+1},\frac{2\pi}{n+1}]} [B_{n,1,f}(x) - f(x)],$$

$$\lim_{n\to\infty} \min_{x\in[-\pi,0)} [B_{n,1,f}(x) - f(x)] = \lim_{n\to\infty} \min_{x\in[\frac{-2\pi}{n+1},\frac{-\pi}{n+1}]} [B_{n,1,f}(x) - f(x)],$$

and for $x \in (-\infty, \infty)$,

$$\lim_{n\to\infty} \Big[B_{n,1,f}\Big(\frac{\pi x}{n+1}\Big) - f\Big(\frac{\pi x}{n+1}\Big)\Big] = \frac{f(0_+) - f(0_-)}{2} G(x).$$

By integrating by parts it is shown that the function $G(x)$ can be computed as

$$G(x) = \frac{\sin \pi x}{\pi x} \sum_{k=1}^{\infty} \frac{2^{-k} k!}{(1+x)(2+x)\cdots(k+x)}, \quad x > 0.$$

By expanding $\frac{1-u}{1+u}$ in Taylor expansion and then recognizing the Fourier sine series of $\sin tu$, $u \in (-\pi, \pi)$, it is shown that

$$G(x) = \frac{1}{\pi} \int_0^{\pi} \cot(\frac{u}{2}) \sin(xu) du - 1.$$

For $p > 1$, the L_p approximating polynomial is unique, while the L_1 approximating polynomial is not unique. Uniqueness of the best interpolating polynomial in L_1 holds for f continuous or piecewise continuous with precisely one jump mod(2π).

5.2 Gibbs phenomenon in wavelets

General properties of the Fourier transform allows to state that if a function is very localized in time domain, its Fourier transform is very widespread in frequency domain, and viceversa if a function is very localized in frequency domain, its Fourier transform is very widespread in time domain. A local trigonometric regression allowed us to exploit time limited data using a limited range of frequency. A second approach that permits to overcome this duality consists in wavelet analysis.

A wavelet is a family of functions very localized in the time domain whose Fourier transform is very localized in the frequency domain.

The operator translation $T : L^2 \to L^2$ is defined by $(Tf)(x) = f(x-1)$. The dilatation $D_a : L^2 \to L^2$ with scaling function a is defined by $(D_a f)(x) = \frac{1}{\sqrt{a}} f(\frac{x}{a})$. Thus

$$((D_a^{-1})f)(x) = \sqrt{a} f(ax),$$

and

$$(T^n D_a^n f)(x) = \frac{1}{\sqrt{a^m}} f\big(\frac{x - a^m n}{a^m}\big).$$

A dyadic orthonormal wavelet is a function $\psi \in L^2(\mathbb{R})$ such that the set

$$\{2^{\frac{n}{2}}\psi(2^n t - l) : n, l \in \mathbb{Z}\}$$

forms an orthonormal basis for $L^2(\mathbb{R})$. The adjective "dyadic" refers to the scaling operator D_2.

A first family of wavelets can be devised by means of the sinc function...

Some wavelet does not exhibit the Gibbs phenomenon, ex. Haar wavelet.

Let

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \ldots$$

a chain of closed subspaces in $L_2(\mathbb{R})$ such that $\bigcap_{m\in\mathbb{Z}=\{0\}}$ and $\overline{\bigcup_{m\in\mathbb{Z}}} = L_2(\mathbb{R})$, let

$$f \in V_m \Leftrightarrow f(2\cdot) \in V_{m+1}.$$

If ϕ is the scaling function defining the wavelet transform and $\phi_{m,n}(x) = 2^{\frac{m}{2}}\phi(2^m x - n)$ is an orthonormal basis of $V_0 \subset V = L_2(\mathbb{R})$, there exist the subspaces W_m, $m \in \mathbb{Z}$ such that for each m, $V_{m+1} = V_m \oplus W_m$ and

$$\sum_{m=-\infty}^{\infty} \oplus W_m = L_2(\mathbb{R}).$$

For $f \in L_2(\mathbb{R})$ it is defined the projection onto V_m as

$$\begin{aligned} f_m(x) =& P_m f(x) = \sum_{n\in\mathbb{Z}} \phi_{m,n}(x)\phi^*_{m,n} f = \\ & \int \sum_{n\in\mathbb{Z}} \phi_{m,n}(x)\phi_{m,n}(y)f(y)dy = \int K_m(x,y)f(y)dy, \end{aligned}$$

where

$$K_m(x,y) = \sum_{n\in\mathbb{Z}} \phi_{m,n}(x)\phi_{m,n}(y).$$

Kelly (1991) showed that a Gibbs effect occurs near the origin if and only if

$$\int_0^\infty K_0(a,u)du > 1 \quad \text{for some} \quad a > 0,$$

or

$$\int_0^\infty K_0(a,u)du > 0 \quad \text{for some} \quad a < 0.$$

A positive kernel allows to reduce the Gibbs phenomenon for wavelets, such as the Fejer kernel reduces the Gibbs phenomenon for trigonometric expansion.

Wavelet analysis can be used to accommodate structural changes of a more varied nature than LTI structures.

Appendix A

Convolution products

Theorem 5. *If the periodic functions $f(t)$ e $g(t)$ admit the Fourier expansions $f(t) = \sum_{l=-\infty}^{-\infty} c_l e^{ilt}$ and $g(t) = \sum_{l=-\infty}^{-\infty} d_l e^{ilt}$ respectively, then the product function $f(t)g(t)$ admits Fourier expansion, and its Fourier coefficients w_t are obtained as convolution product of $\{c_k\}_k$ and $\{d_k\}_k$: $w_t = \sum_{n=-\infty}^{+\infty} c_n d_{k-n}$.*

PROOF: $f(t)g(t)$ is a periodic function, and admits a Fourier expansion $f(t)g(t) = \sum_{k=-\infty}^{+\infty} w_k e^{ikt}$. Now

$$\begin{aligned} w_k &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)g(t)e^{-ikt}dt = \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{l=-\infty}^{+\infty} c_l e^{ilt}\sum_{r=-\infty}^{+\infty} d_r e^{irt}e^{-ikt}dt \\ &= \frac{1}{2\pi}\sum_{l=-\infty}^{+\infty}\sum_{r=-\infty}^{+\infty} c_l d_r \int_{-\pi}^{\pi} e^{i(l+r-k)}dt = \sum_{l=-\infty}^{+\infty}\sum_{r=-\infty}^{+\infty} c_l d_r \delta_{l+r-k} = \sum_{l+r=k} c_l d_r \end{aligned}$$

that is the thesis.□

If it is calculated the convolution product of $\mathcal{F}(\chi_I)$, χ_I characteristic function of the interval $[-b,-a]\cup[a,b]$ with a trigonometric polynomial $\sum_{k=-\infty}^{+\infty}\psi_k z^k$, with $\sum_{k=-\infty}^{+\infty}\psi_k < \infty$, there are found the identities:

$$\sum_{k=-\infty}^{+\infty}\frac{\sin(bk)-\sin(ak)}{\pi k}\psi_{k-n} = \psi_k, \quad k \in I,$$

$$\sum_{k=-\infty}^{+\infty}\frac{\sin(bk)-\sin(ak)}{\pi k}\psi_{k-n} = 0, \quad k \notin I$$

The best approximation in L^2 of $\chi_I(\omega)\theta(\omega)/\phi(\omega)$ subject to the constraint that the

Fourier coefficient c_j are null for $|j| > K$ is

$$\sum_{m=-K}^{K} \delta_m(B^*\psi)e^{im\omega}.$$

Appendix B

Useful trigonometric Identities

Let $k \in \mathbb{Z}$, $0 \leq k < T$.

$$\sum_{t=0}^{T-1} e^{\frac{i2\pi kt}{T}} = T \quad for \quad k = 0$$

$$\frac{1 - e^{i2\pi k}}{1 - e^{\frac{i2\pi k}{T}}} = 0 \quad for \quad k \neq T.$$

Since

$$\sum_{t=0}^{T-1} e^{\frac{i2\pi kt}{T}} = \sum_{t=0}^{T-1} \cos\left(\frac{2\pi kt}{T}\right) + i \sum_{t=0}^{T-1} \sin\left(\frac{2\pi kt}{T}\right)$$

we obtain

$$\sum_{t=0}^{T-1} \cos\left(\frac{2\pi kt}{T}\right) = T \quad for \quad k = 0$$

$$0 \quad for \quad k \neq 0,$$

and

$$\sum_{t=0}^{T-1} \sin\left(\frac{2\pi kt}{T}\right) = 0.$$

Recall the Werner identities:

$$\sin nx \sin mx = \frac{1}{2}(\cos(m-n)x - \cos(m+n)x),$$

$$\sin nx \cos mx = \frac{1}{2}(\sin(m-n)x + \sin(m+n)x),$$

$$\cos nx \cos mx = \frac{1}{2}(\cos(m-n)x + \cos(m+n)x).$$

Assume $k, l \in \mathbb{Z}$, $0 < k, l < T/2$, in order to avoid problems of aliasing. We have

$$\begin{aligned}\sum_{t=0}^{T-1} \sin\Big(\frac{2\pi kt}{T}\Big) \sin\Big(\frac{2\pi lt}{T}\Big) =& \frac{1}{2}\sum_{t=0}^{T-1}\Big(\cos\frac{2\pi(k-l)t}{T} - \cos\frac{2\pi(k+l)t}{T}\Big) = \\ & \frac{T}{2} \quad if \quad k = l, \\ & 0 \quad if \quad k \neq l.\end{aligned}$$

$$\begin{aligned}\sum_{t=0}^{T-1} \cos\Big(\frac{2\pi kt}{T}\Big) \cos\Big(\frac{2\pi lt}{T}\Big) =& \frac{1}{2}\sum_{t=0}^{T-1}\Big(\cos\frac{2\pi(k-l)t}{T} + \cos\frac{2\pi(k+l)t}{T}\Big) = \\ & \frac{T}{2} \quad if \quad k = l, \\ & 0 \quad if \quad k \neq l.\end{aligned}$$

$$\sum_{t=0}^{T-1} \sin\Big(\frac{2\pi kt}{T}\Big) \cos\Big(\frac{2\pi lt}{T}\Big) = \frac{1}{2}\sum_{t=0}^{T-1}\Big(\sin\frac{2\pi(k-l)t}{T} + \sin\frac{2\pi(k+l)t}{T}\Big) = 0.$$

If $k = l = 0$ we have

$$\sum_{t=0}^{T-1} \cos\Big(\frac{2\pi kt}{T}\Big)^2 = T,$$

$$\sum_{t=0}^{T-1} \sin\Big(\frac{2\pi kt}{T}\Big)^2 = 0.$$

If $f \notin \mathbb{Z}$, then

$$\sum_{t=1}^{T} e^{2\pi i f t} = e^{i\pi f(T-1)}\frac{\sin(T\pi f)}{\sin(\pi f)},$$

hence, if $f + f', f - f' \notin \mathbb{Z}$,

$$\begin{aligned}\sum_{t=1}^{T} \cos(2\pi f t)\cos(2\pi f' t) =& \frac{1}{2}\frac{\sin(T\pi(f-f'))}{\sin(\pi(f-f'))}\cos((T+1)\pi(f-f')) + \\ & + \frac{1}{2}\frac{\sin(T\pi(f+f'))}{\sin(\pi(f+f'))}\cos((T+1)\pi(f+f')),\end{aligned}$$

$$\begin{aligned}\sum_{t=1}^{T}\cos(2\pi ft)\sin(2\pi f't) =&\frac{1}{2}\frac{\sin(T\pi(f+f'))}{\sin(\pi(f+f'))}\sin((T+1)\pi(f+f'))+\\&+\frac{1}{2}\frac{\sin(T\pi(f-f'))}{\sin(\pi(f-f'))}\sin((T+1)\pi(f-f')),\end{aligned}$$

$$\begin{aligned}\sum_{t=1}^{T}\sin(2\pi ft)\sin(2\pi f't) =&\frac{1}{2}\frac{\sin(T\pi(f-f'))}{\sin(\pi(f-f'))}\cos((T+1)\pi(f-f'))+\\&-\frac{1}{2}\frac{\sin(T\pi(f+f'))}{\sin(\pi(f+f'))}\cos((T+1)\pi(f+f'))\end{aligned}$$

In $f \notin \mathbb{Z}$ then

$$\sum_{t=1}^{T}\cos(2\pi ft)^2 = \frac{N}{2}+\frac{1}{2}\frac{\sin(T2\pi f)}{\sin(2\pi f))}\cos((T+1)2\pi f)),$$

$$\sum_{t=1}^{T}\cos(2\pi ft)\sin(2\pi ft) = \frac{1}{2}\frac{\sin(T2\pi f)}{\sin(2\pi f)}\sin((T+1)2\pi f),$$

$$\sum_{t=1}^{T}\sin(2\pi ft)^2 = \frac{N}{2}-\frac{1}{2}\frac{\sin(T\pi(f+f'))}{\sin(\pi(f+f'))}\cos((T+1)\pi(f+f')).$$

Bibliography

[1] An, Hong-Zi, Chen, Zhao-Guo, Hannan, E. J., (1983), The maximum of the periodogram. *J. Multivar. Anal.* **13** 383-400

[2] Baxter, M., King R. G., (1999) Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series, *The Review of Economics and Statistics*, **81** 575-93.

[3] Bell, W., (1984), Signal Extraction for Nonstationary Time Series, *The Annals of Statistics*, **12**, 646-664.

[4] Box,G. E. P., Jenkins, G. M., (1970) *Time Series Analysis: Forecasting and Control*, San Francisco: Holden Day.

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

[6] Christiano, L. J., Fitzgerald, T. J., (2003) The Band Pass Filter, *Intarnational Economic Review* **44** 435-65.

[7] Cox,D.D. (1988), Approximation of Least Squares Regression on Nested Subspaces, *The Annals of Statistics* **16**, 713-732.

[8] Dagum, E. B., (2001), Time Series: Seasonal Adjustment, in *International Encyclopedia of the Social and Behavioral Science*, vol. II, Metodology, Fienberg, S. and Kadane, J. Edotors, Elsevier Science Pub.

[9] Dagum, E. B., Modelling, Forecasting and Seasonally Adjusting conomic Time Series with the X-11 ARIMA Method, *Journal of the Royal Statistical Society, Series D*, **26** , 203-216.

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

[11] Eubank, R. L., Speckman, P.(1990), Curve Fitting by Polynomial-Trigonometric Regression, *Biometrika* **77** 1-9.

[12] Eubank, R. L., Speckman, P.(1991) Convergence rates for trigonometri and polynomial-trigonometric regression estimators, *Statist. Proba. Letters* **11**, 119-124.

[13] Fan, J., Gijbels, I.,(1996), *Local Polynomial Modelling and Its Applications*, Chapman & Hall.

[14] Fisher, R. A., Test of significance in Harmonic analysis, *Proceedings of the Royal Society of London* s. A **125** 54-59.

[15] Foster, J., Richard, F. B., (1991), The Gibbs phenomenon for piecewise-linear approximation, *Amer. Math Monthly* **98**, 47-49.

[16] Grenander, U., Rosenblatt, M., (1957), *Statistical Analysis of Stationary Time Series*, Wiley.

[17] Gallavotti, G., (2001), Quasi-periodic motions from Hypparcus to Kolmogorov, *Rend. Mat. Acc. Lincei* s.9 v.12, 125-152.

[18] Hall, P.(1981), On trigonometric series estimates of densities, *Ann. Statist.* **95** 683-5.

[19] Hall, P. (1983), Measuring the efficiency of trigonometric series estimates of a density, *J. Mult. Anal* **13** 234-56.

[20] Hannan, E., J., Terrell,R. D., Tuckwell, N. E., (1970), The Seasonal Adlustment of Economic Series, *International Economic Review*, **11** 24-52.

[21] Hannan, E. J., (1973), The estimation of frequencies, *J. Appl. Prob.* **10**, 510-19

[22] Hannan, E. J., Nicholls, D. F.,(1977) The Estimation of the Prediction Error Variance. *JASA*, **72** 834-840.

[23] Hannan, E. J., (1993) Determining the number of jumps in a spectrum. *Developments in Time Series Analysis* (Ed. T. Subba Rao) London: Chapman and Hall, pp. 127-38.

[24] Harvey, A., C., (1993) *Time Series Models*, Harvester-Wheatsheaf.

[25] Harvey, A., Striebel, M., (1998), Tests for Deterministic Versus Indeterministic Cycles, *Journal of Time Series Analysis*, **19** 505-529.

[26] Harvey, A. C., Trimbur, T. M.,(2003) General model-based filters for extracting cycles and trends in economic time series, *The Review of Economics and Statistics*, **85** 244-255.

[27] Hodrick, R. J., Prescott, E. C. (1997), Postwar U.S. Business Cycle: An Empyrical Investigation, *Journal of Money, Credit and Banking*, **29** 1-16.

[28] Jerri, A. J. (1998), *The Gibbs Phenomenon in Fourier Analysis, Splines and Wavelet Approximation*, Kluwer Akademic Publisher.

[29] Kaiser, R., Maravall, A., (1999)Estimation of the bisuness cycle: A modified Hodrick-Prescott filter, *Spanish Economic Review*,**1**, 175-200.

[30] Kaiser, R., Maravall, A., (2005)Combining filter design with model-based filtering (with an application to business cycle estimation), *International Journal of Forecasting*,**21**, 691-710.

[31] Koopmans, L. H., (1974), *The Spectral Analysis of Time Series*, Academic Pess Inch.

[32] Kavalieris, L., Hannan, E. J., (1994), Determining the number of terms in a Trigonometric Regression, *Journal of Time Series Analysis*, **15** 613-625

[33] King, R. G., Rebelo, S., Low Frequency Filtering and Real Business Cycle, *Journal of Economic Dynamic and Control*, **17**, 207-231.

[34] Kolmogorov, A. N. (1941), Interpolation and Extrapolation, *Bulletin de l'Academie des Science des U.R.S.S.*, **5**, 3-14.

[35] Loader, C., (1999), *Local Regression and Likelyhood*, Springer.

[36] Luati, A., Proietti, T., (2010), On the equivalence of weighted least squares and the generalised least squares estimators, with appplication to kernel smoothing, *Ann. Ist. Stat. Math.* **63** 851-71.

[37] Mallows, C. L., (1973), Some Comments on C_p. *Technometrics* **15** 661-75

[38] Moskona, E,, Petrushev, P., Saff, E. B., (1995), Gibbs phenomenon for best L_1-trigonometric polynomial approximation, *An Inter. J. of Approx. and Expansion*, **11**, 391-416.

[39] Percival D. B.; Walden A. T., (1998) *Spectral Analysis for Phisical Applications*, Cambridge University Press.

[40] Pollock. D.S.G., (1999) *A Handbook of Time Series Analysis, Signal Processing and Dynamics*, Academic Press.

[41] Pollock,D.S.G., (2012), Band-Limited Stochastic Processes in Discrete and Continuous Time. *Studies in Nonlinear Dynamics & Econometrics*, **16**, 1-27.

[42] Popinski, W., (1999), Least-Squares Trigonometric Regression Estimation, *Ann. Ist. Stat. Math.* **63** 851-71.

[43] Pötscher, B. M., Leeb. H., Model Selection and Inference: Facts and Fiction, *Econometric Theory* **1**, 21-59.

[44] Proietti, T., Luati, A., (2007), Least Squares Regression: Graduation and Filters, in *Measurement in Economics:A Handbook* Marcel Boumans Editor, Elsevier, 377-411.

[45] Proietti, T., Luati, A., (2010), Low-pass filter design using locally weighted polynomial regression and discrete prolate spheroidal sequences, *Journal of Statistical Planning and Inference* **141** 831-45.

[46] Rafajlowicz, E.(1987), Nonparametric orthogonal series estimators of regression: a class attaining the optimal convergence rate in L_2, *Statist. Prob. Letters***5** 219-24.

[47] Quinn, B.G., (1989) Estimating the number of terms in a sinusoidal regression, *Journal of Time Series Analysis* **10** 71-75.

[48] Richard, F. B. (1991), A Gibbs phenomenon for splines functions, *Journal of approximation theory* **66**, 344-351.

[49] Rudin, W. (1986), *Real and Complex Analysis*, Mc Graw-Hill Book Company.

[50] Ruppert, D., Wand, M. P., Carroll, R. J., (2003), *Semiparametric Regression*, Cambridge University Press.

[51] Schuster, A., (1897), On Lunar and Solar Periodicity of Earthquakes, *Proceedings of the Royal Society of London* **61** 455-465.

[52] Schuster, A., (1906) On the periodicity of the sunspot, *Philosophical Transaction of the Royal Society of London* s.A **206**, 69-100.

[53] Speckman, P. (1985), Spline smoothing and optimal rates of convergence in nonparametric regression models, *Ann. Statist.***13** 970-83.

[54] Stone, C. (1982), Optimal global rate of convergence for nonparametric regression. *Ann. Statist.***10** 1040-53.

[55] Wahba, G., (1990) *Spline Methods for Observational data*, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

[56] Walker, A., M., (1971), On estimation of Harmonic Component in a Time Series with Stationary Independent Residuals, *Biometrika* **58**, 21-36.

[57] Walter, G., Blum, J.,(1979) Probablity density estimation using delta sequencies, *Ann. Statist.***7** 328-40.

[58] Wang, Xiaobao (1993) An AIC type estimator for the number of cosinusoids. *J. Time Ser. Anal.*, **14** 431-40.

[59] Whittle, P., (1983) *Prediction and Regulation, II ed.*, University of Minnesota Press, Minneapolis, Minnesota.

[60] Wiener, N., (1949) *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, John Wiley and Sons.

[61] Zygmund, A:, (1959) *Trigonometric Series*, Cambridge University Press.

Index