AARHUS
UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

Generalised partial autocorrelations and the mutual information between past and future

Tommaso Proietti and Alessandra Luati

CREATES Research Paper 2015-24

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Generalised partial autocorrelations and the mutual information between past and future.*

Alessandra Luati†
University of Bologna

Tommaso Proietti‡
University of Rome Tor Vergata and CREATES

May 25, 2015

Abstract

The paper introduces the generalised partial autocorrelation (GPAC) coefficients of a stationary stochastic process. The latter are related to the generalised autocovariances, the inverse Fourier transform coefficients of a power transformation of the spectral density function. By interpreting the generalised partial autocorrelations as the partial autocorrelation coefficients of an auxiliary process, we derive their properties and relate them to essential features of the original process.

Based on a parameterisation suggested by Barndorff-Nielsen and Schou (1973) and on Whittle likelihood, we develop an estimation strategy for the GPAC coefficients. We further prove that the GPAC coefficients can be used to estimate the mutual information between the past and the future of a time series.

Keywords: Generalised autocovariance, Spectral models, Whittle likelihood, Reparameterisation.

JEL codes: C22, C52.

*Paper prepared for the *Aarhus Conference on Probability, Statistics and their Applications: Celebrating the scientific achievements of Ole E. Barndorff-Nielsen*, Aarhus, 15–19 June 2015.

†Dipartimento di Scienze Statistiche "Paolo Fortunati", Università di Bologna, Via Belle Arti 41, 40126 Bologna, Italy. E-mail: alessandra.luati@unibo.it.

‡Dipartimento di Economia e Finanza, Università di Roma Tor Vergata, Via Columbia 2, 00133 Rome, Italy. E-mail: tommaso.proietti@uniroma2.it. Tommaso Proietti acknowledges support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

1 Introduction

The spectral density of a covariance stationary process subsumes all the information that is needed for best linear prediction of the future process realisations. In this paper, we consider a class of spectral estimates that is based on a generalised power autoregressive scheme, depending on a finite dimensional parameter set and on a power transformation parameter, which can encompass different spectral estimation methods, among which autoregressive and moving average estimation.

In an earlier paper (Proietti and Luati, 2015), spectral estimation was performed by a generalised set of Yule-Walker equations, based on the generalised autocovariance function, the inverse Fourier transform of a power transformation of the spectrum, which was estimated non-parametrically by a power transformation of the periodogram.

In this paper, we consider a parameterisation of spectral models based on a set of unrestricted coefficients taking values in the interval $(-1, 1)$, that are interpretable as generalised partial autocorrelations (GPAC). The GPAC are then related to the mutual information between past and future, a measure of uncertainty about the future based on knowledge of the past, considered in Ibragimov and Rozanov (1978), and recently revisited by Li and Xie (1996) and Li (2005). The novelty here is in the estimation of the mutual information based on the GPAC, via the Whittle likelihood. The computation of the mutual information between past and future entails the availability of the full partial autocorrelation sequence, unless the process is autoregressive, in which case the partial autocorrelation is truncated. Our approach amounts to determining a scale, implied by the power transformation parameter, along which the GPAC sequence is finite.

The paper is organised as follows. Section 2 reviews the generalised autocovariance function and introduces the GPAC as the partial autocorrelation coefficients of an auxiliary power process. The relation between the GPAC and the mutual information between the past and the future is established in section 4, along with an optimality property of the class of generalised spectral autoregressive models, formerly introduced in section 3 and parameterised by the GPAC, based on Barndorff-Nielsen and Schou (1973). Whittle likelihood estimation is dealt with in section 5. An illustration featuring the time series of the U.S. monthly inflation rate, levels and first differences, is reported in section 6. Finally, section 7 draws some conclusions.

2 Generalised autocovariances

Let $\{x_t\}_{t\in T}$ be a stationary zero-mean stochastic process indexed by a discrete time set T, with spectral density function $f(\omega)$ such that $\int_{-\pi}^{\pi} \log f(\omega) \mathrm{d}\omega > -\infty$, and $\int_{-\pi}^{\pi} f(\omega)^p \mathrm{d}\omega < \infty$, $p \in \mathbb{R}$.

The generalised autocovariance (GACV) function is defined (Proietti and Luati, 2015) as the inverse

Fourier transform of the p-th power of the spectral density function,

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

for $k = 0, 1, \ldots$ and $\gamma_{p,-k} = \gamma_{pk}$. The discrete Fourier transform of γ_{pk} gives

$$[2\pi f(\omega)]^p = \gamma_{p0} + 2\sum_{k=1}^{\infty}\gamma_{pk}\cos(\omega k).$$

The coefficients γ_{pk} depend on two arguments, the integer lag k and the real power p. As a matter of fact, for $p = 1, \gamma_{1k} = \gamma_k$, the autocovariance of the process at lag k; for $p = 0, \gamma_{0k} = 0$, for $k \neq 0$ and $\gamma_{00} = 1$, up to a constant, the autocovariance function of a white noise process; for $p = -1, \gamma_{-1k} = \gamma_{ik}$, the inverse autocovariance function (Cleveland, 1972; see also Battaglia, 1983).

The function γ_{pk} lends itself to a further interpretation as the autocovariance function of an auxiliary process derived from x_t. This interpretation turns out to be useful in the derivation of the analytic form of γ_{pk}, as a function of the parameters that govern the process dynamics, by evaluating an expectation in the time domain, rather than solving (1) directly. In addition, the generalised partial autocorrelation function of x_t will be directly defined based on the auxiliary process.

Assuming that $\{x_t\}_{t\in T}$ is purely non-deterministic, its Wold representation is written as

$$x_t = \psi(B)\xi_t, \tag{2}$$

where $\xi_t \sim \mathrm{WN}(0, \sigma^2)$ and $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots$, with coefficients satisfying $\sum_{j=0}^{\infty}\psi_j^2 < \infty$, and such that all the roots of the characteristic equation $\psi(B) = 0$ are in modulus greater than one; here, $\mathrm{WN}(0, \sigma^2)$ denotes a white noise process, a sequence of zero mean and uncorrelated random variables with constant variance σ^2 and B is the backshift operator, $B^k x_t = x_{t-k}$. The autocovariance function of the linear process (2) is $\gamma_k = \sigma^2\sum_{j=0}^{\infty}\psi_j\psi_{j+k}$ for $k = 0, 1, \ldots$, and $\gamma_{-k} = \gamma_k$.

Let us consider the power-transformed process:

$$u_{pt} = \begin{cases} \psi(B)^p\xi_t &= \psi(B)^p\psi(B)^{-1}x_t, \quad \text{for } p \geq 0 \\ \psi(B^{-1})^p\xi_t &= \psi(B^{-1})^p\psi(B)^{-1}x_t, \quad \text{for } p < 0. \end{cases} \tag{3}$$

For a real $p > 0$, the power of $\psi(B)$ in (3) is still a power series,

$$\psi(B)^p = \sum_{j=0}^{\infty}\varphi_j B^j,$$

with coefficients given by the recursive relation

$$\varphi_j = \frac{1}{j}\sum_{h=1}^{j}[h(p+1) - j]\psi_h\varphi_{j-h}, j > 0, \;\; \varphi_0 = 1 \tag{4}$$

(Gould, 1974). When $p < 0$, the coefficients of $\psi(B^{-1})^p = \sum_{j=0}^{\infty} \varphi_j B^{-j}$ are also given by (4).

The spectral density of u_{pt} is $f_u(\omega) = (2\pi)^{-1}|\psi(e^{\imath\omega})|^{2p}\sigma^2$, and satisfies

$$2\pi f_u(\omega)(\sigma^2)^{p-1} = [2\pi f(\omega)]^p \,. \tag{5}$$

Equation (5) establishes the relation between the spectrum of the original process and that of the power process u_{pt}.

It follows from (1) and (5) that the generalised autocovariance function of x_t can be interpreted as the autocovariance function of the process u_{pt}, denoted as γ_u,

$$\gamma_{pk} = (\sigma^2)^{p-1}\gamma_u.$$

It is then straightforward to compute the GACV of x_t as the autocovariance of a linear process,

$$\gamma_{pk} = \sigma^{2p} \sum_{j=0}^{\infty} \varphi_j \varphi_{j+k}.$$

The generalised variance γ_{p0} is related to the variance profile, defined in Luati, Proietti and Reale (2012) as the Hölder, or power, mean of the spectrum of x_t:

$$v_p = \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} [2\pi f(\omega)]^p \right\}^{\frac{1}{p}} .$$

Specifically, for $p \neq 0$, $v_p = \gamma_{p0}^{\frac{1}{p}}$. As a particular case, $v_{-1} = \gamma_{-1,0}^{-1}$ is the interpolation error variance $\mathrm{Var}(x_t|\mathcal{F}_{\setminus t})$, where $\mathcal{F}_{\setminus t}$ is the past and future information set excluding the current x_t; this is also interpreted as the harmonic mean of the spectrum. The limit of v_p for $p \to 0$ yields the prediction error variance, $\lim_{p\to 0} v_p = \sigma^2$, which is the geometric average of the spectral density, $\sigma^2 = \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log 2\pi f(\omega) \mathrm{d}\omega \right\}$, i.e. the Szegö-Kolmogorov formula.

By dividing the generalised autocovariance by the generalised variance, one gets the generalised autocorrelation function (GACF), taking values in $[-1, 1]$,

$$\rho_{pk} = \frac{\gamma_{pk}}{\gamma_{p0}}, \quad k = 0, \pm 1, \ldots .$$

If the GACV of x_t is proportional to the autocovariance function of the auxiliary process u_{pt}, the GACF is equal to the ACF of the auxiliary process. The same holds for the generalised partial autocorrelation coefficients of x_t that are defined here as the sequence of the partial autocorrelation coefficients of u_{pt} and are denoted as π_{pk}.

3 Generalised autoregressive spectral models

The generalised partial autocorrelations are central for estimating the spectral density of a stochastic process according to the following class of models,

$$2\pi f(\omega) = \left[\frac{\sigma_p^2}{\phi_p(e^{-\imath\omega})\phi_p(e^{\imath\omega})} \right]^{\frac{1}{p}} \tag{6}$$

where $\phi_p(e^{-\imath\omega}) = 1 - \phi_{p1}e^{-\imath\omega} - \phi_{p2}e^{-\imath\omega 2} - \cdots - \phi_{pK}e^{-\imath\omega K}$. As p varies, (6) defines a set of spectral models encompassing the AR(K) case ($p = 1$), the MA(K) case ($p = -1$), as well as the fractional case (consider, for instance, the combination $K = 1$, $p = 1/d$ and $\phi_{p1} = 1$).

The coefficients $\{\phi_{pk}, k = 1, \ldots, K\}$ satisfy the following moment conditions (Yule Walker equations): $\Gamma_{p,K}\phi_{p,K} = \gamma_{p,K}$, where $\Gamma_{p,K}$ denotes the Toeplitz matrix, formed from the GACV, with generic element $\gamma_{p,|h-k|}, h, k = 0, \ldots, K-1$, $\gamma_{p,K} = (\gamma_{p1}, \ldots, \gamma_{pK})'$, and $\phi_{p,K} = (\phi_{p1}, \ldots, \phi_{pK})'$; moreover, $\sigma_p^2 = \gamma_{p0} - \phi'_{p,K}\gamma_{p,K}$.

For $p > 0$, the polynomial $\phi_p(B) = 1 - \phi_{p1}B - \cdots \phi_{pK}B^K$ characterises the AR approximation of the process u_{pt}, and provides the spectral factorisation $[2\pi f(\omega)]^p = \sigma_p^2[\phi_p(e^{-\imath\omega})\phi_p(e^{\imath\omega})]^{-1}$. By equations (3)-(5), we obtain the AR approximation of the original process, $\pi(B)x_t = \xi_t$, $\pi(B) = [\phi_p(B)]^{1/p}$, or, equivalently, the moving average representation $x_t = \psi(B)\xi_t$, $\psi(B) = [\phi_p(B)]^{-1/p}$. For $p < 0$, the polynomials are in the forward operator B^{-1}.

According to this parameterisation, the GPAC forms a finite sequence. Moreover, as will be stated formally in section 4, processes having spectrum specified as in (6) are optimal with respect to a mutual information criterion.

For a given p, the model (6) features $K+1$ parameters, which can be estimated by maximising the Whittle likelihood (see details on section 5), under the restriction that the spectral density $[2\pi f(\omega)]^p$ is bounded away from zero and infinity at all frequencies. This is achieved by enforcing the constraints $0 < |\phi_p(e^{-\imath\omega})|^2 < \infty$, $\forall\omega \in (-\pi, \pi)$ or, equivalently, that the roots of the lag polynomials $\phi_p(B)$ do not lie on the unit circle.

Hence, for the purpose of estimation, we reparameterise the AR coefficients in terms of partial autocorrelations (Barndorff-Nielsen and Schou, 1973). This is done by a recursive algorithm: letting $\pi_{pk}, k = 1, \ldots, K$ denote K coefficients, such that $|\pi_{pk}| < 1$, then for $k = 1, \ldots, K$ and $j = 1, \ldots, k-1$, compute

$$\phi_{pj}^{(k)} = \phi_{pj}^{(k-1)} - \pi_{pk}\phi_{p,k-j}^{(k-1)}, \quad \phi_{pk}^{(k)} = \pi_{pk}. \tag{7}$$

The final iteration returns coefficients that are in the stationary region.

The coefficients π_{pk}, that are constrained in the range (-1,1), are in turn obtained as the Fisher inverse transformations of unconstrained real parameters ϑ_{pk}, $k = 1, \ldots, K$, e.g. $\pi_{pk} = \frac{\exp(2\vartheta_{pk})-1}{\exp(2\vartheta_{pk})+1}$ for $k = 1, \ldots, K$. Also, we set $\vartheta_{p0} = \ln(\sigma_p^2)$.

4 Mutual information between past and future

The generalised partial autocorrelation coefficients play a role in the estimation of the mutual information between the past and the future of a stationary Gaussian time series. The latter, denoted as I_{p-f}, is a measure of uncertainty about the future $\mathcal{F}_t$, the sigma-algebra generated by $\{x_{t+1}, x_{t+2}, \dots\}$, when the past $\mathcal{P}_t$, the sigma-algebra generated $\{x_t, x_{t-1}, \dots\}$, is known.

More formally, let us consider two stochastic processes, $\{x_t\}_{t\in T}$ and $\{y_s\}_{s\in S}$, defined on a probability space $(\Omega, \mathcal{F}, P)$, and denote as $\mathcal{S}_1$ and $\mathcal{S}_2$ the minimal sigma-algebra generated by $\{x_t\}_{t\in T}$ and $\{y_s\}_{s\in S}$, respectively. The amount of information of the random process $\{x_t\}_{t\in T}$ given by the process $\{y_s\}_{s\in S}$ is (see Ibragimov and Rozanov, 1978, chapter IV),

$$I(x, y) = \sup \sum P(A_i \cap B_j) \ln \frac{P(A_i \cap B_j)}{P(A_i)P(B_j)}, \tag{8}$$

where the supremum is taken over all the possible finite partitions of Ω in the non intersecting events $(A_i)_{i=1,\dots,n}, (B_j)_{j=1,\dots,m}$, where $A_i \in \mathcal{S}_1$ for all $i = 1, \dots, n$ and $B_j \in \mathcal{S}_2$ for all $j = 1, \dots, m$.

It is easy to verify that the following properties hold: $I(x, y) \geq 0$, with $I(x, y) = 0$ when the sigma-algebra $\mathcal{S}_1$ is independent of $\mathcal{S}_2$; $I(x, y) = I(y, x)$, which motivates the name of mutual information for the quantity in equation (8). An important concept related to mutual information is the information regularity of a stochastic process. A stationary random process is said to be information regular if

$$I_\tau = I(\{x_t\}_{t<s}, \{x_t\}_{t\geq s+\tau}) \to 0, \quad \tau \to \infty.$$

The value I_τ is sometimes referred to as the information regularity coefficient and I_0 is the mutual information between past and future, that we shall denote here as I_{p-f}.

For Gaussian processes, I_{p-f} may be defined in terms of the cepstral coefficients,

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln[2\pi f(\omega)] \cos(\omega k) \mathrm{d}\omega, \quad k = 1, 2, \dots \tag{9}$$

as (Li, 2005)

$$I_{p-f} = \frac{1}{2} \sum_{k=1}^{\infty} k c_k^2 \tag{10}$$

and a necessary condition for information regularity is $\sum_{k=1}^{\infty} k c_k^2 < \infty$. Li (2005) also shows that the following relation holds between cepstral coefficients and the partial autocorrelation coefficients, $\{\pi_k\}_{k=1,2,\dots}$, the so called reflectrum identity

$$\sum_{k=1}^{\infty} k c_k^2 = -\sum_{k=1}^{\infty} k \ln(1 - \pi_k^2) \tag{11}$$

and $c_0 = \ln \gamma_0 + \sum_{k=1}^{\infty} \ln(1 - \pi_k^2)$, the latter being a consequence of the Kolmogorov-Szegö formula.

It is shown in Proietti and Luati (2014) that the cepstral coefficients are related to the GACV via

$$c_k = \lim_{p\to 0} \frac{\gamma_{pk}}{p}, \quad k = 1, 2, \ldots.$$

The next theorem shows that the mutual information can be written as a function of the generalised partial autocorrelation coefficients.

Theorem 1 *Let π_{pk} denote the generalised partial autocorrelations of the stationary process $\{x_t\}_{t\in T}$. The mutual information between past and future is*

$$I_{p\text{-}f} = -\frac{1}{2p^2}\sum_{k=1}^{\infty} k \ln(1-\pi_{pk}^2)$$

and the equality holds for all p.

Proof Let us denote as I^u_{p-f} the mutual information between past and future of the process u_{pt}, defined by equation (3), having spectrum $2\pi f_u(\omega) = [2\pi f(\omega)]^p(\sigma^2)^{1-p}$, as in (5), and prediction error variance σ^2.

By (10), I^u_{p-f} may be written as a function of the cepstral coefficients of u_{pt}, denoted here as c_{uk}, the inverse Fourier coefficients of the logarithm of the spectral density function of u_{pt}, see equation (9),

$$\log 2\pi f_u(\omega) = c_{u0} + 2\sum_{k=1}^{\infty} c_{uk}\cos(\omega k). \tag{12}$$

It follows from (5) that $\log 2\pi f_u(\omega) = p\log 2\pi f(\omega) + (1-p)\log\sigma^2$ and, from (9),

$$\log 2\pi f_u(\omega) = p\left(c_0 + 2\sum_{k=1}^{\infty} c_k\cos(\omega k)\right) + (1-p)\log\sigma^2. \tag{13}$$

Combining (12) and (13), and noticing that $c_{u0} = c_0 = \log\sigma^2$, we get the relation between the cepstral coefficients of the original process and the auxiliary process

$$c_{uk} = pc_k$$

which clearly implies that $\sum_{k=1}^{\infty} kc_{uk}^2 = \sum_{k=1}^{\infty} kp^2c_k^2$ and $p^2 I_{p-f} = I^u_{p-f}$. By the reflectrum identity applied to the process u_{pt}, we are able to conclude that

$$p^2 I_{p-f} = -\frac{1}{2}\sum_{k=1}^{\infty} k\ln(1-\pi_{pk}^2)$$

which completes the proof of the theorem. ■

The computation of I_{p-f} entails the availability of the full partial autocorrelation sequence, unless the process is autoregressive, in which case the partial autocorrelation is truncated at K. The approach followed

in this paper amounts to determining a scale, determined by the transformation parameter p, along which the GPAC sequence is finite. The following theorem, that generalises theorem 3.1 of Li and Xie (1996), establishes the optimality of the generalised spectral autoregressive models (6) with respect to the minimum mutual information principle.

Theorem 2 *A process with given generalised autocovariances $\gamma_{pk}, k = 0, 1, \ldots, K$, $p \neq 0$ and minimal information between past and future belongs to the class of spectral models (6).*

Proof The theorem follows straightforwardly from theorem 3.1 of Li and Xie (1995), who state and prove that given $K + 1$ autocovariances forming a positive definite sequence, a Gaussian process achieves the minimum information I_{p-f} when it follows an AR(K) model. The theorem directly applies to an AR(K) auxiliary process with prediction error variance equal to σ_p^2 whose first $K + 1$ autocovariances are γ_{pk}, $k = 0, 1, \ldots, K$. ■

In the case when $p = 0$, the process with minimal mutual information belongs to the class of EXP(K) models of Bloomfield (1973), obtained by truncating to the lag K the Fourier expansion of the logarithm of the spectral density function (see Li and Xie, 1995).

Model selection by the minimum mutual information (MMI) principle is one of the applications of I_{p-f} traced by Li and Xie (1995). As a matter of fact, MMI is related to the maximum entropy principle, by means of the complementarity of I_{p-f} and the entropy rate of a regular stationary and Gaussian process (see equation (9) in Li, 2005). Once the class of fitting models is selected through the MMI, a penalised criterion based on I_{p-f}, called LIC, has been proposed by Li and Xie (1995) to select the order of the fitting model.

Other applications of the mutual information between past and future arise in connection with the equivalent concept of excess entropy, an intuitive measure of memory stored in a stochastic process: fractional ARMA processes have infinite mutual information. Recently, Hidden Markov models with infinite mutual information between past and future have been investigated and examples of processes with an infinite number of countable states are presented with reference with modelling texts in natural language and complex system analysis (Debowski, 2014).

5 Estimation

The main tool for estimating the spectral density function and its functionals is the periodogram.

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

The periodogram (sample spectrum) is defined as

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

where $\bar{x} = n^{-1}\sum_t x_t$.

The following large sample distributional result holds in the short memory case (Brockwell and Davis, 1991, ch. 10):

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

whereas $\frac{I(\omega_j)}{f(\omega_j)} \sim \chi^2_1, \omega_j = 0, \pi$, where χ^2_r denotes a chi-square random variable with r degrees of freedom; equivalently, $I(\omega_j)$ is exponentially distributed with mean $f(\omega_j)$.

For a given transformation parameter p, the log-likelihood function of the unconstrained parameters ϑ_{pk}, $k = 1, \ldots, K$ based on the exponential density for $I(\omega_j)$, for $0 < \omega_j < \pi$ and $j = 1, \ldots, N$ where $N = [(n-1)/2]$, is

$$\ell(\vartheta_{p,K}) = -\sum_{j=1}^{N}\left(\ln f(\omega_j) + \frac{I(\omega_j)}{f(\omega_j)}\right)$$

The latter can be maximised with respect to the unconstrained parameter vector $\vartheta_{p,K}$ by a quasi-Newton optimisation algorithm, using numerical first and second derivatives. The GPAC π_{pk} and the parameters $\phi_{p1}, \ldots \phi_{pK}$ lying in the stationary region are then obtained by (7).

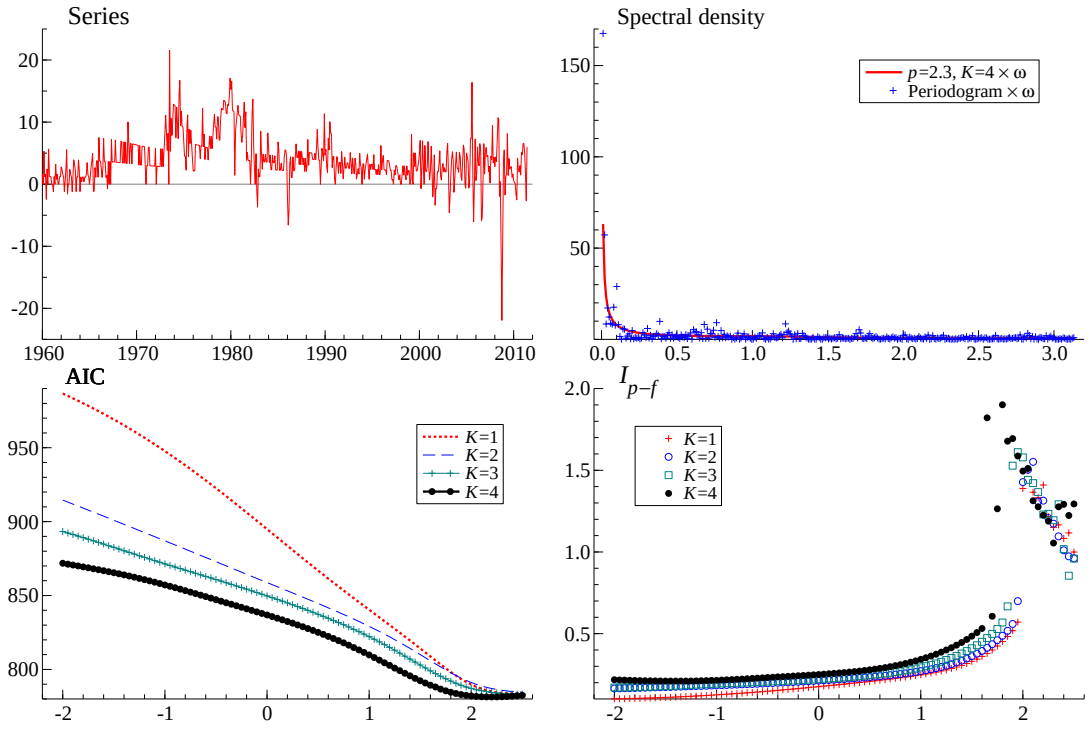
The spectral models for different combinations of K and p can be compared using an information criterion, such as AIC or BIC.

As long as the summability conditions for the GPAC is satisfied, $\sum_{k=1}^{\infty} k\ln(1-\pi^2_{pk}) < \infty$, i.e. as long as the mutual information as in theorem 1 is finite, the strong Szegö theorem holds (see Bingham, 2012) and under the usual regularity conditions on the parameter space and the spectral density function, theorems 2.1 and 2.2 on chapter 2 of Dzhaparidze (1986) hold, which imply that $\tilde{\vartheta}_{p,K} \to_p \vartheta_{p,K}$ and $\sqrt{n}(\tilde{\vartheta}_{p,K} - \vartheta_{p,K}) \to_d N(0, V_p)$ where

$$V_p^{-1} = \frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{1}{[2\pi f(\omega)]^{2p}}z(\omega)z(\omega)'\mathrm{d}\omega$$

with $z(\omega) = [1, 2\cos(\omega), 2\cos(2\omega), \ldots, 2\cos(K\omega)]'$. Given that the function that maps the partial autocorrelation coefficients to the model parameters is one to one and smooth (see Barndorff-Nielsen and Schou, 1973, Theorem 2), the asymptotic properties of the Whittle estimator continue to hold in the reparameterized model.

Figure 1: U.S. monthly inflation rate (annualised rate). Plot of the series (top left). Periodogram and fitted GAR spectrum with $p = 2.4, K = 1$ (top right). Values of the AIC as a function of p for values of $K = 1, 2, 3, 4$ (bottom left). Mutual information estimates as a function of p for different values of K (bottom right).



6 Illustration

Our illustration deals with the estimation of the mutual information between past and future for the monthly U.S. inflation rate. The latter is computed as the logarithmic change over the previous month of the Consumer Price Index (CPI), multiplied by 1200, and is considered for the period running from January 1960 to December 2012, for a total of 624 observations. The plot of the series is available in the top right panel of figure 1. A widely debated issue deals with the covariance stationarity of the underlying generating process, as the evidence arising from either unit root and stationarity tests is not clear cut. The periodogram, displayed in the right top plot of figure 1, takes very large values around the zero frequency.

The left bottom graph displays the behaviour of the Akaike's information criterion, AIC $= -2\ell(\tilde{\vartheta}_{p,K}) + 2K$, where $\tilde{\vartheta}_{p,K}$ is the Whittle likelihood estimator of the parameters of the generalised AR spectral model (6) fitted to the series for different values of p and K. In particular, for K ranging from 1 to 4, it displays

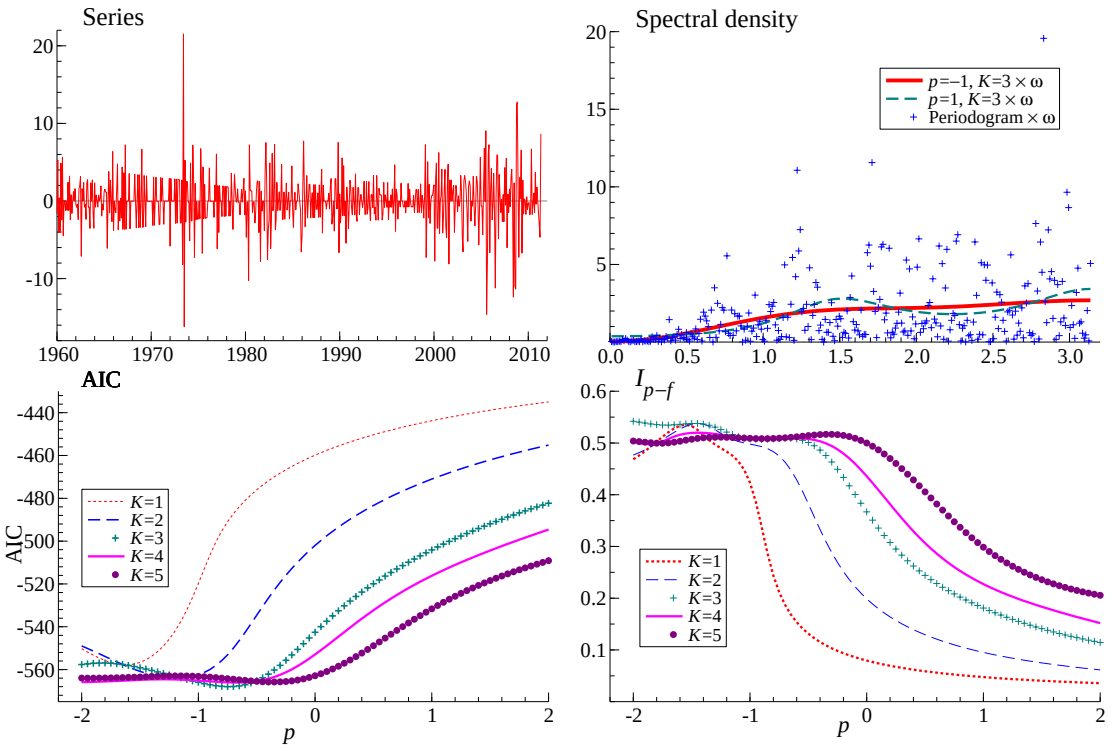
the pattern of the AIC for values of p in the interval [-2, 2.5]. The AIC is similar for the values of K considered for p in the range [2, 2.5]; when $K = 1$, the estimated generalised autoregressive parameter is equal to 1 for p in this range; similarly, for higher values of K, the generalised polynomial $\phi_p(z)$ has a unit root. The combination of values minimising the AIC is $K = 4$, $p = 2.3$; the estimated GAR polynomial is $\tilde{\phi}_p(e^{-\imath\omega}) = 1 - 1.0383e^{-\imath\omega} + 0.2028e^{-2\imath\omega} + 0.0251e^{-3\imath\omega} - 0.1896e^{-4\imath\omega}$; the corresponding GPACs are $\tilde{\pi}_{p1} = 0.9999, \tilde{\pi}_{p2} = -0.0612, \tilde{\pi}_{p3} = 0.1782, \tilde{\pi}_{p4} = 0.1896$. Notice that if $\phi_p(e^{-\imath\omega}) = (1 - e^{-\imath\omega})\phi_p^*(e^{-\imath\omega}))$, where $|\phi_p^*(e^{-\imath\omega})|^2$ is bounded away from zero and finite, and $p > 2$, y_t is long memory process, with a spectral density being $O(|\omega|^{-2d})$ as $|\omega| \to 0$, with $d = 2/p$. The fitted spectral density is superimposed to the periodogram in the top left panel. As a result, likelihood inference points decisively towards a nonstationary or a long memory model. Although the AIC is concave in p and has a minimum within the assumed range, the properties of the estimate of p cannot be ascertained. The bottom right graph displays the pattern of the estimated I_{p-f} as p varies in its range and for the same values of K. The plot highlights a neat discontinuity occurring at around $p = 2$, which is a reflection of the fact that the estimated model is long memory.

We then turn our attention to the first differences of the monthly inflation rates, which are plotted in figure in the top left panel of figure 2. Model selection performed according to the AIC leads to choosing the GAR specification with $p = -0.8$ and $K = 3$ (see the bottom left plot which depicts the AIC as a function of p for $K = 1, 2, 3, 4, 5$). The selected values of p is not significantly different from -1, which in turn leads to a moving average model for the spectrum of the time series. The fitted spectral density has a low dynamic range and is superimposed to the periodogram of the series in the top right plot. It can be seen from the same plot that the spectral fit arising from the AR spectrum of the same order, i.e. $p = 1, K = 3$, is likely to overfit some features of the periodogram. The estimated GAR polynomial is $\tilde{\phi}_p(e^{-\imath\omega}) = 1 - 0.5419e^{-\imath\omega} - 0.1989e^{-2\imath\omega} - 0.0960e^{-3\imath\omega}$; the corresponding GPACs are $\tilde{\pi}_{p1} = 0.7583, \tilde{\pi}_{p2} = 0.2532, \tilde{\pi}_{p3} = 0.0960$.

The behaviour of the estimated mutual information between past and future is displayed in the bottom right panel as p varies in [-2, 2] and for $K = 1, \ldots, 5$. The estimates are rather insensitive to p in the negative range (in fact, also the AIC is almost constant), hovering around 0.5.

We conclude this section with a remark on the estimation of I_{p-f} based on the GPAC. When a generalised spectral AR model is fitted to a series, the selection the best pair (p, K) in terms of minimum AIC implies that the estimation of I_{p-f} is based on a *finite* number of generalised partial autocorrelation coefficients, $\tilde{\pi}_{p1}, \ldots, \tilde{\pi}_{pK}$. Except in the case when $p = 1$, which identifies a pure autoregressive model in the general class (6), the computation of I_{p-f} based on the ordinary partial autocorrelations or cepstral coefficients involves an infinite number of coefficients and its estimation necessarily requires a truncation of the infinite sequence.

Figure 2: Changes in U.S. monthly inflation rate (first differences). Plot of the series (top left). Periodogram and fitted GAR spectrum with $p = 2.4, K = 1$ (top right). Values of the AIC as a function of p for values of $K = 1, 2, 3, 4, 5$ (bottom left). Mutual information estimates as a function of p for different values of K (bottom right).



7 Conclusions

In this paper we have introduced the generalised partial autocorrelation function. This can be useful for parameterising a model for the spectrum of a time series, which encompasses autoregressive and moving average spectral estimation, and for characterising the properties of a random stationary process. We have considered in particular its use for computing (and estimating from a sample time series) the mutual information between the past and the future. The latter is an interesting concept which complements the traditional one step ahead prediction error variance, looking at the joint predictability at the different many horizons that arise for the future.

References

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

[2] Battaglia, F.: Inverse Autocovariances and a Measure of Linear Determinism For a Stationary Process. Journal of Time Series Analysis. **4**, 79–87 (1983)

[3] Bingham, N.H.: Szegö Theorems and its Probabilistic Descendants. Probability Surveys **9**, 287–324 (2012)

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

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

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

[7] Debowski, L.: On Hidden Markov Processes with Infinite Excess Entropy. Journal of Theoretical Probability, **27**, 539–551 (2014)

[8] Dzhaparidze, K.:Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series. Springer-Verlag, New York (1986)

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

[10] Ibragimov, I.A., Rozanov, Y.A.: Gaussian Random Processes. Springer-Verlag, New York (1978)

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

[12] Li, L.M., Xie, Z.: Model Selection and Order Determination for Time Series by Information between the Past and the Future. Journal of Time Series Analysis, **17**, 65–84 (1996)

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

[14] Proietti, T., Luati, A.: The Exponential Model for the Spectrum of a Time Series: Extensions and Applications. MPRA working paper 45280 (2014)

[15] Proietti, T., Luati, A.: The Generalised Autocovariance function. Journal of Econometrics, **187**, 245–257 (2015)

Research Papers
2013

2015-07: Peter Christoffersen, Bruno Feunou and Yoontae Jeon: Option Valuation with Observable Volatility and Jump Dynamics

2015-08: Alfonso Irarrazabal and Juan Carlos Parra-Alvarez: Time-varying disaster risk models: An empirical assessment of the Rietz-Barro hypothesis

2015-09: Daniela Osterrieder, Daniel Ventosa-Santaulària and Eduardo Vera-Valdés: Unbalanced Regressions and the Predictive Equation

2015-10: Laurent Callot, Mehmet Caner, Anders Bredahl Kock and Juan Andres Riquelme: Sharp Threshold Detection Based on Sup-norm Error rates in High-dimensional Models

2015-11: Arianna Agosto, Giuseppe Cavaliere, Dennis Kristensen and Anders Rahbek: Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX)

2015-12: Tommaso Proietti, Martyna Marczak and Gianluigi Mazzi: EuroMInd-D: A Density Estimate of Monthly Gross Domestic Product for the Euro Area

2015-13: Michel van der Wel, Sait R. Ozturk and Dick van Dijk: Dynamic Factor Models for the Volatility Surface

2015-14: Tim Bollerslev, Andrew J. Patton and Rogier Quaedvlieg: Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting

2015-15: Hossein Asgharian, Charlotte Christiansen and Ai Jun Hou: Effects of Macroeconomic Uncertainty upon the Stock and Bond Markets

2015-16: Markku Lanne, Mika Meitz and Pentti Saikkonen: Identification and estimation of non-Gaussian structural vector autoregressions

2015-17: Nicholas M. Kiefer and C. Erik Larson: Counting Processes for Retail Default Modeling

2015-18: Peter Reinhard Hansen: A Martingale Decomposition of Discrete Markov Chains

2015-19: Peter Reinhard Hansen, Guillaume Horel, Asger Lunde and Ilya Archakov: A Markov Chain Estimator of Multivariate Volatility from High Frequency Data

2015-20: Henri Nyberg and Harri Pönkä: International Sign Predictability of Stock Returns: The Role of the United States

2015-21: Ulrich Hounyo and Bezirgen Veliyev: Validity of Edgeworth expansions for realized volatility estimators

2015-22: Girum D. Abate and Niels Haldrup: Space-time modeling of electricity spot prices

2015-23: Eric Hillebrand, Søren Johansen and Torben Schmith: Data revisions and the statistical relation of global mean sea-level and temperature

2015-24: Tommaso Proietti and Alessandra Luati: Generalised partial autocorrelations and the mutual information between past and future