Generalised partial autocorrelations and the mutual information between past and future

Alessandra Luati [1] Tommaso Proietti [2]

[1]University of Bologna

[2]University of Rome "Tor Vergata" and CREATES

Aarhus conference on Probability, Statistics and their Applications
Aarhus, 15–19 June 2015

Celebrating the scientific achievements of Ole Barndorff-Nielsen

Outline

1. Introduction
2. Generalised autocovariances, generalised autocorrelations and generalised partial autocorrelations
3. Generalised autoregressive spectral models
4. Estimation, based on Barndorff-Nielsen and Schou (1973)
5. Mutual information between past and future
6. Empirical application
7. Conclusive remarks and further research

Introduction

- Let $\{x_t\}_{t \in T}$ be a purely non deterministic stationary zero-mean process, with Wold representation

$$x_t = \psi(B)\xi_t,$$

$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \ldots$, infinite polynomial in the backshift operator B, $\xi_t \sim \mathrm{WN}(0, \sigma^2)$.

- Assume that the spectral density function $f(\omega)$ exists

$$F(\omega) = \int_{-\pi}^{\omega} f(\lambda) d\lambda,$$

that the process is regular and that the powers $f(\omega)^p$ exist and are integrable.

The generalised autocovariance function (GACV)

- The GACV is defined (Proietti and Luati, 2015) as

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

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

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

Motivation

The underlying idea, which has a well established tradition in statistics and time series analysis (Box and Cox, 1964), is that taking powers of the spectral density function allows one to emphasise certain features of the process and mute other features.

Applications

- White noise and goodness of fit tests (fractional p)
- Feature matching estimation of the spectrum ($p > 1$)
- Cluster and discriminant analysis ($p < 0$)

Interpretation of the GACV

Let us consider the auxiliary process u_{pt},

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

where, for real $p > 0$,

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

with coefficients given by the recursive relation

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

(see Gould, 1974). The same recursion holds for $p < 0$.

Interpretation of the GACV

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

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

- The GACV of x_t can be interpreted as the autocovariance function of the process u_{pt}, denoted as γ_u,

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

  and it is 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 and the variance profile

When $k = 0$, the generalised variance γ_{p0} is related to the variance profile, defined in Luati, Proietti and Reale (2012) as the Hölder mean of the spectrum of x_t:

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

In particular, for $p \neq 0$, $v_p = \gamma_{p0}^{\frac{1}{p}}$.

- $v_{-1} = \gamma_{-1,0}^{-1}$ 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.
- $\lim_{p\to 0} v_p = \sigma^2 = \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log 2\pi f(\omega) \mathrm{d}\omega \right\}$ (Szegö-Kolmogorov), is the prediction error variance.

Generalised partial autocorrelations

- The generalised autocorrelation function (GACF) of x_t is

$$\rho_{pk} = \frac{\gamma_{pk}}{\gamma_{p0}},$$

  $k = 0, \pm 1, \pm 2, \ldots,$ taking values in $[-1, 1]$.
- If the GACV of x_t is proportional to the autocovariance function of the auxiliary process u_{pt}, the GACF is equal to the autocorrelation function 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}.

Generalised autoregressive spectral models

The GPAC are central for estimating the following class of models,

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

where

$$\phi_p(e^{-\imath\omega}) = 1 - \phi_{p1}e^{-\imath\omega} - \phi_{p2}e^{-\imath\omega 2} - \cdots - \phi_{pK}e^{-\imath\omega K}.$$

Special cases

→ AR(K) spectral models, $p = 1$

→ MA(K) case, $p = -1$

→ fractional case, e.g. $K = 1$, $p = 1/d$ and $\phi_{p1} = 1$.

According to this parameterisation, the GPAC form a finite sequence.

The periodogram and the Whittle likelihood

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 $\bar{x} = n^{-1} \sum_t x_t$ and $\omega_j = \frac{2\pi j}{n}$, $j = 1, \ldots, [n/2]$.

Asymptotically, short range processes,

$$\frac{I(\omega_j)}{f(\omega_j)} \sim \text{IID} \frac{1}{2} \chi_2^2, \quad 0 < \omega_j < \pi$$

For a given transformation parameter p, the log-likelihood function of unconstrained parameters ϑ_{pk} is $k = 1, \ldots, K$ is

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

$j = 1, \ldots, N$ where $N = [(n-1)/2]$.

Reparameterisation

Solution Reparameterise the AR coefficients in terms of partial autocorrelations (Barndorff-Nielsen and Schou, 1973).

Letting $\pi_{pk}, k = 1, \ldots, K$, $|\pi_{pk}| < 1$, compute, for $j = 1, \ldots, k-1$,

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

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

The coefficients π_{pk}, 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 $\vartheta_{p0} = \ln(\sigma_p^2)$.

Mutual information

Let $\{x_t\}_{t\in T}$ and $\{y_s\}_{s\in S}$ on $(\Omega, \mathcal{F}, P)$, and $\mathcal{S}_1$ and $\mathcal{S}_2$ minimal sigma-algebra.

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),

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

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 $j = 1, \dots, m$.

Properties

- $I(x, y) \geq 0$, with $I(x, y) = 0$ when $\mathcal{S}_1 \perp\!\!\!\perp \mathcal{S}_2$
- $I(x, y) = I(y, x)$, which motivates the name of mutual information
- Information regularity coefficient

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

- $I_0 = I_{p-f}$ is the mutual information between past and future
- Reflectrum identity
- Relation with the generalised partial autocorrleation coefficients

Reflectrum identity

- For Gaussian processes (Li, 2005)

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

where c_k are the cepstral coefficients of the process,

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

- Necessary condition for information regularity: $\sum_{k=1}^{\infty} k c_k^2 < \infty$.
- Reflectrum identity, π_k partial autocorrelation coefficients

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

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.

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

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

and the equality holds for all p.

Why estimating the mutual information by the GPAC?

- 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 with respect to the minimum mutual information principle.

Theorem 2

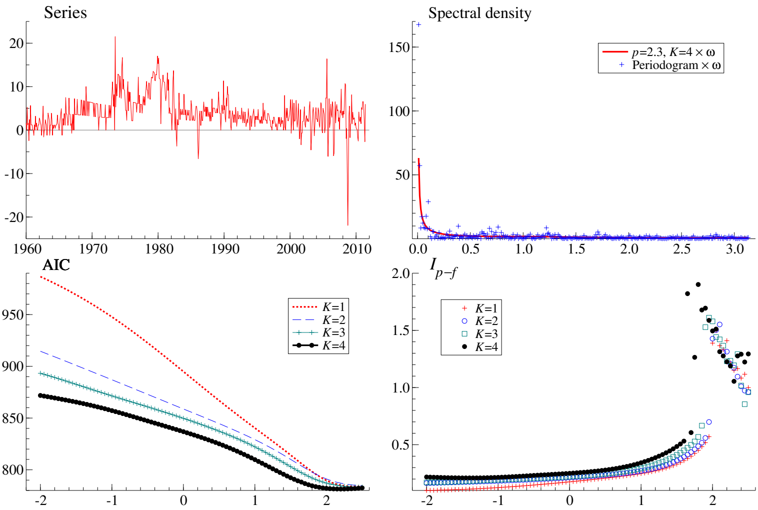
A process with given generalised autocovariances $\gamma_{pk}, k = 0, 1, \ldots, K$, $p \neq 0$ *and minimal information between past and future belongs to the class of generalised autoregressive spectral models.*

Illustration

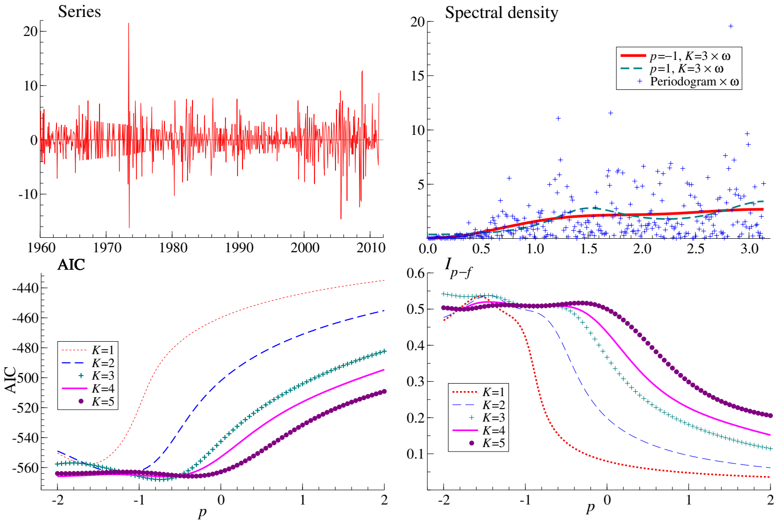
Our illustration deals with the estimation of the mutual information between past and future for the monthly U.S. inflation rate.

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.

U.S. monthly inflation rate



U.S. monthly inflation rate, first differences



Concluding remarks

- A class of models for estimating the spectrum of a stationary process has been introduced encompassing AR ($p = 1$) and MA ($p = -1$) estimation of the spectrum. The class is optimal in the sense of minimal mutual information.
- The models are estimated via the Whittle likelihood based on a reparameterisation of the (generalised) AR coefficients based on the (generalised) partial autocorrelation, due to Barndorff-Nielsen and Schou (1973).
- A relation has been derived between the mutual information between past and future of a Gaussian process, I_{p-f}, and the generalised partial autocorrelation coefficients, π_{pk}, enabling estimation of I_{p-f} based on a finite sequence of π_{pk}.

Further research

- General linear models for the spectrum of a time series: consider the Box-Cox transform and get Bloomfield (1973) exponential model as a particular case
- Extension to the locally stationary case: a dynamic, logistic smooth transition is assumed for the coefficients of the Fourier expansion of the Box-Cox transform of the spectrum. Estimation is carried out based on the pre-periodogram
- Multivariate extension