The Generalised Autocovariance Function

Tommaso, Proietti and Alessandra, Luati

6 June 2012

The Generalised Autocovariance Function

Tommaso Proietti
Department of Economics and Finance
University of Rome Tor Vergata, Italy

Alessandra Luati
Department of Statistics
University of Bologna, Italy

Abstract

The generalised autocovariance function is defined for a stationary stochastic process as the inverse Fourier transform of the power transformation of the spectral density function. Depending on the value of the transformation parameter, this function nests the inverse and the traditional autocovariance functions. A frequency domain non-parametric estimator based on the power transformation of the pooled periodogram is considered and its asymptotic distribution is derived. The results are employed to construct classes of tests of the white noise hypothesis, for clustering and discrimination of stochastic processes and to introduce a novel feature matching estimator of the spectrum.

Keywords: Stationary Gaussian processes. Non-parametric spectral estimation. White noise tests. Feature matching. Discriminant Analysis.

1 Introduction

The temporal dependence structure of a stationary stochastic process is characterised by the autocovariance function, or equivalently by its Fourier transform, the spectral density function. We generalise this important concept, by introducing the generalised autocovariance function (GACV), which we define as the inverse Fourier transform of the p-th power of the spectral density function, where p is a real parameter. The GACV depends on two arguments, the power parameter p and the lag k. Dividing by the GACV at lag zero for p given yields the generalised autocorrelation function (GACF).

For $k = 0$ the GACV is related to the variance profile, introduced by Luati, Proietti and Reale (2012) as the Hölder mean of the spectrum. For $p = 1$, it coincides with the traditional autocovariance function, whereas for $p = -1$ it yields the inverse autocovariance function, as k varies. The extension to any real power parameter p is fruitful for many aspects of econometrics and time series analysis. We focus in particular on model identification, time series clustering and discriminant analysis, the estimation of the spectrum for cyclical time series, and on testing the white noise hypothesis and goodness of fit.

The underlying idea, which has a well established tradition in statistics and time series analysis (Tukey, 1957, Box and Cox, 1964), is that taking powers of the spectral density function allows one to emphasise certain features of the process. For instance, we illustrate that setting $p > 1$ is useful for the identification of spectral peaks, and in general for the extraction of signals contaminated by noise. Moreover, fractional values of $p \in (0, 1)$ enable the definition of classes of white noise tests with improved size and power properties, with respect to the case $p = 1$, as the finite sample distribution can be made closer to the limiting one by the transformation that is implicit in the use of the GACV.

For given stochastic processes the GACV can be analytically evaluated in closed form in the time domain by constructing the standard autocovariance function of an auxiliary stochastic process, whose Wold representation is obtained from the original one, by taking a power transformation of the Wold polynomial.

As far as estimation from a time series realisation is concerned, we consider a nonparametric estimator based on the power transformation of the pooled periodogram. For a given p, the estimator is asymptotically normally distributed around the population value, with a variance that depends on the GACV evaluated at $2p$; as a result, a consistent estimator of the asymptotic variance is readily available. We also show that Bartlett's formula generalises to any value of p. As a related result we derive the asymptotic distribution of a ratio estimator of the GACF.

These results open the way to the application of the GACV for the analysis of stationary time series. In addition to the possible uses hinted above (model identification, testing for white noise, and feature extraction), we consider the possibility of defining measures of pairwise distance based on the GACV or GACF, encompassing the Euclidean and the Hellinger distances, and we illustrate their use for discriminant and cluster analysis of time series. Negative values can be relevant as they nest the Euclidean and the Hellinger distances based on the inverse autocorrelation functions.

The structure of the paper is the following. The GACV and the GACF are formally defined in section 2. The interpretation in terms of the autocovariance function of a suitably defined power-transformed process is provided in section 3. This is used for the analytical derivation of the GACV for first order autoregressive (AR) and moving average processes, as well as long memory processes (section 4). Estimation is discussed in section 5. Sections 6-8 focus on three main uses of the GACV and the GACF. The first deals with testing for white noise: two classes of tests, generalising the Box and Pierce (1970) test and the Milhøj (1981) statistics, are proposed and their properties discussed. A Yule-Walker estimator of the spectrum based on the GACV is presented in section 7: in particular, the GACV for $p > 1$ will highlight the cyclical features of the series; this property can be exploited for the identification and estimation of spectral peaks. We finally consider measures of distance between two stochastic processes based on the GACV or GACF and we illustrate their use for time series discriminant analysis. In section 9 we provide some conclusions and directions for future research.

2 The generalised autocovariance function

Let $\{x_t\}_{t\in T}$ be a stationary zero-mean stochastic process indexed by a discrete time set T, with spectral distribution function $F(\omega)$. We assume that the spectral density function of the process exists, $F(\omega) = \int_{-\pi}^{\omega} f(\lambda)\mathrm{d}\lambda$, and that the process is regular (Doob, 1953, p. 564), i.e. $\int_{-\pi}^{\pi} \log f(\omega)\mathrm{d}\omega > -\infty$. We further assume that the powers $f(\omega)^p$ exist, are integrable with respect to $\mathrm{d}\omega$ and bounded for p in (a subset of) the real line.

The generalised autocovariance (GACV) function is defined as the inverse Fourier transform of the p-th power of the spectral density function,

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

where we have replaced $\exp(\imath\omega k)$ by $\cos(\omega k)$ since the spectral density and the cosine are even

functions while the sine function is odd. Taking the Fourier transform of γ_{pk} gives

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

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).

The GACV satisfies all the properties of an autocovariance function: an obvious property is $\gamma_{pk} = \gamma_{p,-k}$; moreover, $\gamma_{p0} > 0$ and $|\gamma_{pk}| \leq \gamma_{p0}$, for all integers k. Non-negative definiteness of the GACV follows from the assumptions on $f(\omega)$. These properties enable to define the generalised autocorrelation function (GACF) as

$$\rho_{pk} = \frac{\gamma_{pk}}{\gamma_{p0}}, \quad k = 0, \pm 1, \pm 2, \ldots, \tag{3}$$

taking values in $[-1, 1]$.

Other relevant properties are nested in the following lemma, which is a consequence of the fact that the spectral density of a convolution is the product of the spectral densities (see corollary 3.4.1.1. in Fuller, 1996).

Lemma 1 *Let γ_{pk} be defined as in* (1) *and* (2). *Then,*

$$\sum_{j=-\infty}^{\infty} \gamma_{p,j+k}\gamma_{q,j+l} = \frac{1}{2\pi}\int_{-\pi}^{\pi} [2\pi f(\omega)]^{p+q} \cos(\omega(k-l))\mathrm{d}\omega. \tag{4}$$

An important special case of Lemma 1, that will be exploited later in the derivation of goodness of fit tests, relates the GACV with transformation parameter $2p$ to the GACV at p and is obtained by setting $p = q$ and $l = 0$ in lemma 1:

$$\gamma_{2p,k} = \sum_{j=-\infty}^{\infty} \gamma_{pj}\gamma_{p,j+k}, \tag{5}$$

which for $k = 0$ specialises as

$$\gamma_{2p,0} = \gamma_{p0}^2 + 2\sum_{j=1}^{\infty} \gamma_{pj}^2.$$

Furthermore, setting $q = -p$ and $l = 0$ in lemma 1 we obtain

$$\sum_{j=-\infty}^{\infty} \gamma_{pj}\gamma_{-p,j-k} = 1_{k=0}, \tag{6}$$

where 1_A indicates the indicator function on the set A. Property (6) extends the well known orthogonality between the autocovariance function and the inverse autocovariance function (see Pourahmadi, 2001, theorem 8.12).

3 The power process and its autocovariance function

The function γ_{pk} lends itself to a further interpretation as the autocovariance function of a power 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.

The Wold representation of $\{x_t\}_{t\in T}$ will be written as

$$x_t = \psi(B)\xi_t, \tag{7}$$

where $\xi_t \sim \text{IID}(0, \sigma^2)$ and $\psi(B) = 1+\psi_1 B+\psi_2 B^2+\ldots$, with coefficients satisfying $\sum_{j=0}^{\infty}|\psi_j| < \infty$, and such that all the roots of the characteristic equation $\psi(B) = 0$ are in modulus greater than one; B is the backshift operator, $B^k x_t = x_{t-k}$ and IID stands for independent and identically distributed. The autocovariance function of the linear process (7) is $\gamma_k = \sigma^2 \sum_{j=0}\psi_j\psi_{j+k}$ for $k = 0, 1, \ldots$ and $\gamma_{-k} = \gamma_k$.

Let us consider the power-transformed process:

$$u_{pt} = \begin{cases} \psi(B)^p\xi_t &= \psi(B)^p\psi(B)^{-1}x_t, \quad \text{for } p \geq 0 \\ \psi(B^{-1})^p\xi_t &= \psi(B^{-1})^p\psi(B)^{-1}x_t, \quad \text{for } p < 0. \end{cases} \tag{8}$$

For arbitrary p, the power of $\psi(B)$ in (8) is still a power series,

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

with coefficients given by the recursive relation

$$\varphi_j = \frac{1}{j}\sum_{k=1}^{j}[k(p+1)-j]\psi_k\varphi_{j-k}, j > 0, \;\; \varphi_0 = 1 \tag{9}$$

(see Gould, 1974). In most practical applications, a finite version or approximation of $\psi(B)$ can be considered, say a q dimensional polynomial $\psi_q(B)$, with q roots $-\zeta_1^{-1}, -\zeta_2^{-1}, \ldots, -\zeta_q^{-1}$ lying outside the unit circle to ensure invertibility. Hence, $\psi_q(B)^p = (1+\zeta_1 B)^p(1+\zeta_2 B)^p \ldots (1+\zeta_q B)^p$,

where each factor can be expanded using the binomial theorem holding for $p \in \mathbb{R}$ and $\zeta_i \in \mathbb{C}$, $(1+\zeta_i B)^p = \sum_{k=0}^{\infty} \binom{p}{k} (\zeta_i B)^k$, where

$$\binom{p}{k} = \frac{p(p-1)(p-2)\dots(p-k+1)}{k(k-1)(k-2)\dots 1} \tag{10}$$

with initial conditions $\binom{p}{0} = 1$, $\binom{p}{1} = p$, and where absolute convergence is implied by invertibility (see Graham, Knuth and Patashnik, 1994, ch. 5).

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

$$2\pi f_u(\omega)(\sigma^2)^{p-1} = [2\pi f(\omega)]^p \,. \tag{11}$$

It follows from (1) and (11) that $(\sigma^2)^{1-p}\gamma_{pk}$ is the autocovariance function of the power process u_{pt}.

The 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:

$$v_p = \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} [2\pi f(\omega)]^p \right\}^{\frac{1}{p}} . \tag{12}$$

In particular, 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 by the Szegö-Kolmogorov formula is the geometric average of the spectral density, $\sigma^2 = \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log 2\pi f(\omega) \mathrm{d}\omega \right\}$.

4 Illustrations

4.1 The generalised autocovariance function of AR(1) and MA(1) processes

Let us consider the stationary AR(1) process $x_t = (1-\phi B)^{-1}\xi_t$, $|\phi| < 1$, $\xi_t \sim \mathrm{WN}(0, \sigma^2)$. The generalised autocovariance function of this process is given by

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

The power process associated with $x_t \sim$AR(1) is $u_{pt} = (1-\phi B)^{-p}\xi_t$. Given that, in the present case, $\psi_0 = 1, \psi_1 = -\phi, \psi_k = 0$ for $k > 1$, the recursive formula (9) becomes $\varphi_j = \frac{1}{j}(-p+1-j)(-\phi)\varphi_{j-1}$ and thus we obtain $\varphi_j = \frac{(-\phi)^j}{j!}(-p)(-p-1)(-p-2)\dots(-p-j+1) = (-\phi)^j \binom{-p}{j}$,

see equation (10) and note that for $p = 0$, $\psi_j = 0$ for all j, since $\binom{0}{j} = 0$. The GACV of x_t is $(\sigma^2)^{p-1}$ times the autocovariance of the process u_{pt} and therefore

$$\gamma_{pk} = \sigma^{2p}(-\phi)^k \sum_{j=0}^{\infty} (-\phi)^{2j} \binom{-p}{j} \binom{-p}{j+k} \tag{13}$$

with

$$\gamma_{p0} = \sigma^{2p} \sum_{j=0}^{\infty} \binom{p+j-1}{k}^2 \phi^{2j},$$

where we have applied the basic identity, $\binom{p}{j} = (-1)^j \binom{-p+j-1}{k}$. Straightforward algebra allows us to verify that for $p = 1$, $\gamma_{1k} = \sigma^2 \frac{\phi^k}{1-\phi^2}$.

The GACF is

$$\rho_{pk} = (-\phi)^k \frac{\sum_{j=0}^{\infty} (-\phi)^{2j} \binom{-p}{j}^2 \left((j+1)(j+2)\dots(j+k)\right)^{-1}}{\sum_{j=0}^{\infty} (-\phi)^{2j} \binom{-p}{j}^2}.$$

Similarly to the AR(1) case, for the invertible MA(1) process $x_t = (1 - \theta B)\xi_t$, $|\theta| < 1$, $\xi_t \sim \mathrm{WN}(0, \sigma^2)$, with associated power process $u_t = (1 - \theta B)^p \xi_t$, we find:

$$\gamma_{pk} = \sigma^{2p}(-\theta)^k \sum_{j=0}^{\infty} (-\theta)^{2j} \binom{p}{j} \binom{p}{j+k} \tag{14}$$

and

$$\gamma_{p0} = \sigma^{2p} \sum_{j=0}^{\infty} (-\theta)^{2j} \binom{p}{j}^2.$$

For $p = 1$, binomial coefficients of the form $\binom{1}{j}$ are involved, which are null whenever $j > 1$ and therefore it is immediate to see that $\gamma_{10} = \sigma^2(1 + \theta^2)$ and $\gamma_{11} = -\sigma^2\theta$ while $\gamma_{1k} = 0$ for $k > 1$, as expected.

In general, for integer $p > 0$, the GACV of an MA(1) process has a cutoff point at $k = p$. As an example, let us consider the case of a square transformation, that is $p = 2$, for which:

$$\begin{aligned}
\gamma_{20} &= \sigma^4(1 + 4\theta^2 + \theta^4), \\
\gamma_{21} &= \sigma^4(-\theta)(2 + 2\theta^2), \\
\gamma_{22} &= \sigma^4\theta^2, \\
\gamma_{2k} &= 0, k > 2.
\end{aligned}$$

Equations (13) and (14) generalise to any fractional p equations 3.616.7 and 3.616.4 of Gradshteyn and Ryzhik (1994) that hold for AR(1) and MA(1) processes in the case of a positive integer power p .

4.2 Long memory processes

For the fractional noise (FN) process, $(1-B)^d x_t = \xi_t$, where $\xi_t \sim \mathrm{WN}(0,\sigma^2)$, $d < 0.5$, the GACV and GACF are defined for $pd < 0.5$ and are given respectively by

$$\gamma_{pk} = \sigma^{2p}\frac{\Gamma(1-2dp)\Gamma(k+dp)}{\Gamma(1-dp)\Gamma(dp)\Gamma(1+k-dp)}, \quad \rho_{pk} = \frac{\Gamma(1-dp)\Gamma(k+dp)}{\Gamma(1+k-dp)\Gamma(dp)}.$$

This is easily established from the autocovariance of u_{pt}, which is a FN process with memory parameter pd. For $p = -1/d$, $\rho_{p1} = -0.5, \rho_{pk} = 0, k = 2, \ldots,$ as u_{pt} ha a non-invertible MA(1) representation.

Let us consider the Gegenbauer process

$$x_t = (1 - 2\nu B + B^2)^{-d}\xi_t,$$

where $\xi_t \sim \mathrm{WN}(0,\sigma^2)$; $\nu = \cos\lambda$, determines the frequency at which a long-memory behavior occurs. The process is stationary for $d < 0.5$ if $|\nu| < 1$ and for $d < 1/4$ for $\nu = \pm 1$. See Gray, Zhang, and Woodward (1989) for further details. The Wold representation of the process x_t is obtained from the series expansion of the Gegenbauer polynomial (Erdélyi et.al, 1953, 10.9),

$$x_t = \sum_{j=0}^{\infty} G_j(\nu,d)B^j\xi_{t-j},$$

with coefficients

$$G_j(\nu,d) = \sum_{q=0}^{[j/2]} \frac{(-1)^q(2\nu)^{j-2q}\Gamma(d-q+j)}{q!(j-2q)!\Gamma(d)}$$

that are derived from the recursive formula:

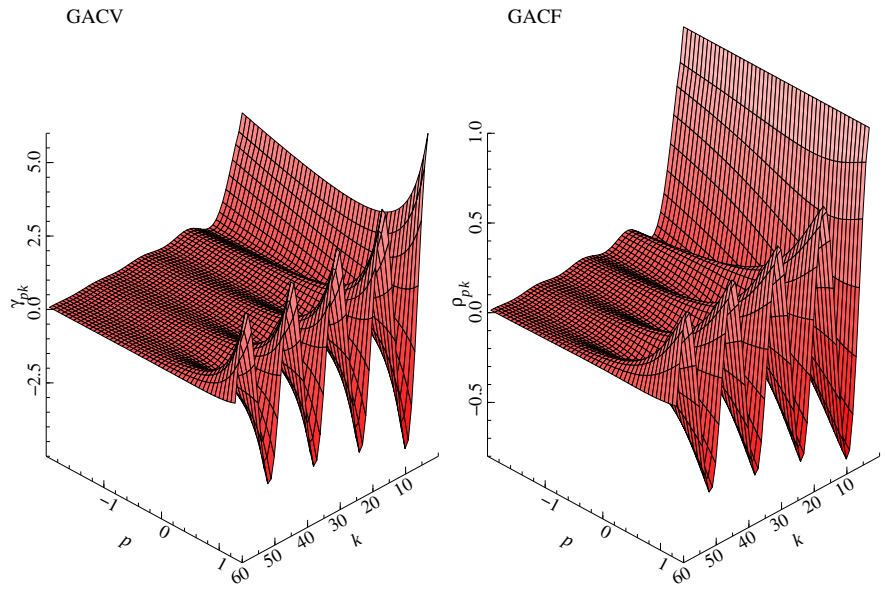
$$G_j(\nu,d) = 2\nu\left(\frac{d-1}{j}+1\right)G_{j-1}(\nu,d) - \left(2\frac{d-1}{j}+1\right)G_{j-2}(\nu,d),$$

with initial conditions $G_0(\nu,d) = 1$ and $G_1(\nu,d) = 2d\nu$. Hence, provided that $pd < 0.5$, the generalised autocovariance function of x_t for $p \neq 0$ is given by

$$\gamma_{pk} = \sigma^{2p}\sum_{j=0}^{\infty} G_j(\nu,dp)G_{j+k}(\nu,dp).$$

For $p = 1$, the above series is the autocovariance function of the Gegenbauer process and it is known that it can converge very slowly and several techniques have been implemented with the aim of increasing the rate of convergence (see Woodward, Cheng and Gray, 1998, and references therein). The generalised autocovariance function overcomes the problem, since for values of

Figure 1: Generalised autocovariances and autocorrelations for the Gegenbauer process $x_t = (1 - 2\nu B + B^2)^{-d}\xi_t, \xi_t \sim \mathrm{WN}(0, \sigma^2)$ with $d = 0.4, \nu = 0.9, \sigma^2 = 1$.



$p < 0.5/d$, it converges at a faster rate than $p = 1$. Figure 1 illustrates the behavior of the GACV and GACF of a Gegenbauer process with $\nu = 0.9$ and $d = 0.4$, for different values of $p < 1.25$ and $k = 1, 2, \ldots, 60$. For $p = -1/d$, $\rho_{pk} = 0, k = 3, \ldots,$ as u_{pt} ha a non-invertible MA(2) representation.

5 Estimation

We shall consider a nonparametric estimator of the generalised autocovariance function based on the periodogram of $(x_1, x_{,2}, \ldots, x_n)$,

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

evaluated at the Fourier frequencies $\omega_j = \frac{2\pi j}{n} \in (0, \pi)$, $1 < j < [n/2]$.

Specifically, estimation of γ_{pk}, as defined in (1), will be based on the nonparametric estimator

defined as follows. Let $M = [(n-1)/(2m)]$, then

$$\hat{\gamma}_{pk} = \frac{1}{M} \sum_{j=0}^{M-1} Y_j^{(p)} \cos(\bar{\omega}_j k), \tag{15}$$

where

$$Y_j^{(p)} = \left(2\pi \bar{I}_j\right)^p \frac{\Gamma(m)}{\Gamma(m+p)}$$

and

$$\bar{I}_j = \sum_{l=1}^{m} I(\omega_{jm+l}),$$

is the pooled periodogram over m non overlapping contiguous frequencies, whereas

$$\bar{\omega}_j = \omega_{jm+(m+1)/2}$$

are the mid range frequencies.

The estimator (15) is constructed based on the same principles of the variance profile estimator considered by Luati, Proietti and Reale (2012), which is an extension of Hannan and Nicholls (1977) frequency domain estimator of the prediction error variance, which, in turn, generalised the Davis and Jones (1968) estimator based on the raw periodogram. For simplicity of exposition, we have ruled out from estimation the frequencies 0 and π, which require a special treatment, as the asymptotic theory based on the periodogram ordinates is slightly different in 0 and π. The latter can be included without substantially modifying the estimator, see the discussion in Hannan and Nicholls (1977).

The factor $\frac{\Gamma(m)}{\Gamma(m+p)}$ serves to correct for the asymptotic bias, $\mathrm{E}\left(Y_j^{(p)}\right) = (2\pi f(\bar{\omega}_j))^p$, and pooling is required since the bias correction term exists only for $p > -m$, that for $p = -1$ requires $m > 1$. Furthermore, we shall prove that the asymptotic variance of the estimator (15) exists only for $p > -\frac{m}{2}$. The underlying assumption is that the spectral density is constant over frequency intervals of length $\frac{2\pi m}{M}$. Notice that in the definition of $\hat{\gamma}_{pk}$ the dependence on m is implicitly considered. The asymptotic properties of the estimator (15) are established by the following theorem.

Theorem 1 *Let $\{x_t\}_{t\in T}$ be the process $x_t = \sum_{j=0}^{\infty} \psi_j \xi_{t-j}$ where $\xi_t \sim NID(0, \sigma^2)$, $\sum_{j=0}^{\infty} |\psi_j| < \infty$, $\sum_{j=0}^{\infty} |\psi_j||j|^{\frac{1}{2}+\delta} < \infty$, $\delta > 0$, and with absolutely continuous spectral density function $f(\omega)$, whose powers $f(\omega)^p$ are integrable and uniformly bounded. Let us denote the vector of generalised autocovariance functions up to lag K as $\boldsymbol{\gamma}_p' = [\gamma_{p0}, \gamma_{p1}, \ldots, \gamma_{pK}]'$ and the corresponding estimator with elements given by (15) as $\hat{\boldsymbol{\gamma}}_p' = [\hat{\gamma}_{p0}, \hat{\gamma}_{p1}, \ldots, \hat{\gamma}_{pK}]'$. Then, $\hat{\boldsymbol{\gamma}}_p \to_p \boldsymbol{\gamma}$ and*

$$\sqrt{n^*}\left(\hat{\boldsymbol{\gamma}}_p - \boldsymbol{\gamma}_p\right) \to_d N(\mathbf{0}, \mathbf{V}) \tag{16}$$

where $\mathbf{V} = \{v_{kl}; k, l = 0, 1, 2, \ldots, K\}$, *with*

$$v_{kl} = \frac{2}{2\pi} \int_{-\pi}^{\pi} [2\pi f(\omega)]^{2p} \cos(\omega k) \cos(\omega l)\, d\omega \tag{17}$$

and $n^* = \frac{n}{m[C(m;p,p)-1]}$, *with*

$$C(m; p, q) = \frac{\Gamma(m + p + q)\Gamma(m)}{\Gamma(m + p)\Gamma(m + q)}. \tag{18}$$

The proof, given in Appendix A.1, is based on the asymptotic properties of the periodogram of a linear process, that require the strong convergence assumption on the coefficients of the linear process, on the fractional moments of Gamma random variables and on a central limit theorem for non linear functionals of the periodogram due to Faÿ, Moulines and Soulier (2002), which can be applied when some regularity conditions on the functional of the spectrum and on the moments of the noise process are satisfied. The latter are easy to verify for a power function and a Gaussian process. Notice that the strong convergence condition on the filter coefficient implies short-range dependence.

For $m = 1$ and $p > 0$, $Y_j^{(p)}$ is the inverse Laplace transform of $[2\pi f(\omega_j)]^{-(p+1)}$ evaluated at $2\pi I(\omega_j)$, that gives an estimator of $[2\pi f(\omega_j)]^p$ as in Taniguchi (1980), so that the consistency and the asymptotic normality of (15) follows from Taniguchi (1980). For large m, using Stirling's approximation, $\Gamma(m)/\Gamma(m+p) \approx m^{-p}$, $Y_j^{(p)} \approx \left(2\pi \bar{I}_j\right)^p$, and interpreting $\bar{I}_j$ as a kernel (Daniell) estimator of the spectral density at $\bar{\omega}_j$, theorem 6.1.2 in Taniguchi and Kakizawa (2000) can be applied, as the power transformation is a continuously twice differentiable function of ω and $\cos(k\omega)$ is even and continuous in $[-\pi, \pi]$. Since our result rests on the normality assumption, the additive component of asymptotic variance depending on the fourth cumulant vanishes. Although our result is derived under more restrictive assumptions, it embodies a finite sample bias correction and establishes a lower bound for m in the case of a negative p.

For $m = 1$ and $p = 1$ the estimator (15) gives the sample autocovariance at lag k, that is $\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$ for $k = 0, \ldots, n-1$ and $\hat{\gamma}_{-k} = \hat{\gamma}_k$, which follows from the relation

$$I(\omega_j) = \frac{1}{2\pi} \sum_{|h|<n} \hat{\gamma}_h \cos(\omega_j h).$$

Moreover, equation (17) gives the Bartlett's formula for the generic element of the asymptotic covariance matrix of sample covariance, $\hat{\gamma}_k$. In fact, by lemma 1, equation (4), and by the

prostapheresis formulae, equation (17) can be written as

$$v_{kl} = \sum_{j=-\infty}^{\infty} (\gamma_{p,j+k}\gamma_{p,j+l} + \gamma_{p,j+k}\gamma_{p,j-l}) \tag{19}$$

which for $m = 1$ and $p = 1$ coincides with the asymptotic covariance of $\hat{\gamma}_k$.

In addition, the arguments of the proof allow us to derive the asymptotic covariance between generalised autocovariance estimators across different power transformations, i.e.

$$\mathrm{Cov}(\hat{\gamma}_{pk}, \hat{\gamma}_{ql}) = \frac{1}{n}(C(m;p,q) - 1)\frac{2m}{2\pi}\int_{-\pi}^{\pi} [2\pi f(\omega)]^{p+q} \cos(\omega k)\cos(\omega l)\mathrm{d}\omega. \tag{20}$$

A consistent estimator of (20) is

$$\widehat{\mathrm{Cov}}(\hat{\gamma}_{pk}, \hat{\gamma}_{ql}) = (C(m;p,q) - 1)\frac{1}{M}\sum_{j=1}^{M-1} (2\pi \bar{I}_j)^{p+q} \cos(\bar{\omega}_j k)\cos(\bar{\omega}_j l)\mathrm{d}\omega. \tag{21}$$

Consistency follows from the same arguments that imply consistency of (15), see the last paragraph of the proof of theorem 1 in appendix A.1.

Under the assumptions of theorem 1, similar results can be derived for the generalised autocorrelation function, that is estimated based on (15) by

$$\hat{\rho}_{pk} = \frac{\hat{\gamma}_{pk}}{\hat{\gamma}_{p0}}. \tag{22}$$

Theorem 2 *Let us consider the vectors* $\boldsymbol{\rho}_p' = [\rho_{p1}, \rho_{p2}, \ldots, \rho_{pK}]'$ *and* $\hat{\boldsymbol{\rho}}_p' = [\hat{\rho}_{p1}, \hat{\rho}_{p2}, \ldots, \hat{\rho}_{pK}]'$ *having components as in (3) and (22), respectively. Under the assumptions of theorem 1,*

$$\sqrt{n^*}\left(\hat{\boldsymbol{\rho}}_p - \boldsymbol{\rho}_p\right) \to_d N(\mathbf{0}, \mathbf{W}) \tag{23}$$

where $\mathbf{W} = \{w_{kl}; k, l = 1, 2, \ldots, K\}$, *with generic element given by the Bartlett's formula*

$$w_{kl} = \sum_{j=-\infty}^{\infty} (\rho_{p,j+k}\rho_{p,j+l} + \rho_{p,j+k}\rho_{p,j-l} + 2\rho_{p,k}\rho_{p,l}\rho_{p,j}^2 - \rho_{p,k}\rho_{p,j}\rho_{p,j+l} - \rho_{p,l}\rho_{p,j}\rho_{p,j+k}). \tag{24}$$

The proof is in Appendix A.2 and it is a standard proof based on the same arguments that lead to the proof of the Bartlett's formula in the case when $p = 1$. The asymptotic covariance matrix is estimated by replacing $\hat{\boldsymbol{\rho}}_p$ for the population quantities into the expression for $\mathbf{W}$.

In finite samples, the mean square errors of the GACV and GACF estimators, $\hat{\gamma}_{pk}$ and $\hat{\rho}_{pk}$, are a rather complicated function of p, m, and the spectral properties of x_t. Luati, Proietti and Reale (2012) propose the use of the the jackknife (Quenouille, 1949, see Miller, 1974, and Efron and Tibshirani, 1993, for reviews) for selecting the bandwidth parameter m.

6 GACV-based Tests for White Noise

Two classes of tests for white noise can be based on the GACV. When applied to the residuals from a time series model, they serve as goodness of fit tests.

6.1 Generalised Portmanteau Tests

The generalised Portmanteau test statistic for lack of serial correlation, $H_0 : \rho_{p1} = \rho_{p2} = \cdots = \rho_{pK} = 0$, is

$$\mathrm{BP}_p = n^* \sum_{k=1}^{K} \hat{\rho}_{pk}^2. \tag{25}$$

By Theorem 2, (25) provides an asymptotically χ_K^2 test, generalising the Box-Pierce (1970) statistic $\mathrm{BP} = n \sum_{k=1}^{K} \tilde{\rho}_k^2$ where $\tilde{\rho}_k = \sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x}) / \sum_{t=1}^{n}(x_t - \bar{x})^2$. The generalisation of the modified statistic $\mathrm{LB} = n(n+2) \sum_{k=1}^{K} (n-k)^{-1} \tilde{\rho}_k^2$, known as the Ljung-Box (1978) statistic, is also possible.

6.2 Generalised Milhøj Goodness of Fit Tests

A class of test statistics, generalising Milhøj (1981) goodness of fit test, exploits an important property of the GACV, which is a direct consequence of Lemma 1: $\gamma_{2p,0} = \gamma_{p0}^2 + 2 \sum_{j=1}^{\infty} \gamma_{pj}^2$. Hence, the ratio

$$\mathcal{R}_p = \frac{\gamma_{2p,0}}{\gamma_{p0}^2} = 1 + 2 \sum_{j=1}^{\infty} \rho_{pj}^2$$

equals 1 for a WN process. A test of the null $H_0 : \mathcal{R}_p = 1$ against $H_1 : \mathcal{R}_p > 1$ can then be based on the estimated ratio $\widehat{\mathcal{R}}_p = \frac{\hat{\gamma}_{2p,0}}{\hat{\gamma}_{p0}^2}$, whose null distribution has variance

$$\mathrm{Var}(\widehat{\mathcal{R}}_p) = \frac{1}{M} \left\{ 4C(m; p, p) + C(m; 2p, 2p) - 4C(m; 2p, p) - 1 \right\}.$$

Hence, the test statistic

$$\mathcal{M}_p = \frac{\widehat{\mathcal{R}}_p - 1}{\sqrt{\mathrm{Var}(\widehat{\mathcal{R}}_p)}} \tag{26}$$

provides an asymptotically standard normal test.

The test (26) has the advantage of being independent of the choice of a particular truncation lag K, and of depending of the full generalized autocorrelation function. For $m = p = 1$ it is coincident with the goodness of fit test of Milhøj (1981). It is related to the classes of serial correlation tests proposed by Hong (1996), and in particular that based on the statistic

$H_n = n\sum_{j=1}^{B} \mathcal{K}^2(j)\tilde{\rho}_j^2$, where $\mathcal{K}(j)$ is a lag window, e.g. the Tukey-Hanning kernel $\mathcal{K}(j) = 0.5[1 + \cos(\pi j/\tau)]$, for $|j|/\tau \leq 1$, $\mathcal{K}(j) = 0$, for $|j|/\tau > 1$, and τ is the truncation parameter. The relationship has been made clear by Chen and Deo (2004a), see also Beran (1992), who propose a test based on $\mathcal{T}_n = \left[\frac{2\pi}{n}\sum_{j=0}^{n-1}\tilde{f}(\omega_j)\right]^{-2}\frac{2\pi}{n}\sum_{j=0}^{n-1}\tilde{f}^2(\omega_j)$, where $\tilde{f}(\omega_j)$ is an estimate of the spectral density at the Fourier frequency ω_j, and showed that H_n and $n(\pi\mathcal{T}_n - 0.5)$ have the same asymptotic distribution. Our test statistic depends on m and p. Their role will be illustrated by a Monte Carlo (MC) experiment. Notice that (26) with $m > 1$ implies a Daniell-type estimation of the spectral density (the corresponding lag window is the sinc function; see Priestley, 1981, p 440).

Table 1 reports the size of the WN tests proposed so far when $x_t \sim \text{NID}(0, 1)$, estimated by MC simulation using 50,000 replications, for three different significance levels (10%, 5% and 1%), two sample sizes (128 and 512), using K autocorrelations (BP_p tests and LB) or $\tau = K$ truncation parameter (for the Hong and Chen-Deo statistics), and pooling parameter m. For BP_p we report only the case $m = 1$. The column "Dist" provides the Euclidean distance between the upper tail quantiles of the MC distribution and those of the asymptotic distribution (from 0.80 to 0.995 with step 0.005); hence, it measures the discrepancy between the empirical distribution and the asymptotic approximation in the upper 20% tail.

While it can be seen that the size properties of the BP_p test are only marginally improved by choosing $p < 1$, as far as $\mathcal{M}_p$ is concerned, having $p < 1$ yields more substantial gains. The heuristic explanation is that in finite sample fractional values of p have a normalisation effect on the distribution It should be recalled that the cube root transformation is the normalising transformation for a χ_1^2 random variable. The Hong (1996) and Chen and Deo (2004a) tests suffer from size distortions, which are resolved in Chen and Deo (2004b) by taking a power transformation of the test statistic, aiming at reducing the skewness of the distribution. In our case, the idea of transforming the periodogram is already embodied in GACV estimate. The null distribution of the $\mathcal{M}_p$ tests are displayed in figure 2, and compared to the Hong (1996) and Chen and Deo (2004a) tests, whose distribution before the correction is markedly right skewed.

Figure 3 displays the logarithm of the power function of the $\mathcal{M}_p$ test conducted at the 5% significance level, when the data are generated by a first order AR process. The plot confirms that choosing $p < 1$ yields a test statistic with improved finite sample properties.

Table 1: Effective sizes of WN tests. The data are generated as $x_t \sim \text{NID}(0,1)$

	$n = 128$											
	$K = 8$				$K = 13$				$K = 21$			
	10%	5%	1%	Dist	10%	5%	1%	Dist	10%	5%	1%	Dist
$\text{BP}_{1/3}$	14.58	7.95	1.84	1.75	15.25	8.30	1.99	1.78	17.17	9.40	2.25	2.03
$\text{BP}_{1/2}$	11.50	6.05	1.24	0.65	11.82	6.12	1.34	0.66	12.42	6.41	1.49	0.83
$\text{BP}_{2/3}$	10.41	5.26	1.12	0.22	10.56	5.43	1.18	0.34	10.88	5.84	1.41	0.59
$\text{BP}_{3/4}$	10.09	5.09	1.08	0.14	10.31	5.32	1.21	0.32	10.56	5.69	1.46	0.63
BP_1	9.35	4.73	1.14	0.30	9.82	5.31	1.43	0.60	10.39	5.92	1.86	1.20
BP	8.41	4.34	1.03	0.51	7.87	4.13	1.05	0.65	6.74	3.67	1.08	1.07
LB	10.40	5.51	1.37	0.62	11.05	6.10	1.75	1.13	12.01	7.00	2.30	1.87
Hong	9.69	6.50	3.21	2.47	10.13	6.67	3.06	2.31	10.67	6.97	3.16	2.29
Chen-Deo	10.10	6.81	3.42	2.68	10.73	7.06	3.27	2.56	11.37	7.43	3.44	2.57
	$m = 1$				$m = 3$				$m = 5$			
$\mathcal{M}_{1/3}$	9.65	5.27	1.46	0.42	8.63	4.97	1.57	0.62	7.84	4.70	1.62	0.77
$\mathcal{M}_{1/2}$	9.40	5.29	1.48	0.47	8.45	4.84	1.58	0.59	7.69	4.63	1.60	0.76
$\mathcal{M}_{2/3}$	8.99	5.23	1.61	0.64	8.06	4.63	1.62	0.68	7.37	4.44	1.58	0.83
$\mathcal{M}_{3/4}$	8.70	5.10	1.63	0.73	7.90	4.57	1.59	0.76	7.15	4.32	1.52	0.88
$\mathcal{M}_1$	7.65	4.61	1.73	1.04	7.02	4.17	1.55	1.03	6.42	3.85	1.46	1.12
	$n = 512$											
	$K = 11$				$K = 20$				$K = 37$			
	10%	5%	1%	Dist	10%	5%	1%	Dist	10%	5%	1%	Dist
$\text{BP}_{1/3}$	11.16	5.66	1.32	0.48	11.59	5.97	1.27	0.59	12.31	6.45	1.40	0.73
$\text{BP}_{1/2}$	10.38	5.19	1.14	0.20	10.54	5.41	1.08	0.23	10.96	5.60	1.12	0.32
$\text{BP}_{2/3}$	10.04	5.05	1.09	0.10	10.29	5.16	1.06	0.14	10.50	5.37	1.17	0.26
$\text{BP}_{3/4}$	9.90	4.97	1.08	0.11	10.25	5.11	1.08	0.13	10.47	5.48	1.22	0.31
BP_1	9.66	4.94	1.12	0.17	10.17	5.22	1.26	0.33	10.80	5.82	1.50	0.66
BP	9.29	4.75	1.03	0.21	9.10	4.58	1.08	0.28	8.21	4.28	1.05	0.46
LB	9.98	5.15	1.18	0.24	10.58	5.47	1.34	0.47	11.29	6.16	1.67	0.84
Hong	10.29	6.90	3.18	2.29	10.36	6.81	2.87	1.91	10.84	6.71	2.68	1.63
Chen-Deo	10.43	7.00	3.23	2.35	10.55	6.93	2.93	1.97	11.10	6.90	2.77	1.71
	$m = 1$				$m = 3$				$m = 5$			
$\mathcal{M}_{1/3}$	9.95	5.16	1.20	0.18	9.46	5.21	1.46	0.38	9.13	5.24	1.51	0.46
$\mathcal{M}_{1/2}$	9.89	5.17	1.27	0.24	9.37	5.19	1.43	0.36	9.06	5.14	1.50	0.45
$\mathcal{M}_{2/3}$	9.75	5.37	1.42	0.36	9.30	5.13	1.47	0.41	8.97	5.00	1.49	0.48
$\mathcal{M}_{3/4}$	9.65	5.47	1.50	0.44	9.11	5.19	1.54	0.46	8.86	5.01	1.53	0.51
$\mathcal{M}_1$	9.35	5.37	1.81	0.76	8.75	5.13	1.67	0.68	8.57	4.85	1.65	0.70

Figure 2: Distribution of white noise tests statistics based on 50,000 simulations of Gaussian $x_t \sim \text{NID}(0,1)$ with $n = 128, m = 1, K = 8$.

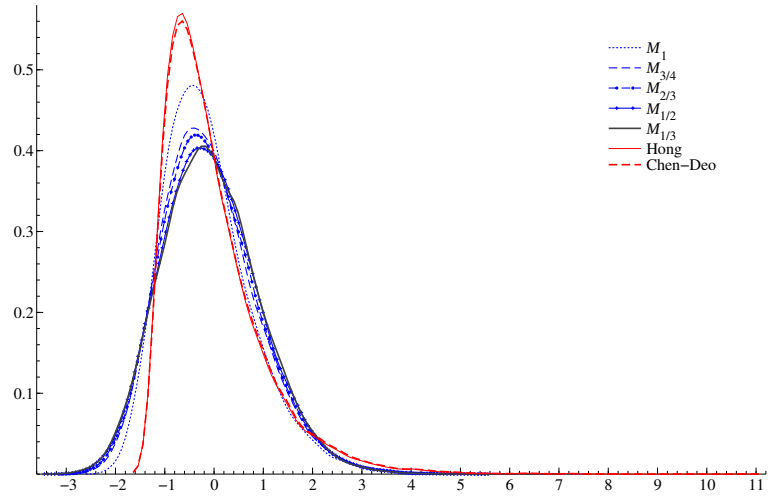


Figure 3: Logarithm of the power function of the test $\mathcal{M}_p$ based on 20,000 simulations from the AR(1) process $x_t = \phi x_{t-1} + \xi_t, \xi_t \sim \text{NID}(0,1)$ with $n = 128, m = 1$. The straight line is drawn at the log-size $\ln(0.05)$.

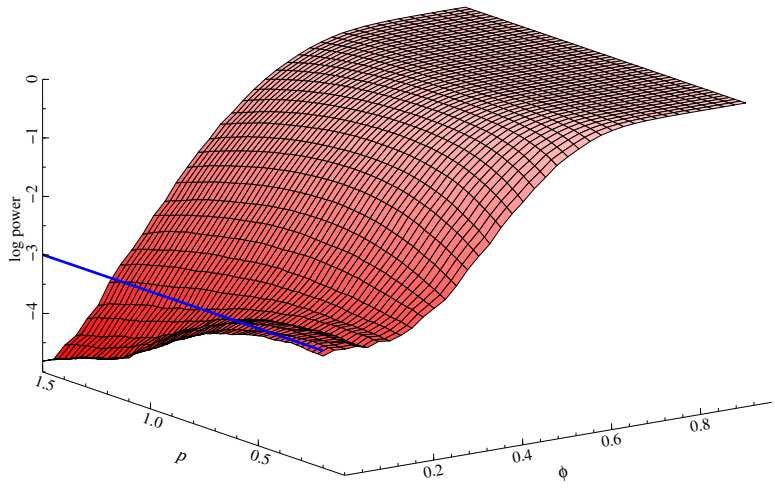
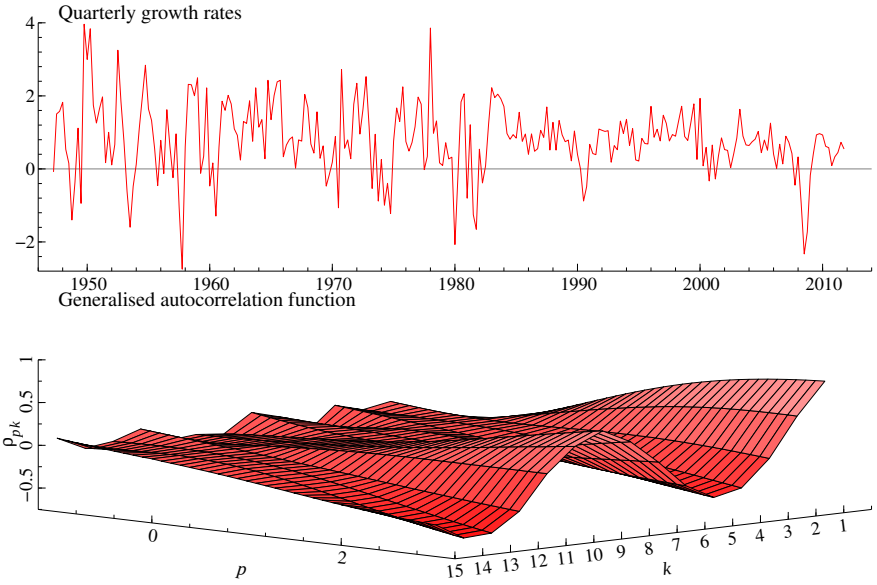


Figure 4: US Gross domestic product: quarterly growth rates (1947.2-2012.1), periodogram and GACF, estimated with $m = 3$.



7 Feature matching: a Yule-Walker spectral estimator based on the GACV

An important use of the GACV is in extracting features of interest from a time series. Figure 4 displays the quarterly growth rate of the US Gross Domestic Product (1947.2-2012.1), along with its estimated GACF, using $m = 3$ for p ranging from -1 to 3 (recall that $m \geq 3$ is needed to estimate the GACV at $p = -1$ with finite asymptotic variance). The cyclical nature of this series has represented a long debated issue. See Harvey and Jäger (1997) and the references therein. The periodogram (see also figure 5) does indeed display large ordinates at low frequencies and $\hat{\rho}_{pk}$ describes a pseudo-cyclical pattern for values of p greater than 1. However, parsimonious ARMA models, selected on the basis of information criteria, fail to capture the cyclical feature of GDP growth and fit a monotonically decreasing spectrum with a global maximum at the origin.

In this section we propose a Yule-Walker spectral estimator based on the GACV. We illustrate that allowing for a power transformation parameter greater than 1 amounts to boosting the cyclical features of the series, as large periodogram ordinates will receive an higher weight.

The Yule-Walker estimation method is very popular in time series estimating the autoregressive parameters (see Percival and Walden, 1993, for a review). Recently, Xia and Tong (2011) have introduced an approach for time series modelling, aiming at matching stylised features of the time series, such as the autocorrelation structure. We consider here a feature matching Yule-Walker estimate of the spectrum with a similar intent, which uses the GACV at different values of p.

Let $\mathbf{\Gamma}_{p,K}$ denote the Toeplitz matrix, formed from the GACV, with generic element $\gamma_{p,|h-k|}, h, k = 0, \ldots, K-1$, let $\boldsymbol{\gamma}_{p,K} = (\gamma_{p1}, \ldots, \gamma_{pK})'$, and $\boldsymbol{\phi}_{p,K} = (\phi_{1p}, \ldots, \phi_{pK})'$. The latter is a $K \times 1$ vector of AR coefficients satisfying the Yule-Walker equations $\mathbf{\Gamma}_{p,K}\boldsymbol{\phi}_{p,K} = \boldsymbol{\gamma}_{p,K}$. The polynomial $\phi_p(B) = 1 - \phi_{1p}B - \cdots \phi_{pK}B^K$, characterises the AR approximation of the process u_{pt}, and provides directly the spectral factorisation $[2\pi f(\omega)]^p \propto [\phi_p(e^{-\imath\omega})\phi_p(e^{\imath\omega})]^{-1}$. By (9) we can obtain the AR approximation of order $K' > K$ for the original process, $\pi(B)x_t = \xi_t$, $\pi(B) = [\phi_p(B)]^{1/p}$, or, equivalently, the moving average representation $x_t = \psi(B)\xi_t$, $\psi(B) = [\phi_p(B)]^{-1/p}$. Given a time series realisation, we replace the theoretical GACF with the estimated one to get $\hat{\boldsymbol{\phi}}_{p,K} = \hat{\mathbf{\Gamma}}_{p,K}^{-1}\hat{\boldsymbol{\gamma}}_{p,K}$, applying (9), we obtain different estimates for the spectrum of the time series according to the value of p, $\hat{f}_p(\omega)$.

Figure 5 displays the periodogram of the US GDP quarterly growth rate series and the spectral estimates corresponding to $p = 0.5, 1, 2, 3, 4$, using $K = 3$ sample GACVs. No cyclical peaks is identified for $p \leq 1$, but as p increases, the cyclical properties of GDP growth become prominent. For judging which spectral estimate is more suitable, we consider a measure of deviance equal to minus twice the Whittle's likelihood (Whittle, 1961), as advocated by Xia and Tong (2011), $\text{dev}(p) = \sum_{j=1}^{n}\left[\frac{I(\omega_j)}{\hat{f}_p(\omega_j)} + \ln \hat{f}_p(\omega_j)\right]$. The plot of $\text{dev}(p)$ versus p (right panel of figure 5) suggests the value $\tilde{p} = 2.65$.

The spectral peak, corresponding to a period of roughly 2.5 years (10 quarters), can alternatively be identified by increasing the AR order, as figure 6 shows, but there is a risk of overfitting the sample spectrum in other frequency ranges.

8 Time Series Cluster and Discriminant Analysis

Let us consider two stochastic processes, $\{x_{it}\}_{t\in T}$ and $\{x_{jt}\}_{t\in T}$, and let $f_i(\omega)$ and $f_j(\omega)$ be their spectral densities. The p-squared distance (p-sd, henceforth) between the two processes is defined as the integrated squared difference between their power transformed spectra, which is

Figure 5: US Gross domestic product (1947.2-2012.1): spectrum estimation by the p-Yule-Walker method using $K = 3$.

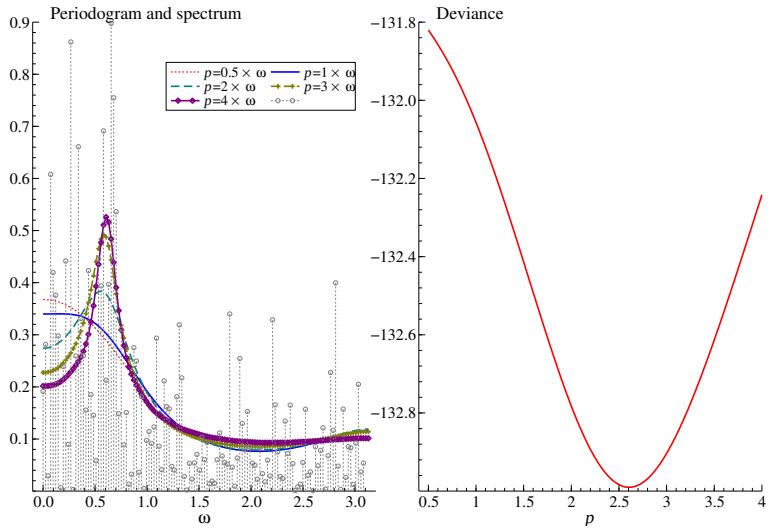
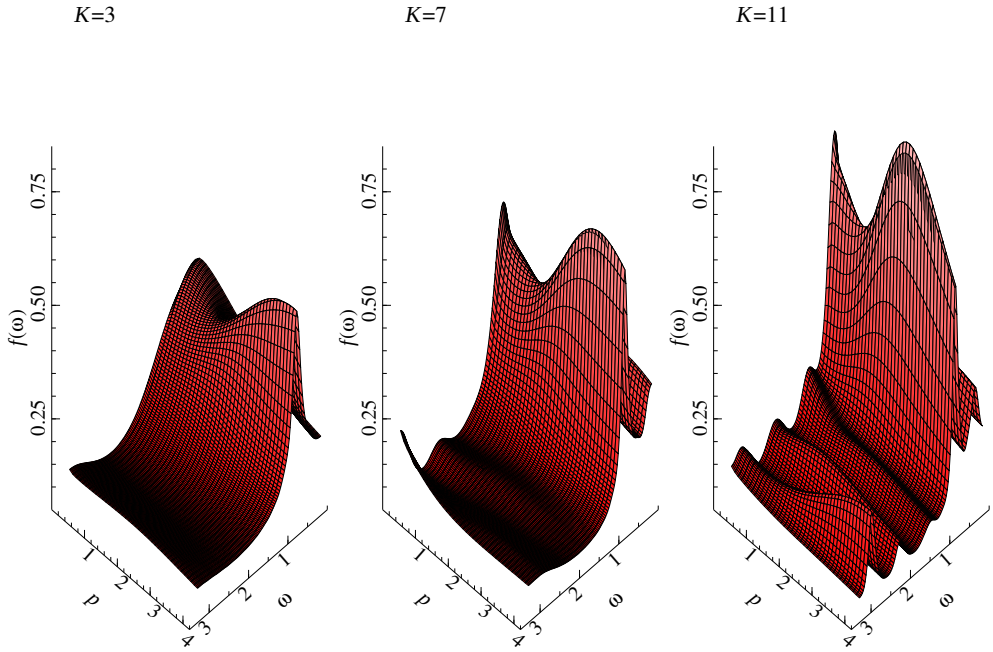


Figure 6: US Gross domestic product (1947.2-2012.1): Yule-Walker estimates of the spectrum (as a function of p and ω), based on different GACV orders.



equivalent to the Euclidean distance between the GACVs $\gamma_{i,pk}$ and $\gamma_{j,pk}$ of the two processes:

$$\begin{aligned} d_{ij,p}^2 &= \tfrac{1}{2\pi}\int_{-\pi}^{\pi}\left\{[2\pi f_i(\omega)]^p - [2\pi f_j(\omega)]^p\right\}^2 \mathrm{d}\omega \\ &= \sum_{k=-\infty}^{\infty}\left(\gamma_{i,pk} - \gamma_{j,pk}\right)^2 \\ &= \gamma_{i,2p,0} + \gamma_{j,2p,0} - 2\sum_{k=-\infty}^{\infty}\gamma_{i,pk}\gamma_{j,pk}. \end{aligned} \tag{27}$$

The p-sd (27) encompasses the Euclidean distance ($p = 1$), referred to as the quadratic distance in Hong (1996) and the Hellinger distance ($p = 1/2$). It can also be based on the normalised spectral densities of the two processes, in which case the autocorrelations ρ_{pk} replace the autocovariances in (27).

The p-sd can be estimated by

$$\hat{d}_{ij,p}^2 = \left(\hat{\gamma}_{i,p0} - \hat{\gamma}_{j,p0}\right)^2 + 2\sum_{k=1}^{K}\left(\hat{\gamma}_{i,pk} - \hat{\gamma}_{j,pk}\right)^2.$$

or, if the autocorrelations are used,

$$\hat{d}_{ij,p}^2 = 2\sum_{k=1}^{K}\left(\hat{\rho}_{i,pk} - \hat{\rho}_{j,pk}\right)^2.$$

The p-sd can be used for clustering time series and estimation by feature matching, if the distance is computed with respect to the theoretical GACF implied by a time series model. In the stationary case, for which $|\rho_{1k}|$ declines at a geometric or hyperbolic rate, when p is larger than 1, the contribution of low order, high autocorrelations to the overall distance will be higher. On the contrary, for values of $0 < p < 1$ less than unity, the contribution of high order, small autocorrelations, will be comparatively larger. Similar considerations hold for negative p, but with reference to the inverse autocorrelations.

Another use is in discriminant analysis. The relevance of generalising the distance to fractional and negative values of p is illustrated by an application of Fisher's linear discriminant analysis (Mardia, Kent and Bibby, 1979) to a simulated data set.

$N = 1,050$ time series of size n were generated under three different models: $N_1 = 600$ AR(1) series with coefficient ϕ randomly drawn from a uniform distribution in [0.1, 0.9], $N_2 = 300$ MA(1) series with coefficient θ uniformly distributed in $[-0.9, -0.1]$, and $N_3 = 150$ fractional noise series with memory parameter uniformly distributed in [0.1, 0.4].

Two-thirds of the series were used as a training sample to estimate the canonical variates, and the remaining third was used as a test sample. The objective is to classify correctly the 350 test series by predicting their generating model. Different values of p were used to compute the GACF up to lag K for both the training and the test sample.

For the training sample the two canonical variates are obtained from the generalised eigenvectors of the between groups deviance matrix, $\mathbf{B}$, satisfying $\mathbf{Ba} = \lambda\mathbf{a}$, $\mathbf{a'Wa} = 1$, where $\mathbf{W}$ is the within groups deviance matrix and λ is the generalised eigenvalue of $\mathbf{B}$, for $\lambda > 0$.

The test series are then classified according to the smallest Mahalanobis' distance to the GACF group means, which amounts to computing the canonical scores for the test series, by combining linearly the GACFs using the weights $\mathbf{a}$ computed on the training sample, and assigning the series to the group for which the distance with the canonical means is a minimum.

Different values of p yield different discriminant functions and different results. We select the optimal solution (across the values of p) as the one minimising the missclassification rate (MR) computed for the test sample.

Figure 7 displays the MRs for values of p in the range [-2, 2] for a simulation dataset with $n = 1,000$, $K = 30$. For estimating the GACF we set $m = 6$. The value of p yielding the smallest MR resulted $\tilde{p} = -0.7$ (replicating the experiment, we always obtain values in the range [-1,-0.2]); the improvement with respect to $p = 1$ is large (around a 5% reduction in the MR). The generalised eigenvectors $\mathbf{a}$, defining the two canonical variates for $\tilde{p}$ are also plotted. Interestingly, the first canonical variable assigns declining (negative) weights to the GACF from 2 to K, whereas the second is a contrast between the first two GACF and the higher order ones. The two canonical variate scores for the training sample are displayed in figure 8: it illustrates that the solution provides an effective separation of the three groups.

9 Conclusions

The paper has defined the generalised autocovariance function and has shown its potential for three different analytic tasks: testing for white noise, the estimation of the spectrum of cyclical time series, and time series methods based on the distance between stochastic processes, like cluster and discriminant analysis.

By tuning the power transformation parameter given features of a time series can be emphasised or muted for a particular purpose. In this respect, we think that the proposed feature matching Yule-Walker spectral estimator based the GACV has very good potential for the identification of spectral peaks of time series affected by noise. As p increases, the contribution of the noise to the spectrum will be subdued to some extent and the AR fit will attempt at matching the cyclical properties of the series more closely.

We have also argued that for fractional values of p in (0,1), the tests for white noise based on the GACV have better finite sample properties than those defined on the untransformed

Figure 7: Canonical analysis of simulated series: missclassification rate as a function of p, and canonical variates weights for the optimal p.

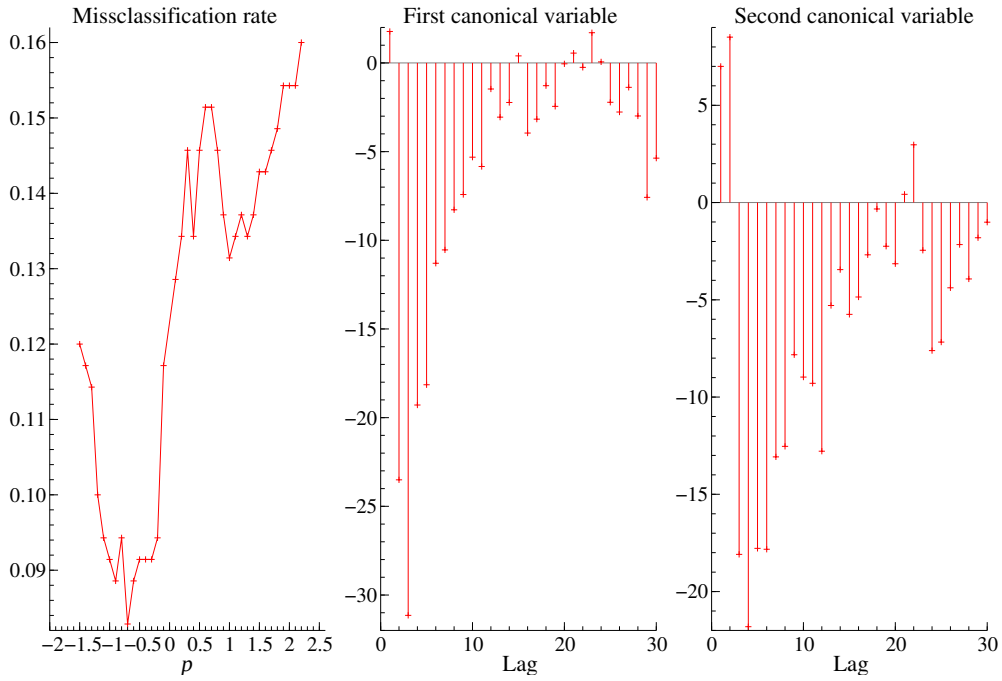
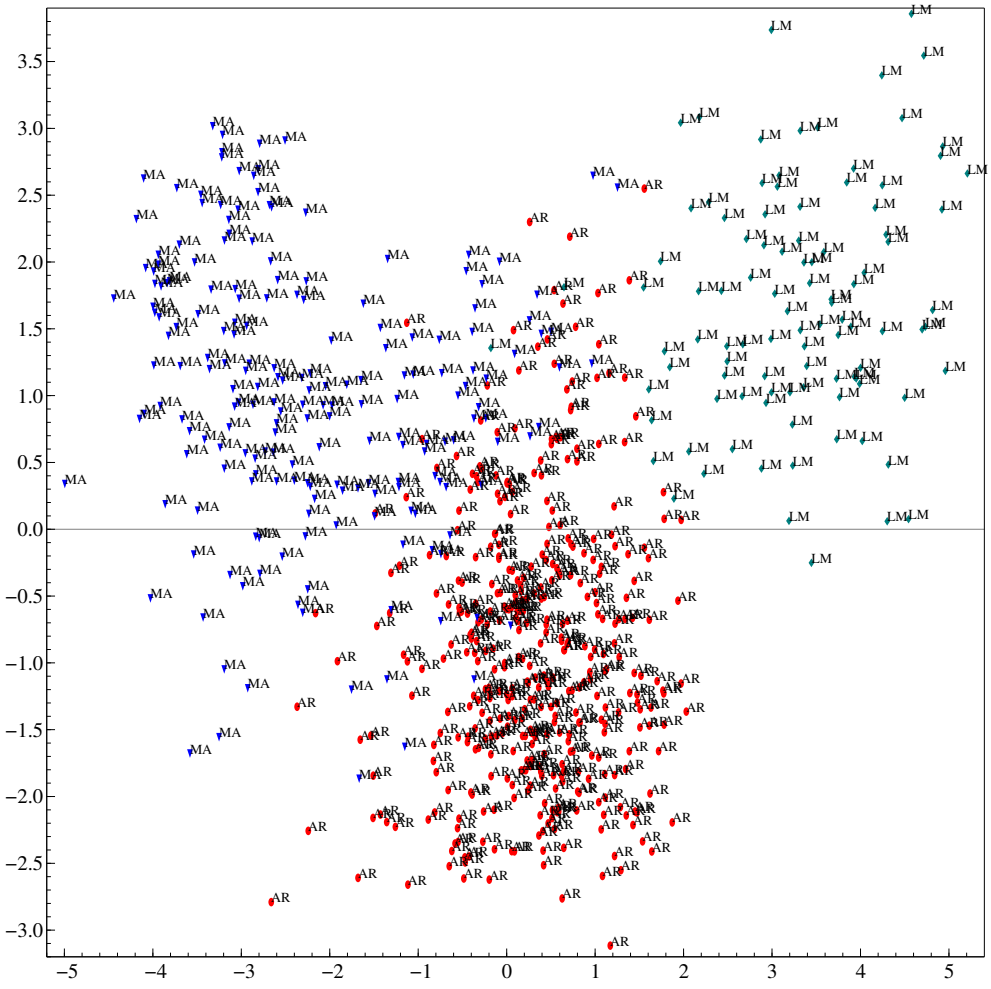


Figure 8: Canonical analysis of simulated series: plot of the two canonical variates for the training sample, consisting of 400 AR series, 200 MA series and 50 fractional noise series. The value of p is -0.7.



spectrum or autocorrelation function. We leave to future research the generalisation of Bartlett's tests (Bartlett, 1954) for white noise based on the normalized integrated power transformation of the spectrum (see Priestley, sec. 6.2.6., Durlauf, 1991, Deo, 2000, and Delgado, Hidalgo and Velasco, 2005).

Another extension that we have not investigated is the partial generalised autocorrelation function, which can be used as an additional model identification tool, complementing the traditional one. El Ghini and Francq (2006) advocated the use of the inverse partial ACF, and their results encourage further investigating this direction.

Finally, we plan to construct an estimator of the long memory parameter based on the GACV, which generalises the minimum distance estimators based on the autocorrelation function (Tieslau, Schmidt and Baillie, 1996) and the variance profile (Luati, Proietti and Reale, 2012).

A Appendix

A.1 Proof of theorem 1

Let us consider the estimator (15). Under the assumption $\sum_{j=0}^{\infty} |j|^{\frac{1}{2}} |\psi_j| < \infty$, the periodogram ordinates of a linear process with finite fourth moment evaluated at the Fourier frequencies $0 < \omega_1 < \cdots < \omega_k < \pi$ are asymptotically IID exponentials with means equal to $f(\omega_j)$, $j = 1, 2, \ldots, k$ (Brockwell and Davis, 1991, Theorem 10.3.2). If M and m are large enough for asymptotics and $\frac{m}{M}$ is small enough for $f(\omega)$ to be constant over frequency intervals of length $\frac{2\pi m}{M}$, then, for fixed m, $\bar{I}_j$ can be interpreted as a Daniell type estimator for $f(\bar{\omega}_j)$ and $2\pi\bar{I}_j \sim 2\pi f(\bar{\omega}_j)\mathcal{G}_m$, where the $\mathcal{G}_m$ are independent and identically distributed basic Gamma random variables with shape parameter equal to m (see Koopmans, 1974, pp. 269-270). Hence, $\mathrm{E}(Y_j^{(p)}) = (2\pi f(\bar{\omega}_j))^p$ and $\mathrm{Cov}(Y_j^{(p)}, Y_j^{(q)}) = (C(m; p, q) - 1)[2\pi f(\bar{\omega}_j)]^{p+q}$. Notice that, for $p = q$, $\mathrm{Var}(Y_j^{(p)})$ exists for $p > -\frac{m}{2}$.

Let us consider the sequence

$$S_n(Y_j^{(p)}) = \sum_{j=0}^{M-1} b_j \left\{ Y_j^{(p)} - (2\pi f(\bar{\omega}_j k))^p \right\} \tag{28}$$

where $b_j = c_j / (s(c_j^2))^{\frac{1}{2}}$, with $c_j = \cos(\bar{\omega}_j k)$ and $s(c_j^2) = \sum_{j=0}^{M-1} \cos^2(\bar{\omega}_j k)$. By construction, the coefficients b_j satisfy $\sum_{j=0}^{M-1} b_j^2 = 1$ and $\max_{0 \le j \le M-1} |b_j| \to 0$, since $\cos(\omega k)$ is a function of bounded variation for $\omega \in (-\pi, \pi)$.

If the $Y_j^{(p)}$ were IID or, equivalently, if the process $\{x_t\}_{t \in T}$ were a Gaussian white noise, then a central limit theorem for linear combinations of independent random variables would directly apply to $S_n(Y_j^{(p)})$, as in Luati, Proietti and Reale (2012). Given that the random variables $Y_j^{(p)}$ are asymptotically independent, an intermediate step is needed, that consists in proving the asymptotic negligibility of the term $S_n(Y_j^{(p)}) - S_n(Z_j^{(p)})$, where $Z_j^{(p)} = (2\pi\bar{I}_{j,\xi})^p \frac{\Gamma(m)}{\Gamma(m+p)}$, $\bar{I}_{j,\xi}$ denoting the pooled periodogram of $(\xi_1, \ldots, \xi_n)$ at $\bar{\omega}_j$. In this way, a central limit theorem for $S_n(Y_j^{(p)})$ can be proved as a consequence of a central limit theorem for $S_n(Z_j^{(p)})$, see Faÿ, Moulines and Soulier (2002). A similar approach is in Klüpperberg and Mikosch (1996), though in a slightly different context. The implications of $S_n(Y_j^{(p)}) - S_n(Z_j^{(p)}) \to_p 0$ on a central limit theorem for $S_n(Y_j^{(p)})$ are proved based on a Bartlett-type decomposition that relates the pooled periodogram of the observed variables with the pooled (pseudo) periodogram of the noise process and requires the strong assumption on the filter coefficients, $\sum_{j=0}^{\infty} |\psi_j| |j|^{\frac{1}{2}+\delta} < \infty$, $\delta < 0$. The latter condition implies short-range dependence on the process. Some additional regularity conditions on the non linear function of the spectrum that is the object of inference are required

to prove the weak convergence of $S_n(Y_j^{(p)})$ to $S_n(Z_j^{(p)})$. In our context, Theorem 1 of Faÿ, Moulines and Soulier (2002) applies to $S_n(Y_j^{(p)})$, since assumptions 1, 2, 4, 7 and S1, that have to be satisfied when the noise term is Gaussian, hold. Specifically, assumptions 1 and 2 concern the coefficients b_j, assumption 4 requires the existence of the asymptotic variance of $S_n(Y_j^{(p)})$, which occurs for $p > -\frac{m}{2}$, assumption 7 is the convergence on the filter coefficients necessary for the Bartlett decomposition and assumption S1, such that $\max_{x\in\mathbb{R}} \frac{|\phi(x)|+|\phi'(x)|+|\phi''(x)|}{1+|x|^\nu} < \infty$, is satisfied for $\phi(x) = x^p$ and for a Gaussian process, having all finite moments $\mu \geq \min\{8\nu, 4\}$. Hence,

$$S_n(Y_j^{(p)}) \rightarrow_d \mathrm{N}\left(0, \sum_{j=0}^{M-1} b_j^2 \mathrm{Var}\left(Y_j^{(p)}\right)\right).$$

As a function of the estimator (15), by multiplying (28) by $\frac{1}{M}(s(c_j^2))^{\frac{1}{2}}$ and rearranging,

$$\frac{1}{M}\sum_{j=0}^{M-1} c_j Y_j^{(p)} \rightarrow_d \mathrm{N}\left(\frac{1}{M}\sum_{j=0}^{M-1} c_j \left(2\pi f(\bar{\omega}_j)\right)^p, \frac{1}{M}\sum_{j=0}^{M-1} c_j^2 \left(2\pi f(\bar{\omega}_j)\right)^{2p} \frac{1}{M}(C(m;p,p)-1)\right). \tag{29}$$

By taking the limits (see also Theorem 2 of Faÿ, Moulines and Soulier, 2002)

$$\sqrt{n^*}(\hat{\gamma}_{pk} - \gamma_{pk}) \rightarrow_d \mathrm{N}\left(0, v_{kk}\right), \tag{30}$$

where $v_{kk} = \frac{2}{2\pi}\int_{-\pi}^{\pi}(2\pi f(\omega))^{2p}\cos^2(\omega k)\mathrm{d}\omega$, $n^* = n(m\left(C(m;p,p)-1\right))^{-1}$ and $C(m;p,q)$ is as in (18). It is straightforward to get the covariance between $\hat{\gamma}_{pk}$ and $\hat{\gamma}_{pl}$ as v_{kl} as in (17). Taking v_{kl} for $k = 0, 1, \ldots, K$ and setting the results in matrix notation complete the proof of the asymptotic distribution of the generalised autocovariance estimator.

Consistency of $\hat{\gamma}_{pk}$ follows by the Chebychev weak law of large numbers, applied to the sequence of random variables in (29) and from the convergence of the Riemannian sum to the integral. This completes the proof of theorem 1.

A.2 Proof of theorem 2

The asymptotic joint normality of the generalised autocorrelation estimators is obtained by applying the delta method to the transformation operated by the function $g : \mathbb{R}^{K+1} \rightarrow \mathbb{R}^K$ which associates the vector $\hat{\boldsymbol{\rho}}_p$, having components $\hat{\rho}_{pj} = \frac{\hat{\gamma}_{pj}}{\hat{\gamma}_{p0}}$, $j = 1, \ldots, K$, with the vector $\hat{\boldsymbol{\gamma}}_{pj}$, $j = 0, 1, \ldots, K$, with components as in (15). The covariance matrix of $\hat{\boldsymbol{\rho}}_p$ is $\mathbf{W} = \mathbf{DVD'}$ with partial derivatives matrix $\mathbf{D} = \frac{1}{\gamma_{p0}}\left[-\boldsymbol{\rho}_p \ \mathbf{I}_K\right]$, so that the generic element of $\mathbf{W}$ is (as in Brockwell and Davis, 1991, proof of theorem 7.2.1) $w_{kl} = v_{kl} - \rho_{pk}v_{0l} - \rho_{pl}v_{k0} + \rho_{pk}\rho_{pl}v_{00}$. By replacing v_{kl} with its expression in equation (19), one gets the generalized Bartlett formula (24).

References

[1] Bartlett, M. S. (1954). Problemes de l'analyse spectral des series temporelles stationnaires. *Publ. Inst. Statist. Univ. Paris*, III–3 119–134.

[2] Battaglia, F. (1983), Inverse Autocovariances and a Measure of Linear Determinism For a Stationary Process, *Journal of Time Series Analysis*, 4, 7987 .

[3] Battaglia, F. (1988), On the Estimation of the Inverse Correlation Function, *Journal of Time Series Analysis*, 9, 1-10.

[4] Battaglia, F., Bhansali, R. (1987), Estimation of the Interpolation Error Variance and an Index of Linear Determinism, *Biometrika*, 74, 4, 771–779.

[5] Beran, J. (1992), A Goodness-of-Fit Test for Time Series with Long Range Dependence, *Journal of the Royal Statistical Society, Series B*, 54, 3, 749–760.

[6] Bhansali, R. (1993), Estimation of the Prediction Error Variance and an R^2 Measure by Autoregressive Model Fitting, *Journal of Time Series Analysis*, 14, 2, 125–146.

[7] Box, G.E.P., and Cox, D.R. (1964), An analysis of transformations (with discussion), *Journal of the Royal Statistical Society, B*, 26, 211–246.

[8] Box, G. E. P., and Pierce, D. A. (1970), Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models, *Journal of the American Statistical Association*, 65, 1509-1526.

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

[10] Chen, W.W. and Deo, R.S. (2004a), A Generalized Portmanteau Goodness-of-Fit Test for Time Series Models, *Econometric Theory*, 20, 382–416.

[11] Chen, W.W. and Deo, R.S. (2004b), Power transformations to induce normality and their applications. *Journal of the Royal Statistical Society*, *Series B*, 66, 117-130.

[12] Cleveland, W.S. (1972), The Inverse Autocorrelations of a Time Series and Their Applications, *Technometrics*, 14, 2, 277–293.

[13] Davis, H.T. and Jones, R.H. (1968), Estimation of the Innovation Variance of a Stationary Time Series, *Journal of the American Statistical Association*, 63, 321, 141–149.

[14] Delgado, M.A., Hidalgo, J. and Velasco, C. (2005), Distribution Free Goodness-of-Fit Tests for Linear Processes, *Annals of Statistics*, 33, 2568–2609.

[15] Deo, R.S. (2000), Spectral Tests of the Martingale Hypothesis under Conditional Heteroscedasticity, *Journal of Econometrics*, 99, 291-315.

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

[17] Durlauf, S.N. (1991). Spectral Based Testing of the Martingale Hypothesis, *Journal of Econometrics*, 50, 355–376.

[18] Efron, B., and Tibshirani R.J. (1993), *An Introduction to the Bootstrap*, Chapman&Hall.

[19] El Ghini, A., and Francq, C. (2006), Asymptotic Relative Efficiency of Goodness-Of-Fit Tests Based on Inverse and Ordinary Autocorrelations, *Journal of Time Series Analysis*, 27, 843–855.

[20] Erdélyi A., Magnus W., Oberhettinger F., and Tricomi F.G. (1953), *Higher Transcendental Functions*, vol. II, Bateman Manuscript Project, McGraw and Hill.

[21] Faÿ, G., Moulines, E. and Soulier, P. (2002), Nonlinear Functionals of the Periodogram, *Journal of Time Series Analysis*, 23, 523–551.

[22] Fuller, W.A. (1996), *Introduction to Statistical Time Series*, John Wiley & Sons.

[23] Gould, H.W. (1974), Coefficient Identities for Powers of Taylor and Dirichlet Series, *The American Mathematical Monthly*, 81, 1, 3–14.

[24] Gradshteyn I.S. and Ryzhik I.M. (1994) *Table of Integrals, Series, and Products* Jeffrey A. and Zwillinger D. Editors, Fifth edition, Academic Press.

[25] Graham, R.L., Knuth, D.E. and Patashnik, O., (1994), *Concrete Mathematics*, Addison-Wesley Publishing Company.

[26] Gray, H. L., Zhang, N.-F. and Woodward, W. A. (1989), On Generalized Fractional Processes. Journal of Time Series Analysis, 10, 233-257.

[27] Hannan, E.J. and Nicholls, D.F. (1977), The Estimation of the Prediction Error Variance, *Journal of the American Statistical Association*, 72, 360, 834–840.

[28] Harvey, A.C., and Jäger, A. (1993), Detrending, Stylized Facts and the Business Cycle, *Journal of Applied Econometrics*, 8, 231–247.

[29] Hong, Y. (1996), Consistent Testing for Serial Correlation of Unknown Form, *Econometrica*, 64, 4 837–864.

[30] Klüpperberg, C., Mikosch, T. (1996), Gaussian Limit Fields for the Integrated Periodogram, *The Annals of Probability*, 6, 3 969–991.

[31] Koopmans, L.H. (1974), *The Spectral Analysis of Time Series*, Academic Press.

[32] Ljung, G. M. and Box, G. E. P. (1978), On a Measure of Lack of Fit in Time Series Models, *Biometrika*, 65, 297-303.

[33] Luati, A., Proietti, T. and Reale, M. (2012), The Variance Profile, *Journal of the American Statistical Association*, 107, 498, 607–621.

[34] Mardia, K., Kent, J. and Bibby, J. (1979). *Multivariate Analysis*, Academic Press.

[35] Milhøj, A. (1981), A Test of Fit in Time Series Models, *Biometrika*, 68, 1, 167–177.

[36] Miller, R.G. (1974), The Jackknife–A Review. *Biometrika*, 61, 1, 1–15.

[37] Percival, D.B. and Walden, A.T. (1993), *Spectral Analysis for Physical Applications*, Cambridge University Press.

[38] Pourahmadi, M. (2001), *Foundations of Time Series Analysis and Prediction Theory*, John Wiley and Sons.

[39] Priestley, M.B. (1981), *Spectral Analysis and Time Series*, John Wiley and Sons.

[40] Quenouille, M.H. (1949), Approximate Tests of Correlation in Time Series. *Journal of the Royal Statistical Society, Series B*, 11, 68-84.

[41] Taniguchi, M. (1980), On Estimation of the Integrals of Certain Functions of Spectral Density, *Journal of Applied Probability*, 17, 1, 73–83.

[42] Taniguchi, M. and Kakizawa, Y. (2000), *Asymptotic Theory for Statistical Inference of Time Series*, Springer.

[43] Tieslau, M.A., Schmidt, P. and Baillie, R.T. (1996), A Minimum Distance Estimator for Long-Memory Processes, *Journal of Econometrics*, 71, 249–264.

[44] Tukey, J. W. (1957), On the comparative anatomy of transformations, *Annals of Mathematical Statistics*, 28, 602–632.

[45] Walden, A. T. (1994), Interpretation of Geophysical Borehole Data via Interpolation of Fractionally Differenced White-Noise, *Journal of the Royal Statistical Society, Series C*, 43, 335–345.

[46] Whittle, P. (1961), Gaussian estimation in stationary time series, *Bulletin of the International Statistical Institute*, 39, 105–129.

[47] Woodward, W. A., Cheng, Q. C. and Gray, H. L. (1998), A k-Factor GARMA Long-memory Model. *Journal of Time Series Analysis*, 19, 485-504.

[48] Xia, Y., Tong, H. (2011), Feature Matching in Time Series Modeling. *Statistical Science*, 26, 1, 21-46.