

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

The generalised autocovariance function

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Tommaso Proietti, Alessandra Luati (2015). The generalised autocovariance function. JOURNAL OF ECONOMETRICS, 186(1), 245-257 [10.1016/j.jeconom.2014.07.004].

Availability:

This version is available at: <https://hdl.handle.net/11585/324315> since: 2016-01-08

Published:

DOI: <http://doi.org/10.1016/j.jeconom.2014.07.004>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

Proietti T, Luati A. The generalised autocovariance function. *J Econom.* 2015;186(1):245-257. doi:10.1016/j.jeconom.2014.07.004

The final published version is available online at:

<https://doi.org/10.1016/j.jeconom.2014.07.004>

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)

When citing, please refer to the published version.

The Generalised Autocovariance Function

Tommaso Proietti*

Department of Economics and Finance
University of Rome Tor Vergata, Italy
and CREATES, Aarhus University, Denmark

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 processes. Spectral estimation. White noise tests. Feature matching. Discriminant analysis.

*Corresponding author, address: Department of Economics and Finance, via Columbia 2, 00133, Rome, Italy.
Tel. +39 06 72595941. E-mail addresses: tommaso.proietti@uniroma2.it, alessandra.luati@unibo.it

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 extend 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. Finally, by solving a generalised Yule-Walker system of equations based on the GACV, we can estimate a general spectral model that, according to the value of p , encompasses both autoregressive and moving average spectral models.

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 autoregressive (AR) and moving average (MA) 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. It emerges that fractional values of p in the $(0,1)$ interval provide finite sample tests of the white noise hypothesis with improved size properties. A generalised 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, $F(\omega) = \int_{-\pi}^{\omega} f(\lambda) d\lambda$, that it is positive, that the process is regular (Doob, 1953, p. 564), i.e. $\int_{-\pi}^{\pi} \log f(\omega) d\omega > -\infty$, and that $\int_{-\pi}^{\pi} f(\omega)^p d\omega < \infty$.

The generalised autocovariance (GACV) function is defined as the inverse Fourier transform

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

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

where we have replaced $\exp(i\omega k)$ 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

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

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; see also Battaglia, 1983).

Other examples where integrals of powers of the spectral density function are of relevant interest may be found in recent advances on probabilistic approximations to Gaussian processes. As an example, for $k = 0$ and integer $p > 0$, the generalised variance gives the p -th cumulant of the sample variance of a stationary zero mean stochastic process (see Nourdin and Peccati, 2012, formula 7.2.2). In turn, the cumulants enter in Berry-Esseen type bounds for the distance between the normalised asymptotic distribution of the sample variance and the standard Gaussian distribution (see Nourdin and Peccati, 2012, formulae 7.3.1 and 9.5.1, which involve fractional powers of the generalised variance).

As defined in (1) and due to the Herglotz theorem, the GACV is a true autocovariance (see also section 3), and as such it inherits all the well known properties of an autocovariance function: an obvious property is symmetry with respect to the lag, $\gamma_{pk} = \gamma_{p,-k}$; 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 $f(\omega)$. The generalised autocorrelation function (GACF) is defined as

$$\rho_{pk} = \frac{\gamma_{pk}}{\gamma_{p0}}, \quad k = 0, \pm 1, \pm 2, \dots, \quad (3)$$

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

Further 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). Then,*

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

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 $p = q$ in Lemma 1:

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

which for $k = 0$ specialises as

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

Furthermore, setting $q = -p$ in Lemma 1, we obtain

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

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.

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

$$x_t = \psi(B)\xi_t, \quad (7)$$

where $\xi_t \sim \text{WN}(0, \sigma^2)$ and $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$, 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, $\text{WN}(0, \sigma^2)$ 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 (7) is $\gamma_k = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}$ for $k = 0, 1, \dots$, and $\gamma_{-k} = \gamma_k$.

Let us consider the power-transformed process:

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

For a real $p > 0$, 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

$$\varphi_j = \frac{1}{j} \sum_{h=1}^j [h(p+1) - j] \psi_h \varphi_{j-h}, \quad j > 0, \quad \varphi_0 = 1 \quad (9)$$

(Gould, 1974). When $p < 0$, the coefficients of $\psi(B^{-1})^p = \sum_{j=0}^{\infty} \varphi_j B^{-j}$ are also given by (9).

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

$$2\pi f_u(\omega) (\sigma^2)^{p-1} = [2\pi f(\omega)]^p. \quad (10)$$

Equation (10) establishes the relation between the spectrum of the original process and that of the power process u_{pt} . Section 7 proposes a method of moments estimator of $f_u(\omega)$ based on a generalised Yule-Walker system of equations.

It follows from (1) and (10) 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,

$$\gamma_{pk} = \sigma^{2p} \sum_{j=0}^{\infty} \varphi_j \varphi_{j+k}. \quad (11)$$

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 :

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

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 $\text{Var}(x_t | \mathcal{F}_{\setminus t})$, 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 $p \rightarrow 0$ yields the prediction error variance, $\lim_{p \rightarrow 0} v_p = \sigma^2$, which is the geometric average of the spectral density, $\sigma^2 = \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log 2\pi f(\omega) d\omega \right\}$, i.e. the Szegő-Kolmogorov formula.

4 Illustrations

We illustrate how the auxiliary power process is used to derive and characterise the GACV and GACF of important classes of stationary processes.

In the general case when x_t is a stationary autoregressive fractionally integrated moving average process, an ARFIMA(r, d, q) process, the GACV is given by

$$\gamma_{pk} = \frac{\sigma^{2p}}{2\pi} \int_{-\pi}^{\pi} |1 - e^{-i\omega}|^{-2pd} \frac{|\theta_q(e^{-i\omega})|^{2p}}{|\phi_r(e^{-i\omega})|^{2p}} \cos(\omega k) d\omega. \quad (13)$$

The solution of the above integral is obviously not straightforward. Nevertheless, by exploiting the relation between the original process x_t and the auxiliary process u_{pt} , we are able to get explicit solutions of (13) for fixed values of r, d, q . The key arguments are equations (8), (9) and (11). In the case of AR and MA processes, the next sections aim at illustrating the methods that can be used to obtain closed form solutions for the GACV and GACF of a given process. When closed form expressions are not straightforward to obtain, like in the case of ARMA processes, our suggestion is to truncate the infinite summation (11) or to numerically solve the integral (13) for fixed values of r and q and for $d = 0$. In the case of long memory processes, we shall be able to provide some results on the summability of the GACV.

4.1 The GACV of MA(q) processes

Let us consider the invertible MA(q) process, $x_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \xi_t$, where $\xi_t \sim \text{WN}(0, \sigma^2)$. The GACV of the MA(q) process is (11), with $\varphi_0 = 1$ and $\varphi_j = \frac{1}{j} \sum_{h=1}^j [h(p+1) - j](-\theta_h) \varphi_{j-h}$, $j > 0$.

When p is a positive integer, given that $\theta_j = 0$ for $j > q$, the GACV of x_t becomes zero after the pq -th lag and in this very specific case the power process turns out to be a MA(pq) process. In all the other cases, involving negative and/or fractional values of p , the GACV of a MA(q) process is never zero.

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

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

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:

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

4.2 The GACV of AR(r) processes

We shall start by considering a stationary AR(1) process, that corresponds to the case when $q = d = 0$ and $r = 1$ in (13), i.e. $x_t = (1 - \phi B)^{-1}\xi_t$, $|\phi| < 1$, $\xi_t \sim \text{WN}(0, \sigma^2)$. The generalised autocovariance function of this process is given by

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

The power process associated with $x_t \sim \text{AR}(1)$, e.g. for $p > 0$, 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 $\varphi_j = \frac{(-\phi)^j}{j!}(-p)(-p-1)(-p-2)\dots(-p-j+1) = (-\phi)^j \binom{-p}{j}$, where we have used $\binom{p}{k} = \frac{p(p-1)(p-2)\dots(p-k+1)}{k(k-1)(k-2)\dots 1}$, that holds with initial conditions $\binom{p}{0} = 1, \binom{p}{1} = p$ (see Graham, Knuth and Patashnik, 1994, ch. 5); note that for $p = 0$, $\psi_j = 0$ for all j , since $\binom{0}{j} = 0$. By (11), the GACV of x_t is

$$\gamma_{pk} = \sigma^{2p} (-\phi)^k \sum_{j=0}^{\infty} \phi^{2j} \binom{-p}{j} \binom{-p}{j+k}, \quad (15)$$

with

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

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

The GACF for $k > 0$ is

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

The above expressions hold also for $p < 0$. Recalling that for integer $p < 0$, $\binom{-p}{j+k} = 0$ for all $k > -p-j$, we have that $\gamma_{-1,k} = 0$ for $|k| > 1$, whereas, $\gamma_{-1,0} = (1+\phi^2)/\sigma^2$ and $\gamma_{-1,1} = -\phi/\sigma^2$, so that only the first inverse autocorrelation is nonzero and equal to $-\phi/(1+\phi^2)$. When $p = -2$, $\gamma_{-2,0} = (1+4\phi^2+\phi^4)/\sigma^4$, $\gamma_{-2,1} = (-2\phi)(1+\phi^2)/\sigma^4$, $\gamma_{-2,2} = \phi^2/\sigma^4$, $\gamma_{-2,k} = 0, k > 2$. Note that, as implied by Lemma 1, $\gamma_{-2,0} = \gamma_{-1,0}^2 + 2\gamma_{-1,1}^2$.

More generally, for integer $p < 0$ the GACV (GACF) of an AR(r) process is identically zero for all lags $|k| > -pr$, which follows by the same considerations holding for a MA(q) process and positive integer power parameter. On the other hand, for fractional and/or negative value of p , the GACV (GACF) of an AR(r) process is never zero.

Equations (14) and (15) generalise to any fractional p equations 3.616.4 and 3.616.7 of Gradshteyn and Ryzhik (1994), that hold for AR(1) and MA(1) processes in the case of a positive integer power p .

4.3 Long memory processes

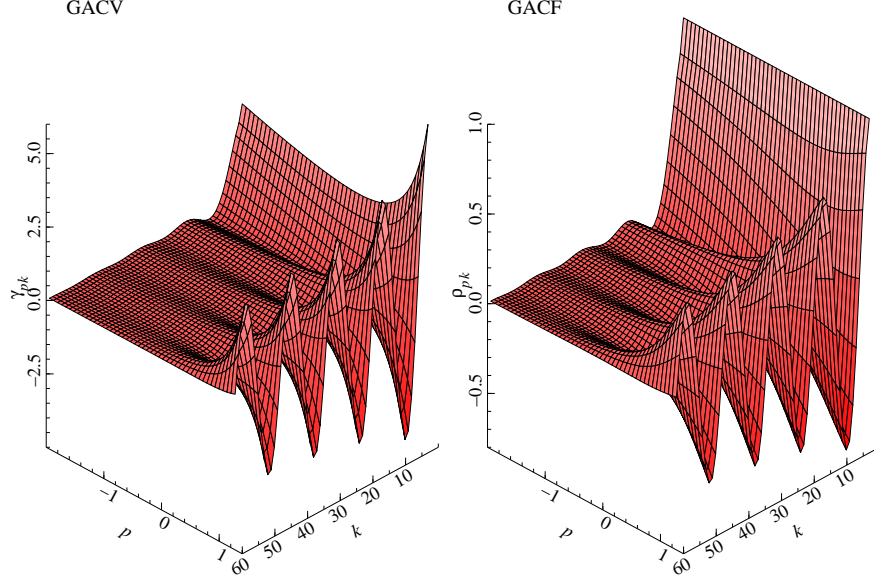
For the fractional noise (FN) process, $(1-B)^d x_t = \xi_t$, where $\xi_t \sim \text{WN}(0, \sigma^2)$, $d < 0.5$, the GACV and GACF are defined for $pd < 0.5$ and are given respectively by

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

This is easily established from the autocovariance of u_{pt} , which is a FN process with memory parameter pd . If $p < 0.25/d$, the GACV of a FN process with $0 < d < 0.5$ is square-summable: as a matter of fact, as implied by Lemma 1, $\sum_{k=-\infty}^{\infty} \gamma_{pk}^2 = \gamma_{2p,0} = \sigma^{4p} \frac{\Gamma(1-4dp)}{[\Gamma(1-2dp)]^2}$ is finite for $2pd < 0.5$.

For $p = -1/d$, $\rho_{p1} = -0.5, \rho_{pk} = 0, k = 2, \dots$, as u_{pt} has a non-invertible MA(1) representation.

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



Let us consider the Gegenbauer process

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

where $\xi_t \sim \text{WN}(0, \sigma^2)$; $\nu = \cos \lambda$, 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. In the sequel we consider the case $|\nu| < 1$. 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 $p < 0.5/d$ it converges at a faster rate than $p = 1$. Indeed, for the Gegenbauer process the GACV is square summable for $p < 0.25/d$: by Lemma 1,

$$\sum_{k=-\infty}^{\infty} \gamma_{pk}^2 = \gamma_{2p,0} = \sigma^4 \sum_{j=0}^{\infty} G_j^2(\nu, 2pd)$$

and the square summability of the coefficients $G_j(\nu, 2pd)$ holds for $2pd < 0.5$.

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 $k = 1, 2, \dots, 60$. For $p = -1/d$, $\rho_{pk} = 0, k = 3, \dots$, as u_{pt} has 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, \dots, x_n) ,

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

evaluated at the Fourier frequencies $\omega_j = \frac{2\pi j}{n} \in (0, \pi)$, $1 \leq j \leq \lfloor \frac{n-1}{2} \rfloor$, where $\lfloor \cdot \rfloor$ is the largest integer not greater than the argument.

Specifically, estimation of γ_{pk} , as defined in (1), will be based on the nonparametric estimator defined as follows. Let $M = \lfloor (n-1)/(2m) \rfloor$, then

$$\hat{\gamma}_{pk} = \frac{1}{M} \sum_{j=0}^{M-1} Y_j^{(p)} \cos(\bar{\omega}_j k), \quad (16)$$

where

$$Y_j^{(p)} = (2\pi \bar{I}_j)^p \frac{\Gamma(m)}{\Gamma(m+p)} \quad (17)$$

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; m will be referred to as the pooling parameter.

The estimator (16) 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. The latter was introduced to improve the bias-variance trade off of 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 asymptotic properties of the estimator (16) are established for the linear causal process with independent and identically distributed (IID) errors,

$$x_t = \sum_{j=0}^{\infty} \psi_j \xi_{t-j}, \quad \xi_t \sim \text{IID}(0, \sigma^2),$$

based on the following assumptions:

- A1. $\sum_{j=0}^{\infty} j^{\delta} |\psi_j| < \infty$, $\delta > \frac{3}{4}$
- A2. $E(|\xi_t|^s) = \mu_s < \infty$, where $s > \max\{9, 4(p+m), 8(p-1) + 2m, 2(p-1) + 4m\}$
- A3. $\sup_{|\theta| \geq \theta_0} |E(e^{i\theta \xi_t})| = \delta(\theta_0) < 1$
- A4. $\int_{-\infty}^{\infty} |E(e^{i\theta \xi_t})|^r d\theta < \infty$ for some $r \geq 1$
- A5. m is a fixed positive integer, satisfying $m + 4(p-1) > 0$

Theorem 1 *Let $\{x_t\}_{t \in T}$ be the linear process $x_t = \sum_{j=0}^{\infty} \psi_j \xi_{t-j}$, $\xi_t \sim \text{IID}(0, \sigma^2)$, with generalised autocovariance function γ_{pk} defined as in (1). Let us denote the vector of generalised autocovariance functions up to lag K as $\gamma_p = [\gamma_{p0}, \gamma_{p1}, \dots, \gamma_{pK}]'$ and the corresponding estimator with elements given by (16) as $\hat{\gamma}_p = [\hat{\gamma}_{p0}, \hat{\gamma}_{p1}, \dots, \hat{\gamma}_{pK}]'$. Under assumptions A1-A5, $\hat{\gamma}_p \rightarrow_p \gamma_p$ and*

$$\sqrt{n} (\hat{\gamma}_p - \gamma_p) \rightarrow_d N(0, \mathbf{V}), \quad (18)$$

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

$$v_{kl} = m(C(m; p, p) - 1) \frac{1}{\pi} \int_{-\pi}^{\pi} [2\pi f(\omega)]^{2p} \cos(\omega k) \cos(\omega l) d\omega + p^2 \kappa_4 \gamma_{pk} \gamma_{pl}, \quad (19)$$

κ_4 is the fourth cumulant of ξ_t and

$$C(m; p, q) = \frac{\Gamma(m + p + q) \Gamma(m)}{\Gamma(m + p) \Gamma(m + q)}. \quad (20)$$

The proof, given in Appendix A.1, is based on the Bartlett (1955) decomposition of the periodogram of a linear process and is made up of two steps. The first step consists in proving a central limit theorem for the powers of the periodogram of the IID noise process, based on the method of moments and on Edgeworth expansions, as in Faÿ and Soulier (2001). The second step consists in proving the negligibility of the term involving the remainder of the Bartlett decomposition and is based on bounds in probability. This is the method used by Faÿ, Moulines and Soulier (2002) to prove a central limit theorem for non linear functionals of the periodogram of a linear process, which can be applied when some regularity conditions on the functional of the spectrum and on the distribution of the noise process are satisfied. In particular, the Cramér condition A3 and condition A4 ensure the existence of a bounded density and they are satisfied by a broad class of continuous distributions. Given the regularity conditions A3 and A4, assumptions A2 and A5 guarantee the validity of the Edgeworth expansions that are needed for the two steps of the proof. The proof of the negligibility of the remainder term also requires the strong assumption A1 on the Wold coefficients.

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

$$v_{kl}^{(pq)} = m(C(m; p, q) - 1) \frac{1}{\pi} \int_{-\pi}^{\pi} [2\pi f(\omega)]^{p+q} \cos(\omega k) \cos(\omega l) d\omega + pq \kappa_4 \gamma_{pk} \gamma_{ql}. \quad (21)$$

For $m = p = 1$ the estimator (16),

$$\hat{\gamma}_{1k} = \frac{1}{\lfloor (n-1)/2 \rfloor} \sum_{j=1}^{\lfloor (n-1)/2 \rfloor} 2\pi I(\omega_j) \cos(\omega_j k),$$

is the Riemannian sum approximation over the Fourier frequencies of the (time domain) sample autocovariance at lag k , $\tilde{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$, for $k = 0, \dots, n-1$ and $\tilde{\gamma}_{-k} = \tilde{\gamma}_k$. In fact, $\tilde{\gamma}_k = \int_{-\pi}^{\pi} I(\omega) \cos(\omega k) d\omega$, where $I(\omega) = \frac{1}{2\pi} \sum_{|h| < n} \tilde{\gamma}_h \cos(\omega h)$.

Moreover, equation (19) encompasses the formula for the generic element of the asymptotic covariance matrix of the sample covariance $\tilde{\gamma}_k$. In fact, by Lemma 1, equation (4), and by trigonometric identities (see also Kakizawa and Taniguchi, 1995), equation (19) can be written as

$$v_{kl} = \sum_{j=-\infty}^{\infty} (\gamma_{p,j+k}\gamma_{p,j+l} + \gamma_{p,j+k}\gamma_{p,j-l}) + \kappa_4\gamma_k\gamma_l \quad (22)$$

which for $m = 1$ and $p = 1$ coincides with the asymptotic covariance between $\tilde{\gamma}_k$ and $\tilde{\gamma}_l$.

In the case of a Gaussian process assumptions A2-A4 are not needed, while assumptions A1 and A5 may be weakened, so that Theorem 1 is modified as follows.

Theorem 2 *Let $\{x_t\}_{t \in T}$ be the linear process $x_t = \sum_{j=0}^{\infty} \psi_j \xi_{t-j}$, $\xi_t \sim \text{IID } N(0, \sigma^2)$, with $\sum_{j=0}^{\infty} j^{\delta'} |\psi_j| < \infty$ where $\delta' = \frac{1}{2} + \epsilon$, $\epsilon > 0$ and $m > -2p$. The results of Theorem 1 then hold with $\kappa_4 = 0$.*

When the process $\{\xi_t\}_{t \in T}$ is a Gaussian white noise, then a central limit theorem for linear combinations of independent random variables can be directly applied (e.g. Gleser, 1965). The condition imposed to the pooling parameter serves to ensure the existence of the second moment of the p -th power of a gamma random variable with parameter $(m, 1)$. On the other hand, the condition on the memory of the process can be relaxed because of the weaker condition required for proving the negligibility of the remainder term (see Walker, 1965). Theorem 2 extends to the case of a negative power the results of Taniguchi (1980). In addition, our result embodies a finite sample bias correction and establishes a lower bound for m in the case of a negative p .

Under the assumptions of Theorem 1, similar results can be derived for the generalised autocorrelation function, that is estimated based on (16) by

$$\hat{\rho}_{pk} = \frac{\hat{\gamma}_{pk}}{\hat{\gamma}_{p0}}. \quad (23)$$

Theorem 3 *Let us consider the vectors $\boldsymbol{\rho}_p = [\rho_{p1}, \rho_{p2}, \dots, \rho_{pK}]'$ and $\hat{\boldsymbol{\rho}}_p = [\hat{\rho}_{p1}, \hat{\rho}_{p2}, \dots, \hat{\rho}_{pK}]'$ having components as in (3) and (23), respectively. Under the assumptions of Theorem 1,*

$$\sqrt{n} (\hat{\boldsymbol{\rho}}_p - \boldsymbol{\rho}_p) \rightarrow_d N(\mathbf{0}, \mathbf{W}) \quad (24)$$

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

$$w_{kl} = \sum_{j=-\infty}^{\infty} (\rho_{p,j+k}\rho_{p,j+l} + \rho_{p,j+k}\rho_{p,j-l} + 2\rho_{p,k}\rho_{p,l}\rho_{p,j}^2 - 2\rho_{p,k}\rho_{p,j}\rho_{p,j+l} - 2\rho_{p,l}\rho_{p,j}\rho_{p,j+k}). \quad (25)$$

The proof, given in Appendix A.3, is based on the same arguments that lead to the proof of the traditional Bartlett's formula ($p = 1$). The asymptotic covariance matrix is estimated by replacing $\hat{\boldsymbol{\rho}}_p$ for the population quantities into the expression for \mathbf{W} .

Formula (25) generalises Bartlett's formula to any real p satisfying the assumptions of Theorems 1 and 2. Recently, the traditional Bartlett's formula has been extended to an important class of nonlinear processes exhibiting conditional heteroscedasticity or stochastic volatility by Francq and Zakoïan (2009) and Kokoszka and Politis (2011). We think that extending our results to the nonlinear case is an interesting direction for future research.

The pooling parameter m is an important ingredient in the design of the GACV estimator (16). First and foremost, pooling is necessary as we can estimate the GACV only for $p > -m/2$ in the Gaussian case and for $p > 1 - m/4$ in the more general IID case; for instance, the estimation of the inverse autocovariance function by $\hat{\gamma}_{-1,k}$ requires $m \geq 3$ in the Gaussian case and $m \geq 9$ in the IID case.

In finite samples, the mean square errors of the GACV and GACF estimators are a rather complicated function of p , m , and the spectral properties of x_t . In selecting the value of the pooling parameter, we face a bias-variance trade-off. As m increases, the bias of the estimator $\hat{\gamma}_{pk}$ will increase, as the estimation error arising from approximating an integral by a Riemann sum will be larger. The asymptotic variance of the GACV estimator depends on m via the factor $m[C(m; p, p) - 1]$. The latter decreases with m at a rate varying with p . Table 1 displays its values for some combinations of m and p . It is clear that pooling achieves an important variance reduction when p is negative.

Table 1: Values of the factor $m[C(m; p, p) - 1]$.

	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 7$	$m = 9$	$m = 11$	$m = 15$	$m = 20$	$m = 30$
$p = -2$	-	-	-	-	25.00	10.50	7.80	6.68	5.68	5.15	4.70
$p = -1$	-	-	3.00	2.00	1.67	1.40	1.29	1.22	1.15	1.11	1.07
$p = -.5$	-	0.55	0.40	0.35	0.32	0.30	0.29	0.28	0.27	0.26	0.26
$p = .5$	0.27	0.26	0.26	0.26	0.26	0.25	0.25	0.25	0.25	0.25	0.25
$p = 1$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$p = 2$	5.00	4.67	4.50	4.40	4.33	4.25	4.20	4.17	4.12	4.10	4.06

Luati, Proietti and Reale (2012) have proposed the use of the the jackknife (Quenouille, 1949, see Miller, 1974, and Efron and Tibshirani, 1993, for reviews) for selecting the optimal value of 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 periodogram scaled by the fitted spectrum, they serve as goodness of fit tests.

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

$$\text{BP}_p = n \sum_{k=1}^K \hat{\rho}_{pk}^2. \quad (26)$$

By Theorem 2, (26) provides an asymptotically χ_K^2 test, generalising the Box-Pierce (1970) statistic $\text{BP} = n \sum_{k=1}^K \tilde{\rho}_k^2$ 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 $\text{LB} = n(n+2) \sum_{k=1}^K (n-k)^{-1} \tilde{\rho}_k^2$, known as the Ljung-Box (1978) statistic, is also possible. It should be noticed that BP_1 is not identical to the traditional time domain BP statistic, not even when $m = 1$, as $\hat{\rho}_{1k} \neq \tilde{\rho}_k$ (see section 5), although the differences tend to be negligible as the sample size increases.

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 $\hat{\mathcal{R}}_p = \frac{\hat{\gamma}_{2p,0}}{\hat{\gamma}_{p0}^2}$, whose null distribution has variance

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

Hence, the test statistic

$$\mathcal{M}_p = \frac{\hat{\mathcal{R}}_p - 1}{\sqrt{\text{Var}(\hat{\mathcal{R}}_p)}} \quad (27)$$

provides an asymptotically standard normal test.

The test (27) 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 (27) 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 2 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, BP 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.

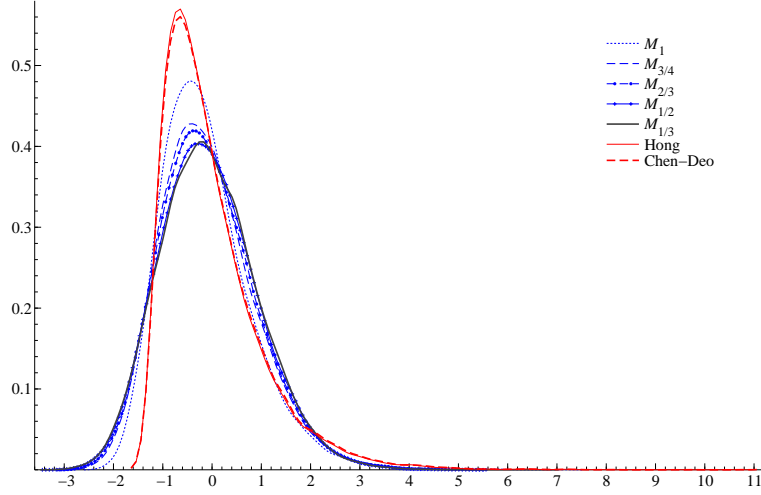
The size properties of the BP_p test are marginally improved by choosing $p < 1$. On the other hand, as far as \mathcal{M}_p is concerned, having $p < 1$ yields more substantial gains. The heuristic explanation is that in finite sample fractional values of p have a normalisation effect on the distribution. It should be recalled that the cube root transformation was already proposed by Wilson and Hilferty (1931) as the normalising transformation for a $\chi_{2m}^2/2$ random variable. Hernandez and Johnson (1980, p. 857-858), later confirmed that if X is a Gamma random variable with shape parameter equal to m , X^p with $p = 1/3$ provides the best approximation to $Y \sim N(\Gamma(m+p)/\Gamma(m), \Gamma(m+2p)/\Gamma(m) - \Gamma^2(m+p)/\Gamma^2(m))$, in the sense of minimising the Kullback-Leibler discrepancy between the two probability density functions. Krishnamoorthy, Mathew and Mukherjee (2008) show the effectiveness of the cube root transformation for constructing prediction intervals of gamma random variables based on the normal approximation. 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.

As for the effect of varying p on the power of the test \mathcal{M}_p , it may be argued that for small p the test will lose power due to the flattening of the periodogram ordinates. To investigate this

Table 2: 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
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
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
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
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
	$m = 1$				$m = 3$				$m = 5$			
	10%	5%	1%	Dist	10%	5%	1%	Dist	10%	5%	1%	Dist
$\mathcal{M}_{1/3}$	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
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
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
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
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
	$m = 1$				$m = 3$				$m = 5$			
	10%	5%	1%	Dist	10%	5%	1%	Dist	10%	5%	1%	Dist
$\mathcal{M}_{1/3}$	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 $n = 128, m = 1, K = 8$.

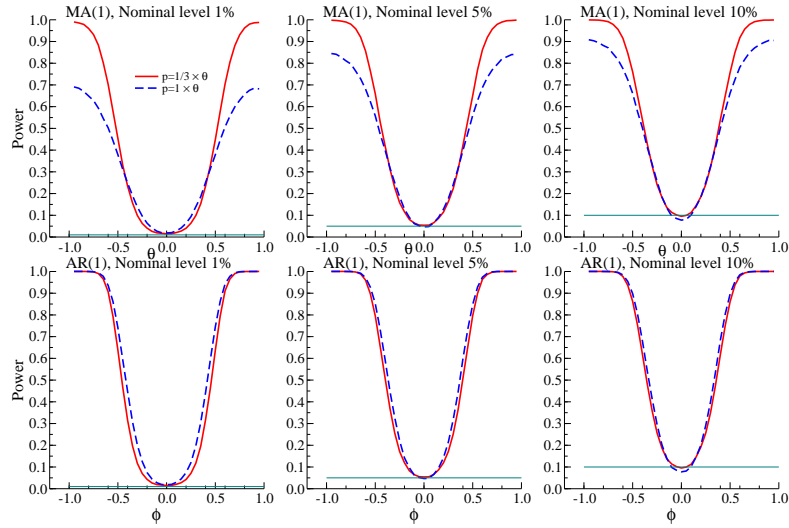


issue we compared the power of $\mathcal{M}_{1/3}$ and \mathcal{M}_1 in the case $m = 1$ by a Monte Carlo simulation according to which 50,000 series of length $n = 128$ were alternatively generated by an MA(1) Gaussian process with parameter $\theta \in (-1, 1)$ and by an AR(1) Gaussian process with parameter $\theta \in (-1, 1)$. The results are presented in figure 3, which shows the powers of the tests conducted using three nominal sizes (1%, 5% and 10%). Interestingly, in the MA case the test statistic for $p = 1/3$ has excellent power properties and clearly outperforms the \mathcal{M}_1 test for values $|\theta| > 0.5$. In the AR(1) case the loss in power is imperceptible. This confirms that choosing p about $1/3$ results in a generalised Milhøj test statistic with improved finite sample properties.

7 A generalised Yule-Walker spectral estimator based on the GACV

The Yule-Walker (YW) method has a long tradition in time series for estimating the parameters of an autoregressive process (see e.g. Brockwell and Davis, 1991, chapter 8, and Percival and Walden, 1993, chapter 9). 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.

Figure 3: Comparison of the power functions of the tests $\mathcal{M}_{1/3}$ (solid line) and \mathcal{M}_1 (dashed line), when the true model is MA(1), $x_t = \xi_t - \theta\xi_{t-1}$, $\xi_t \sim \text{IID } N(0, 1)$, (upper panels, corresponding to three sizes of the tests, 1%, 5% and 10%), and when the true model is AR(1), $x_t = \phi x_{t-1} + \xi_t$, $\xi_t \sim \text{IID } N(0, 1)$, $\phi \in (-1, 1)$. The results refer to the case $n = 128, m = 1$, and are based on 50,000 Monte Carlo replications. The straight lines are drawn at the size of the tests.



We consider here a generalisation of the Yule-Walker method based on the GACV for different values of p , that fits the following class of models for the spectrum of a stationary time series

$$2\pi f(\omega) = \left[\frac{\sigma_p^2}{\phi_p(e^{-i\omega})\phi_p(e^{i\omega})} \right]^{1/p}, \quad (28)$$

where $\phi_p(e^{-i\omega}) = 1 - \phi_{p1}e^{-i\omega} - \phi_{p2}e^{-i2\omega} - \dots - \phi_{pK}e^{-iK\omega}$. As p varies, (28) 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 $\{\phi_{pk}, k = 1, \dots, K\}$ satisfy the following moment conditions (YW equations): $\mathbf{\Gamma}_{p,K}\phi_{p,K} = \gamma_{p,K}$, where $\mathbf{\Gamma}_{p,K}$ denotes the Toeplitz matrix, formed from the GACV, with generic element $\gamma_{p,|h-k|}$, $h, k = 0, \dots, K-1$, $\gamma_{p,K} = (\gamma_{p1}, \dots, \gamma_{pK})'$, and $\phi_{p,K} = (\phi_{p1}, \dots, \phi_{pK})'$; moreover, $\sigma_p^2 = \gamma_{p0} - \phi_{p,K}'\gamma_{p,K}$.

For $p > 0$, the polynomial $\phi_p(B) = 1 - \phi_{p1}B - \dots - \phi_{pK}B^K$ characterises the AR approximation of the process u_{pt} , and provides the spectral factorisation $[2\pi f(\omega)]^p = \sigma_p^2[\phi_p(e^{-i\omega})\phi_p(e^{i\omega})]^{-1}$. By (9) 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} .

Given a time series realisation, we replace the theoretical GACF with the estimated one to get $\hat{\phi}_{p,K} = \hat{\mathbf{\Gamma}}_{p,K}^{-1}\hat{\gamma}_{p,K}$, and by applying (9) we obtain different estimates for the spectrum of the time series according to the value of p , $\hat{f}_p(\omega)$. Also, $\hat{\sigma}_p^2 = \hat{\gamma}_{p0} - \hat{\phi}_{p,K}'\hat{\gamma}_{p,K}$.

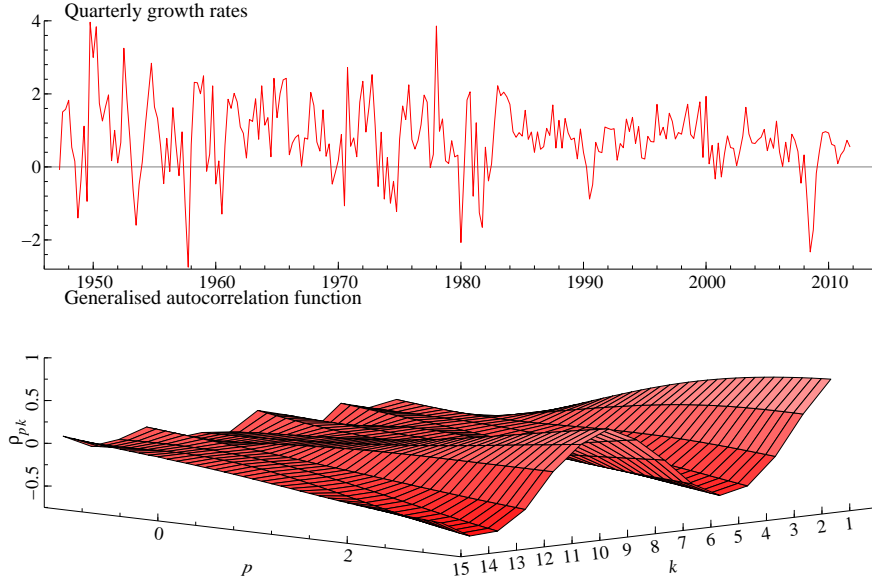
If the process u_{pt} is autoregressive of order K , under the conditions of Theorem 1, $\hat{\phi}_{p,K} \rightarrow_p \phi_{p,K}$, and $\sqrt{n}(\hat{\phi}_{p,K} - \phi_{p,K}) \rightarrow_d N(\mathbf{0}, \sigma_p^2 \mathbf{\Gamma}_{p,K}^{-1})$ (see Brockwell and Davis, 1991, Theorem 8.1.1). Furthermore, as it is well known, the YW estimator coincides with the Whittle estimator of the parameters of an AR(K) process (see, e.g. Dahlhaus, 1988); this result extends to our situation: the generalised YW estimator is equivalent to the Whittle estimator of $(\phi_{p,K}, \sigma_p^2)$ of the AR(K) model for u_{pt} . Notice that the latter is also equivalent to the frequency domain maximum likelihood estimator based upon the distributional assumption

$$\frac{(2\pi\bar{I})^p}{\sigma_p^2} |\phi_p(e^{-i\omega})|^2 \sim X_j^p,$$

where $X_j \sim \frac{1}{2}\chi_{2m}^2$ and χ_{2m}^2 denotes a chi-square random variable with $2m$ degrees of freedom.

The scope of the above generalised YW estimation method is twofold: on the one hand, it is used for extracting features of interest from a time series in much the same spirit as Xia and Tong (2011): values of p larger than 1 will be useful for the detection of spectral peaks, and amount to boosting the cyclical features of the series, as large periodogram ordinates will receive

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

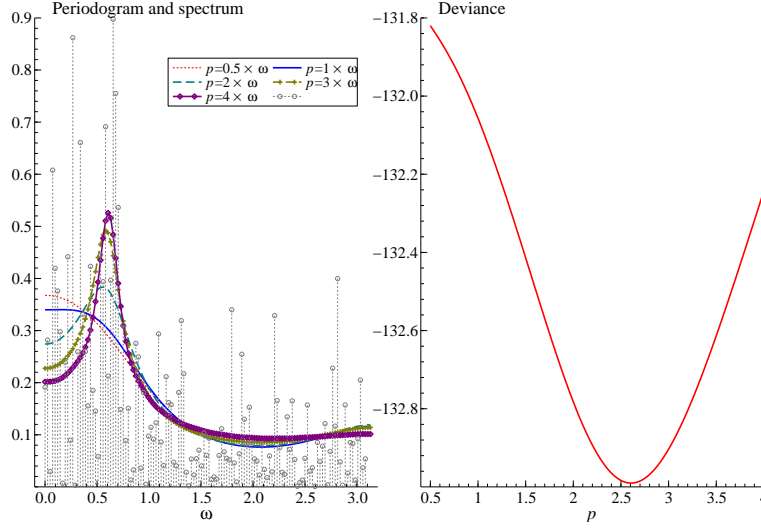


an higher weight; on the contrary, negative values will emphasize the troughs in the spectra. Also, as p varies, a class of spectral fits is obtained and the analyst can select the one optimising a particular criterion function, as for instance a measure of goodness of fit.

Hereby, we provide an illustration of both uses. Figure 4 displays the quarterly growth rate of the US Gross Domestic Product (1947.2-2012.1), hereafter GDP, 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.

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.

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



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 $\tilde{p} = 2.65$.

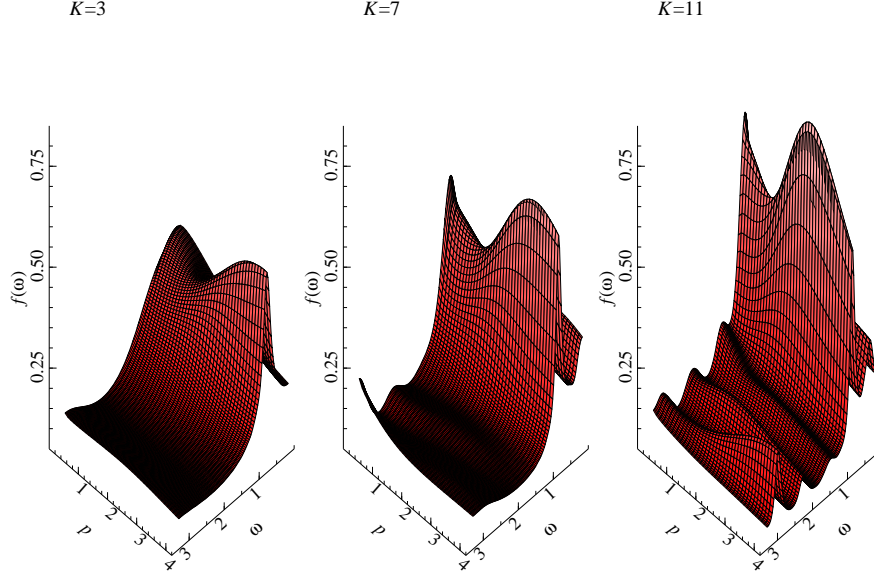
The spectral peak, corresponding to a period of roughly 2.5 years (10 quarters), can alternatively be identified by increasing the AR order, as figure 6 shows, but there is a risk of overfitting the sample spectrum in other frequency ranges.

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

$$\begin{aligned} d_{ij,p}^2 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \{ [2\pi f_i(\omega)]^p - [2\pi f_j(\omega)]^p \}^2 d\omega \\ &= \sum_{k=-\infty}^{\infty} (\gamma_{i,pk} - \gamma_{j,pk})^2 \\ &= \gamma_{i,2p,0} + \gamma_{j,2p,0} - 2 \sum_{k=-\infty}^{\infty} \gamma_{i,pk} \gamma_{j,pk}. \end{aligned} \tag{29}$$

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.



The p -sd (29) 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 (29).

The p -sd can be estimated by

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

or, if the autocorrelations are used,

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

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 p 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.

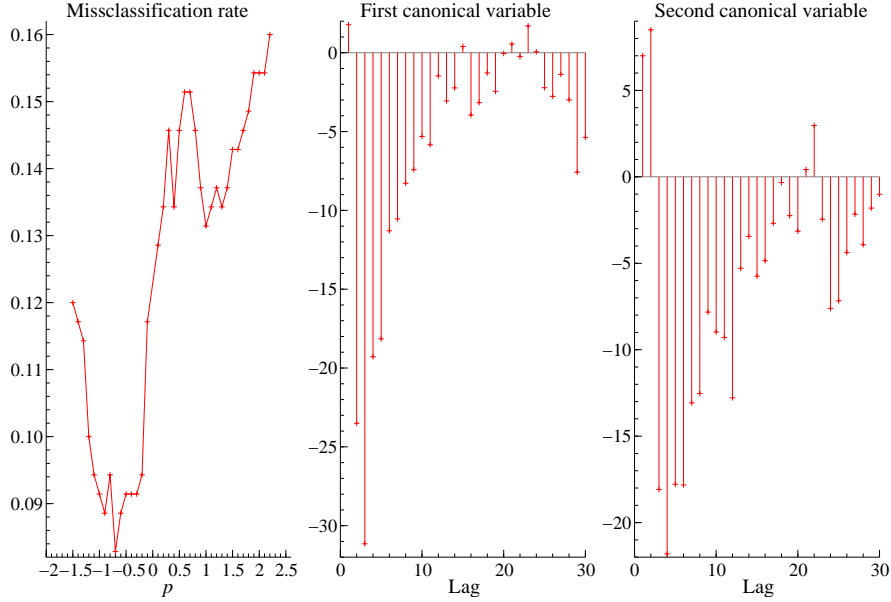
For the training sample the two canonical variates are obtained from the generalised eigenvectors of the between groups deviance matrix, \mathbf{B} , satisfying $\mathbf{B}\mathbf{a} = \lambda\mathbf{a}$, $\mathbf{a}'\mathbf{W}\mathbf{a} = 1$, where \mathbf{W} is the within groups deviance matrix and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

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



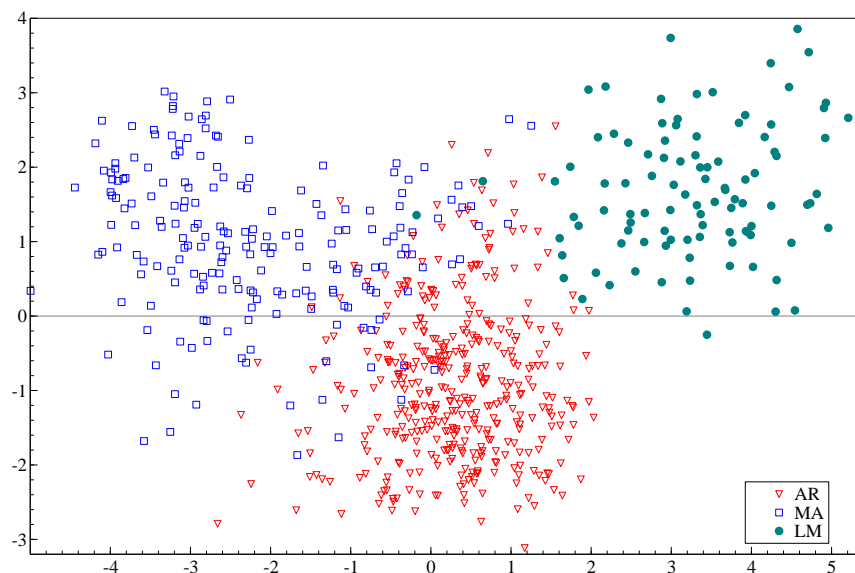
9 Conclusions

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 on 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 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

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



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).

Acknowledgements

The paper was presented at the University of Sydney, University of Technology Sydney, University of New South Wales, University of St Andrews, University of Heidelberg, Collegio Carlo Alberto, CREATES Aarhus, RCEA Rimini, and COMPSTAT.

The authors wish to thank Francesco Battaglia, Bob Bartels, Rainer Dahlhaus, Domenico Marinucci, Victor Solo, the Editor, Prof. Jianqing Fan, an Associate Editor and three Referees, for their helpful comments and suggestions that led to several improvements in this article.

The authors gratefully acknowledge financial support by the Italian Ministry of Education, University and Research (MIUR), PRIN Research Project 2010-2011 - prot. 2010J3LZEN, "Forecasting economic and financial time series". Tommaso Proietti gratefully acknowledges support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

A Appendix

A.1 Proof of Theorem 1

Let us consider the sequence

$$S_n(Y^{(p)}) = \sum_{j=0}^{M-1} b_j \left\{ Y_j^{(p)} - (2\pi f(\bar{\omega}_j))^p \right\}, \quad (30)$$

where $b_j = \cos(\bar{\omega}_j k) \left(\sum_l \cos^2(\bar{\omega}_l k) \right)^{-\frac{1}{2}}$, $\sum_j b_j^2 = 1$ and $Y_j^{(p)}$ is defined in equation (17). As $\sum_j \cos^2(\bar{\omega}_j k) = O(M/2)$, $|b_j| = O(M^{-1/2})$ ($\cos(\bar{\omega}_j k)$ is bounded). Thus, $\sum_j |b_j| = O(M^{1/2})$,

and if m is a fixed integer, then $\sum_j |b_j| = 0(n^{1/2})$. In the sequel, we shall denote $\bar{b} = \max_{0 \leq j \leq (M-1)} |b_j|$, for which $\lim_{M \rightarrow \infty} \bar{b} = O(M^{-1/2})$, so that $M\bar{b}^2 = O(1)$.

Let us define $\tilde{Y}_j^{(p)}$ and $S_n(\tilde{Y}^{(p)})$ as

$$\tilde{Y}_j^{(p)} = (2\pi f(\bar{\omega}_j))^p \frac{\Gamma(m)}{\Gamma(m+p)} \frac{(2\pi \bar{I}_{\xi j})^p}{\sigma^{2p}},$$

and

$$S_n(\tilde{Y}^{(p)}) = \sum_{j=0}^{M-1} b_j \left\{ \tilde{Y}_j^{(p)} - (2\pi f(\bar{\omega}_j))^p \right\}, \quad (31)$$

where $\bar{I}_{\xi j} = \sum_{l=1}^m I_{\xi}(\omega_{jm+l})$ is the pooled periodogram of the process ξ_t .

The proof of the theorem is based on the additive decomposition

$$S_n(Y^{(p)}) = S_n(\tilde{Y}^{(p)}) + T_n, \quad (32)$$

where

$$\begin{aligned} T_n &= \sum_{j=0}^{M-1} b_j \tau_j \\ \tau_j &= Y_j^{(p)} - \tilde{Y}_j^{(p)} = (2\pi f(\bar{\omega}_j))^p \frac{\Gamma(m)}{\Gamma(m+p)} \left[\left(\frac{\bar{I}_j}{f(\bar{\omega}_j)} \right)^p - \left(\frac{2\pi \bar{I}_{\xi j}}{\sigma^2} \right)^p \right], \end{aligned} \quad (33)$$

and consists of two parts. A central limit theorem (CLT) for $S_n(\tilde{Y}^{(p)})$ is proved by the method of moments following Faÿ and Soulier (2001). The second part of the proof is based on Faÿ, Moulines and Soulier (2002, FMS hereafter) and establishes that the remainder term T_n in (33) converges in probability to zero.

Under assumptions A3, A4, $E(|\xi_1|^s) < \infty, s > \max(4, 4p + 4m)$ (implied by A2) and $m > \max(1, -2p)$ (implied by A5), as well as the definition of the coefficients b_j , the conditions A1 (non-smooth case), A2, A5 and A7 of Theorem 2 in Faÿ and Soulier (2001) are satisfied and hence the CLT for $S_n(\tilde{Y}^{(p)})$ holds in the following terms:

$$S_n(\tilde{Y}^{(p)}) \rightarrow_d N \left(0, \left(\frac{\Gamma(m)\Gamma(m+2p)}{\Gamma^2(m+2p)} - 1 \right) \frac{2}{2\pi} \int_{-\pi}^{\pi} [2\pi f(\omega)]^{2p} \cos^2(\omega k) d\omega + \frac{p^2 \kappa_4}{m} \gamma_{pk}^2 \right). \quad (34)$$

In practice, the essence of the proof of the CLT consists in showing that $E[S_n(\tilde{Y}^{(p)})]^q = 0$ for q odd and that $E[S_n(\tilde{Y}^{(p)})]^q = \{\text{Var}[S_n(\tilde{Y}^{(p)})]\}^{q/2} (q-1)!!$ for q even and where $(q-1)!! = \prod_{j=1}^{\frac{q}{2}} (2j-1)$. The evaluation of the moments is done by an Edgeworth expansion for the unknown density of a non linear functional of ξ_t , as in Faÿ and Soulier (2001), based on Bhattacharya and Rao (1976). A similar approach has been followed by Chen and Hannan (1980) and Velasco (2000). The Edgeworth expansion and the moments derivation for the proof of the CLT for $S_n(\tilde{Y}^{(p)})$ can be found in the supplementary file available as Supplement.pdf

The second part of the proof consists in showing that the remainder term T_n in the additive decomposition (32) is asymptotically negligible and relies on the Bartlett decomposition of the periodogram of a linear process.

For a given frequency ω let $J(e^{-i\omega}) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n x_t \exp(-i\omega t)$ and $J_\xi(e^{-i\omega}) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n \xi_t \exp(-i\omega t)$ denote the Fourier transforms of x_t and ξ_t , respectively. The Bartlett decomposition of the periodogram $I(\omega) = |J(e^{-i\omega})|^2$ is

$$I(\omega) = \frac{2\pi f(\omega)}{\sigma^2} I_\xi(\omega) + R(\omega),$$

where $I_\xi(\omega) = |J_\xi(e^{-i\omega})|^2$, $R(\omega) = |r(e^{-i\omega})|^2 + \psi(e^{-i\omega})J_\xi(e^{-i\omega})r(e^{i\omega}) + \psi(e^{-i\omega})J_\xi(e^{i\omega})r(e^{-i\omega})$ and $r(e^{i\omega})$ is the remainder term in the Bartlett decomposition of $J(\omega)$; see Walker (1965).

When the periodogram is pooled, $\bar{I}_j = \sum_{l=1}^m I(\omega_{jm+l})$, the Bartlett decomposition becomes

$$\bar{I}_j = \frac{2\pi f(\bar{\omega}_j)}{\sigma^2} \bar{I}_{\xi j} + \bar{R}_j, \quad (35)$$

where

$$\bar{R}_j = \sigma^{-2} \sum_{l=1}^m [f(\omega_{jm+l}) - f(\bar{\omega}_j)] 2\pi I_\xi(\omega_{jm+l}) + \sum_{l=1}^m R(\omega_{jm+l}).$$

Dividing both sides of equation (35) by $f(\bar{\omega}_j)$,

$$\frac{\bar{I}_j}{f(\bar{\omega}_j)} = \frac{2\pi \bar{I}_{\xi j}}{\sigma^2} + \frac{\bar{R}_j}{f(\bar{\omega}_j)} = \frac{2\pi \bar{I}_{\xi j}}{\sigma^2} (1 + \mathcal{U}_j),$$

where

$$\mathcal{U}_j = \frac{\bar{R}_j \sigma^2}{2\pi \bar{I}_{\xi j} f(\bar{\omega}_j)},$$

we are able to express the term τ_j in (33) as

$$\tau_j = (2\pi f(\bar{\omega}_j))^p \frac{\Gamma(m)}{\Gamma(m+p)} \left(\frac{2\pi \bar{I}_{\xi j}}{\sigma^2} \right)^p [(1 + \mathcal{U}_j)^p - 1].$$

By a second order Taylor expansion of the function $[(1+x)^p - 1]$ at $x=0$,

$$\tau_j = (2\pi f(\bar{\omega}_j))^p \frac{\Gamma(m)}{\Gamma(m+p)} \left\{ p \left(\frac{2\pi \bar{I}_{\xi j}}{\sigma^2} \right)^{p-1} \frac{\bar{R}_j}{f(\bar{\omega}_j)} + \frac{p(p-1)}{2} \left[\left(\frac{2\pi \bar{I}_{\xi j}}{\sigma^2} \right) (1 + \mathcal{U}_j^*) \right]^{p-2} \left(\frac{\bar{R}_j}{f(\bar{\omega}_j)} \right)^2 \right\}$$

where \mathcal{U}_j^* is a suitable value in $(0, \mathcal{U}_j)$. As in FMS, lemma 2, it is sufficient to prove that the quantity

$$\sum_j b_j \left[\left(\frac{2\pi \bar{I}_{\xi j}}{\sigma^2} \right) (1 + \mathcal{U}_j^*) \right]^{p-2} \left(\frac{\bar{R}_j}{f(\bar{\omega}_j)} \right)^2,$$

depending on the second term of the Taylor expansion, is $o_p(1)$ in the range of values $|\mathcal{U}_j| \leq 1/2$, since $\max_j |\mathcal{U}_j| \rightarrow_p 0$, which implies that $P(\max_j |\mathcal{U}_j| > 1/2) \rightarrow 0$ as $n \rightarrow \infty$, so that the probability that $\frac{1}{2} \left(\frac{2\pi \bar{I}_{\xi j}}{\sigma^2} \right) \leq \left(\frac{2\pi \bar{I}_{\xi j}}{\sigma^2} \right) (1 + \mathcal{U}_j^*) \leq \frac{3}{2} \left(\frac{2\pi \bar{I}_{\xi j}}{\sigma^2} \right)$ tends to one.

As the factor $(2\pi f(\bar{\omega}_j))^p \frac{\Gamma(m)}{\Gamma(m+p)}$ is bounded, to show $T_n \rightarrow_p 0$, it suffices to show that

$$\sum_j b_j \left(\frac{2\pi \bar{I}_{\xi j}}{\sigma^2} \right)^{p-k} \bar{R}_j^k \rightarrow_p 0, k = 1, 2.$$

The proof follows by applying Lemma 1 and Lemma 2 of FMS and requires the additional assumption A1 and assumption A2 in its full strength. The details are rather lengthy and are provided in the supplementary file Supplement.pdf

To prove the consistency of $\hat{\gamma}_{pk}$, we write

$$\hat{\gamma}_{pk} = \frac{1}{M} \sum_{j=0}^{M-1} \tilde{Y}^{(p)} \cos(\bar{\omega}_j k) + \frac{1}{M} \sum_{j=0}^{M-1} \tau_j \cos(\bar{\omega}_j k).$$

The second term on the right hand side converges in probability to zero. Given that $E[S_n(\tilde{Y}^{(p)})]^2 = O(1)$ (which comes from the proof of the asymptotic normality of $S_n(\tilde{Y}^{(p)})$), one has that $\frac{1}{\sqrt{2M}} S_n(\tilde{Y}^{(p)}) \rightarrow_p 0$, so that, by (31), $\frac{1}{M} \sum_{j=0}^{M-1} \tilde{Y}^{(p)} \cos(\bar{\omega}_j k) \rightarrow_p \gamma_{pk}$ and therefore we have shown that $\hat{\gamma}_{pk} \rightarrow_p \gamma_{pk}$. This concludes the proof of Theorem 1.

A.2 Proof of Theorem 2

If $\xi_t \sim \text{IID } N(0, \sigma^2)$, then $2\pi \bar{I}_{\xi j}, j = 0, 1, \dots, M-1$, are identically and independently distributed according to a Gamma random variable with shape parameter m . Assuming $m > -2p$, i.e. the mean and variance of $\bar{I}_{\xi j}^p$ exist, then a central limit theorem for linear combinations of independent random variables (see Gleser, 1965) can be applied to $S_n(\tilde{Y}^{(p)})$ that gives equation (34), with $\kappa_4 = 0$.

Further, under the assumption $\sum_{j=0}^{\infty} j^{\delta'} |\psi_j| < \infty$, $\delta' = \frac{1}{2} + \epsilon, \epsilon > 0$, then $T_n \rightarrow_p 0$. The milder condition on the memory of the process arises from the proof of $\max_j |\mathcal{U}_j| \rightarrow_p 0$, which uses Theorem 2b in Walker (1965).

A.3 Proof of Theorem 3

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 $\hat{\rho}_{pj} = \frac{\hat{\gamma}_{pj}}{\hat{\gamma}_{p0}}, j = 1, \dots, K$, with the vector

$\hat{\gamma}_{pj}$, $j = 0, 1, \dots, K$, with components as in (16). The covariance matrix of $\hat{\rho}_p$ is $\mathbf{W} = \mathbf{D}\mathbf{V}\mathbf{D}'$ with partial derivatives matrix $\mathbf{D} = \frac{1}{\gamma_{p0}} [-\boldsymbol{\rho}_p \mathbf{I}_K]$, 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 (22), one gets the generalized Bartlett formula (25).

References

- Bartlett, M. S., 1954. Problemes de l'analyse spectral des series temporelles stationnaires. Publications de l'Institut de statistique de l'Université de Paris, III-3 119–134.
- Bartlett, M. S., 1955. An Introduction to Stochastic Processes, Cambridge University Press.
- Battaglia, F., 1983. Inverse Autocovariances and a Measure of Linear Determinism For a Stationary Process. Journal of Time Series Analysis, 4, 7987.
- 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.
- Bhattacharya, R. N. and Rao, R.R., 1976. Normal Approximations and Asymptotic Expansions, John Wiley and Sons.
- 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.
- 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.
- Brockwell, P.J. and Davis, R.A., 1991. Time Series: Theory and Methods, Springer-Verlag, New York.
- Chen, W.W. and Deo, R.S., 2004a. A Generalized Portmanteau Goodness-of-Fit Test for Time Series Models. Econometric Theory, 20, 382–416.
- 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.
- Chen, Z.G. and Hannan, E.J., 1980. The Distribution of Periodogram Ordinates. Journal of Time Series Analysis, 1, 1, 73–82.

- Cleveland, W.S., 1972. The Inverse Autocorrelations of a Time Series and Their Applications. *Technometrics*, 14, 2, 277–293.
- Dahlhaus, R., 1988. Small Sample Effects in Time Series Analysis: a New Asymptotic Theory and a New Estimate. *Annals of Statistics*, 16, 808–841.
- 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.
- Delgado, M.A., Hidalgo, J. and Velasco, C., 2005. Distribution Free Goodness-of-Fit Tests for Linear Processes, *Annals of Statistics*, 33, 2568–2609.
- Deo, R.S., 2000. Spectral Tests of the Martingale Hypothesis under Conditional Heteroscedasticity. *Journal of Econometrics*, 99, 291–315.
- Doob, J.L., 1953. *Stochastic Processes*, John Wiley and Sons, New York.
- Durlauf, S.N., 1991. Spectral Based Testing of the Martingale Hypothesis. *Journal of Econometrics*, 50, 355–376.
- Efron, B., and Tibshirani R.J., 1993. *An Introduction to the Bootstrap*, Chapman&Hall.
- 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.
- Erdélyi A., Magnus W., Oberhettinger F., and Tricomi F.G., 1953. *Higher Transcendental Functions*, vol. II, Bateman Manuscript Project, McGraw and Hill.
- Faÿ, G., and Soulier, P., 2001. The Periodogram of an I.I.D. sequence. *Stochastic Processes and Their Applications*, 92, 315–343.
- Faÿ, G., Moulines, E. and Soulier, P., 2002. Nonlinear Functionals of the Periodogram. *Journal of Time Series Analysis*, 23, 523–551.
- Francq, C., and Zakoïan, J-M., 2009. Bartlett’s Formula for a General Class of Nonlinear Processes. *Journal of Time Series Analysis*, 30, 449–465.
- Fuller, W.A., 1996. *Introduction to Statistical Time Series*, John Wiley & Sons.

- Gleser L.J., 1965. On the Asymptotic Theory of Fixed-Size Sequential Confidence Bounds for Linear Regression Parameters. *The Annals of Mathematical Statistics*, 36, 2, 463–467.
- Gould, H.W., 1974. Coefficient Identities for Powers of Taylor and Dirichlet Series. *The American Mathematical Monthly*, 81, 1, 3–14.
- Gradshteyn I.S. and Ryzhik I.M., 1994. Table of Integrals, Series, and Products. Jeffrey A. and Zwillinger D. Editors, Fifth edition, Academic Press.
- Graham, R.L., Knuth, D.E. and Patashnik, O., 1994. Concrete Mathematics, Addison-Wesley Publishing Company.
- Gray, H. L., Zhang, N.-F. and Woodward, W. A., 1989. On Generalized Fractional Processes. *Journal of Time Series Analysis*, 10, 233–257.
- 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.
- Harvey, A.C., and Jäger, A., 1993. Detrending, Stylized Facts and the Business Cycle. *Journal of Applied Econometrics*, 8, 231–247.
- Hernandez, F., and Johnson, R. A., 1980. The Large-Sample Behavior of Transformations to Normality. *Journal of the American Statistical Association*, 75, 855–861.
- Hong, Y., 1996. Consistent Testing for Serial Correlation of Unknown Form. *Econometrica*, 64, 4 837–864.
- Kakizawa, Y. and Taniguchi, M., 1995. Asymptotic Efficiency of the Sample Covariances in a Gaussian Stationary Process. *Journal of Time Series Analysis*, 15, 3, 304–311 .
- Krishnamoorthy, K., Mathew, T., and Mukherjee S., 2008. Normal-Based Methods for a Gamma Distribution: Prediction and Tolerance Intervals and Stress-Strength Reliability. *Technometrics*, 50, 69–78.
- Kokoszka, P.S., and Politis, D. N., 2011. Nonlinearity of ARCH and Stochastic Volatility Models and Bartlett’s Formula. *Probability and Mathematical Statistics*, 31, 47–59.
- Ljung, G. M. and Box, G. E. P., 1978. On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65, 297–303.

- Luati, A., Proietti, T. and Reale, M., 2012. The Variance Profile. *Journal of the American Statistical Association*, 107, 498, 607–621.
- Mardia, K., Kent, J. and Bibby, J., 1979. *Multivariate Analysis*, Academic Press.
- Milhøj, A., 1981. A Test of Fit in Time Series Models. *Biometrika*, 68, 1, 167–177.
- Miller, R.G., 1974. The Jackknife—A Review. *Biometrika*, 61, 1, 1–15.
- Nourdin, I. and Peccati, G., 2012. *Normal Approximations with Malliavin Calculus: From Stein’s Method to Universality*, Cambridge University Press.
- Percival, D.B. and Walden, A.T., 1993. *Spectral Analysis for Physical Applications*, Cambridge University Press.
- Pourahmadi, M., 2001. *Foundations of Time Series Analysis and Prediction Theory*, John Wiley and Sons.
- Priestley, M.B., 1981. *Spectral Analysis and Time Series*, John Wiley and Sons.
- Quenouille, M.H., 1949. Approximate Tests of Correlation in Time Series. *Journal of the Royal Statistical Society, Series B*, 11, 68–84.
- Taniguchi, M., 1980. On Estimation of the Integrals of Certain Functions of Spectral Density. *Journal of Applied Probability*, 17, 1, 73–83.
- Tieslau, M.A., Schmidt, P. and Baillie, R.T., 1996. A Minimum Distance Estimator for Long-Memory Processes. *Journal of Econometrics*, 71, 249–264.
- Tukey, J. W., 1957. On the comparative anatomy of transformations. *Annals of Mathematical Statistics*, 28, 602–632.
- Velasco, C., 2000. Non-Gaussian Log-Periodogram regression. *Econometric Theory*, 16, 44–79.
- Walker, A. M., 1965. Some Asymptotic Results for the Periodogram of a Stationary Time Series. *Journal of the Australian Mathematical Society*, 5, 107–128.
- Wilson, E. B., and Hilferty, M. M., 1931. The Distribution of Chi-Squares. *Proceedings of the National Academy of Sciences*, 17, 684–688.

- Whittle, P., 1961. Gaussian estimation in stationary time series. *Bulletin of the International Statistical Institute*, 39, 105–129.
- 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.
- Xia, Y., Tong, H., 2011. Feature Matching in Time Series Modeling. *Statistical Science*, 26, 1, 21-46.