The Variance Profile

Alessandra Luati and Tommaso Proietti and Marco Reale

19. April 2011

The Variance Profile

Alessandra Luati
Department of Statistics
University of Bologna, Italy

Tommaso Proietti
S.E.F. e ME.Q.
University of Rome Tor Vergata, Italy

Marco Reale
Department of Mathematics and Statistics
University of Canterbury, New Zealand

Abstract

The variance profile is defined as the power mean of the spectral density function of a stationary stochastic process. It is a continuous and non-decreasing function of the power parameter, p, which returns the minimum of the spectrum ($p \to -\infty$), the interpolation error variance (harmonic mean, $p = -1$), the prediction error variance (geometric mean, $p = 0$), the unconditional variance (arithmetic mean, $p = 1$) and the maximum of the spectrum ($p \to \infty$). The variance profile provides a useful characterisation of a stochastic processes; we focus in particular on the class of fractionally integrated processes. Moreover, it enables a direct and immediate derivation of the Szegö-Kolmogorov formula and the interpolation error variance formula. The paper proposes a non-parametric estimator of the variance profile based on the power mean of the smoothed sample spectrum, and proves its consistency and its asymptotic normality. From the empirical standpoint, we propose and illustrate the use of the variance profile for estimating the long memory parameter in climatological and financial time series and for assessing structural change.

Keywords: Predictability; Interpolation; Non-parametric spectral estimation; Long memory.

1 Introduction

Essential features of a stationary stochastic process can be defined in terms of averages of the spectral density. In particular, it is well known (Hannan, 1970, p. 166; Whittle, 1983, p. 68; Tong, 1979) that the unconditional variance of the process, the prediction error variance and the interpolation, or cross-validatory, error variance are given respectively by the arithmetic, geometric and harmonic mean of the spectrum.

This recognition motivates the introduction of the variance profile as a tool for characterising a stationary stochastic process. The variance profile is defined as the power mean, or Hölder mean, of the spectral density function of the process. If p denotes the power parameter, the variance profile is a continuous and non-decreasing function of p. For $p = -1$ (harmonic mean) it provides the interpolation error variance, i.e. the variance of the estimation error arising when the process at time t is predicted from the past and future observations. For $p = 0$ (geometric mean, which is the usual Szegö-Kolmogorov formula) it provides the one-step-ahead prediction error variance; for $p = 1$ (arithmetic mean) the unconditional variance of the process is obtained. Also, when $p \to \pm\infty$, the variance profile tends to the maximum and the minimum of the spectrum, so that it provides a a measure of the *dynamic range* of a stochastic process (see Percival and Walden, 1993).

The main theoretical contributions of this paper are three. Firstly, by defining the variance profile in terms of the unconditional variance of a stochastic process characterised by a fractional power transformed Wold polynomial, we offer a direct and simple derivation of the Szegö-Kolmogorov formula and the interpolation error variance. Secondly, we propose a non-parametric estimator of the variance profile based on the power mean of the smoothed sample spectrum, generalising the Davis and Jones (1968) and Hannan and Nicholls (1977) estimators for the prediction error variance. We derive its bias and prove the consistency and the asymptotic normality of the estimator, under mild assumptions on the spectral density function. This enables interval estimation of the variance profile. For $p = -1$ the estimator provides a novel estimator of the interpolation error variance which is an addition to the autoregressive and window estimators proposed by Battaglia and Bhansali (1987). Thirdly, we illustrate that the variance profile provides a useful characterisation of fractionally integrated processes.

From the empirical standpoint, we propose and illustrate the use of the variance profile for estimating the long memory parameter in climatological and financial time series and for assessing structural change.

The content of the paper can be sketched as follows. The variance profile is defined in section 2. In section 3 the definition is used to provide an alternative proof of the the prediction and interpolation error variance formulae. Section 4 deals with the estimation of the variance profile from the time series, proposing a non-parametric estimator and obtaining its asymptotic properties. We move on to illustrate how the variance profile can be used to characterise stationary processes belonging to the autoregressive (AR) and moving average (MA) class as well as long-memory processes (section 5). A strategy for estimating the long memory parameter is considered in section 6. The results are illustrated in section 7 with respect to three case studies dealing with a popular tree rings series characterised by long memory, the choice of the Box-Cox transformation parameter for series of absolute returns and the change in the variance profile in macroeconomic time series that can be ascribed to the so-called Great Moderation.

2 Definition of the Variance Profile

Let $\{x_t\}_{t\in T}$ be a stationary zero-mean stochastic process indexed by a discrete time set T, with covariance function $\gamma_k = \int_{-\pi}^{\pi} e^{\imath\omega k} dF(\omega)$, where $F(\omega)$ is the spectral distribution function of the process. The spectral representation of the process is $x_t = \int_{-\pi}^{\pi} e^{\imath\omega t} dZ(\omega)$, where $\{Z(\omega)\}_{\omega\in[-\pi,\pi]}$ is an orthogonal increment stochastic process and $\mathrm{E}[dZ(\omega)\overline{dZ(\lambda)}] = \delta_{\omega,\lambda} dF(\omega)$, with $\delta_{\omega,\lambda} = 1$ for $\omega = \lambda$ and zero otherwise (see, e.g., Brockwell and Davis, 1991, p. 138-139). We assume that the spectral density function of the process exists, $F(\omega) = \int_{-\pi}^{\omega} f(\lambda) d\lambda$, and that the process is regular (Doob, p.564), i.e. $\int_{-\pi}^{\pi} \log f(\omega) > -\infty$. We further assume that the powers $f(\omega)^p$ exist, are integrable with respect to $d\omega$ and uniformly bounded for p in (a subset of) the real line.

The variance profile, denoted by v_p, is defined as

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

or equivalently $v_p = \{\mathrm{E}[2\pi f(\omega)]^p\}^{\frac{1}{p}}$, where the expectation is taken with respect to the random variable ω, uniformly distributed in $[-\pi, \pi]$.

For $p = 1, 0, -1$, v_p gives the arithmetic, geometric and harmonic mean of the spectral density function, respectively. In these cases v_p has a physical interpretation, since it is known (Hannan, 1970, p. 166; Whittle, 1983, p. 68; Tong, 1979) that the arithmetic, geometric and harmonic mean of the spectral density give the unconditional variance, the one step ahead prediction error variance and the interpolation error variance of the process x_t, respectively.

That the arithmetic mean of the spectral density function is the unconditional variance of the process is a straightforward consequence of the spectral representation of a stationary process and its covariance function. On the other hand, the equality between the geometric mean of $2\pi f(\omega)$ and the one step ahead prediction error variance is the remarkable formula due to Szegö (1920; see English translation, Szegö and Askey, 1982), in the case of an absolutely continuous spectrum, and Kolmogorov (1941; see English translation, 1992), in the general case. We refer the reader to Grenander and Rosenblatt (1957), Hannan (1970), Ash and Gardner (1975), Doob (1953) and Priestley (1981) for alternative derivations and detailed discussions of the Szegö-Kolmogorov formula. In section 3, we shall provide a very simple proof of the Szegö-Kolmogorov formula, based on the variance profile. The equality between the harmonic mean and the interpolation error variance was also derived by Kolmogorov (1941), and we shall provide a proof based on the variance profile as well, but we also refer the reader to Wiener (1949, p. 59), for a discussion on Kolmogorov's approach, Grenander and Rosenblatt (1957, p. 83), for a formal derivation in the frequency domain, Battaglia and Bhansali (1987), Pourahmadi (2001, section 8.5), for a time domain derivation, and Kensahara, Pourahmadi and Inoue (2009) that use a novel approach based on duals of random vectors.

It is relevant to redefine the variance profile in terms of the conditional variance of an auxiliary process. Let $x_t = \psi(B)\xi_t$ denote the Wold representation of the process, with $\psi(B) = 1 + \psi_1 B +$

$\psi_2 B^2 + \ldots, \sum_j |\psi_j| < \infty$, $\xi_t \sim \mathrm{WN}(0, \sigma^2)$, where B is the lag operator, $B^j x_t = x_{t-j}$, and define the stochastic process

$$u_{pt} = \begin{cases} \psi(B)^p \xi_t & = & \psi(B)^p \psi(B)^{-1} x_t, & \text{for } p \geq 0 \\ \psi(B^{-1})^p \xi_t & = & \psi(B^{-1})^p \psi(B)^{-1} x_t, & \text{for } p < 0, \end{cases} \tag{2}$$

with spectral density function $2\pi f_u(\omega) = [\psi(e^{\imath\omega})]^{2p}\sigma^2$, satisfying $2\pi f_u(\omega)(\sigma^2)^{p-1} = [2\pi f(\omega)]^p$. It then holds that

$$v_p = \left\{ \mathrm{Var}(u_{pt}) \frac{1}{\sigma^2} \right\}^{\frac{1}{p}} \sigma^2, \tag{3}$$

where σ^2 is the variance of the innovation process ξ_t.

Hence, the variance profile can be interpreted as the reverse transformation of the unconditional variance of a fractional power transformation of the original process multiplied by a power of the innovation variance. In the next section we shall exploit this interpretation to provide an alternative derivation of the expressions for the unconditional, prediction error and interpolation error variances of x_t, that result from setting $p = 1, 0, -1$.

3 Predictability, Interpolability and the Variance Profile

It is evident from (2) and (3) that, for $p = 1$, $u_{pt} = x_t$ and

$$v_1 = \mathrm{Var}(x_t).$$

When $p = 0$, equation (2) gives $u_{pt} = \xi_t$ and, consequently, $\mathrm{Var}(u_{pt}) = \sigma^2$. It follows that $\mathrm{Var}(u_{pt})\frac{1}{\sigma^2} = 1$ and $\lim_{p\to 0} \left\{\mathrm{Var}(u_{pt})\frac{1}{\sigma^2}\right\}^{\frac{1}{p}} \sigma^2 = \sigma^2$. Hence, we have proved that

$$\lim_{p\to 0} v_p = \sigma^2. \tag{4}$$

The left-hand-side of equation (4) is the geometric average of the spectral density, $\lim_{p\to 0} v_p = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log 2\pi f(\omega) d\omega\right\}$. The right hand side of equation (4) is the variance of the innovation process in the Wold representation of x_t, i.e. the one step ahead prediction error variance

$$\mathrm{Var}(x_t|\mathcal{F}_{t-1}) = \mathrm{E}\left[x_t - \mathrm{E}(x_t|\mathcal{F}_{t-1})\right]^2 = \sigma^2,$$

where $\mathcal{F}_t = \mathfrak{S}\{x_s; s \leq t\}$ is the sigma-algebra generated by the random variables x_s, $s \leq t$. Hence, we have proved that

$$\sigma^2 = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log 2\pi f(\omega) d\omega\right\},$$

the Szegö-Kolmogorov formula.

We now consider the case $p = -1$, which uses the concept of inverse autocovariance, defined by Cleveland (1972) in the frequency domain and then considered by Chatfield (1979) in the time domain. When $p = -1$,

$$u_{pt} = \psi(B^{-1})^{-1}\xi_t \tag{5}$$

and $2\pi f_u(\omega) = \frac{\sigma^4}{2\pi} f_i(\omega)$ where $f_i(\omega) = \frac{1}{f(\omega)}$, satisfying $\gamma_{ik} = \int_{-\pi}^{\pi} e^{\imath\omega k} f_i(\omega) d\omega$, where γ_{ik} is the inverse autocovariance function of x_t and equivalently the autocovariance function of the inverse process, u_{pt}. It follows that $\mathrm{Var}(u_{pt}) = \frac{\sigma^4}{4\pi^2}\gamma_{\imath 0}$, where γ_{i0} is the inverse variance of x_t, and, consequently, $v_{-1} = \frac{4\pi^2}{\gamma_{\imath 0}}$. We now show that $v_{-1} = \frac{\sigma^4}{\mathrm{Var}(u_{pt})}$ is the interpolation error variance of x_t,

$$\mathrm{Var}(x_t|\mathcal{F}_{\setminus t}) = \mathrm{E}\left[x_t - \mathrm{E}(x_t|\mathcal{F}_{\setminus t})\right]^2$$

where $\mathcal{F}_{\setminus t} = \mathfrak{S}\{x_s; s \neq t\}$ is the past and future information set excluding the current x_t. Let us denote by $u^*_{pt} = \frac{u_{pt}}{\sigma^2}$ the inverse process u_t divided by σ^2, so that $f_{u^*}(\omega) = f^*_i(\omega)$ where $f^*_i(\omega) = \frac{1}{4\pi^2} f_i(\omega)$. The key argument of the proof is based on the fact that the stationary process u^*_{pt}, with autocovariance function $\gamma^*_{ik} = \int_{-\pi}^{\pi} e^{\imath\omega k} f^*_i(\omega) d\omega$ and corresponding autocorrelation ρ^*_{ik} can be represented as

$$u^*_{pt} = \frac{1}{\sigma^2}\psi(B^{-1})^{-1}\psi(B)^{-1}x_t,$$

as follows by (5). In fact,

$$\Gamma^*_i(B) = \left[\sigma^2\psi(B^{-1})\psi(B)\right]^{-1}$$

is the autocovariance generating function of u_{pt} and therefore we can write

$$u^*_{pt} = \sum_{k=-\infty}^{\infty} \gamma^*_{ik} x_{t-k}$$

from which

$$\frac{u^*_{pt}}{\gamma^*_{i0}} = x_t + \sum_{k=1}^{\infty} \rho^*_{ik}(x_{t-k} + x_{t+k}). \tag{6}$$

Given that,

$$\mathrm{E}(x_t|\mathcal{F}_{\setminus t}) = -\sum_{k=1}^{\infty} \rho^*_{ik}(x_{t-k} + x_{t+k}), \tag{7}$$

see Masani (1960), Salehi (1979), and Battaglia and Bhansali (1987), it follows from (6) and (7) that

$$u^*_{pt} = \gamma^*_{i0}\left[x_t - \mathrm{E}(x_t|\mathcal{F}_{\setminus t})\right] \tag{8}$$

Turning to the original coordinate system, based on u_{pt} and γ_{i0}, and taking the variance of both sides of equation (8) gives

$$v_{-1} = \mathrm{Var}(x_t|\mathcal{F}_{\setminus t}) = \frac{4\pi^2}{\gamma_{\imath 0}}.$$

The comparison of the values of v_p for $p = -1, 0, 1$ has given rise two important measures of predictability and interpolability. Nelson (1976) proposed

$$P = 1 - \frac{\mathrm{Var}(x_t|\mathcal{F}_{t-1})}{\mathrm{Var}(x_t)} = 1 - \frac{v_0}{v_1}$$

as a measure of relative predictability. See also Granger and Newbold (1981) and Diebold and Kilian (2001). The above measure can be interpreted as a coefficient of determination, i.e. as the proportion of the variance of x_t that can be predicted from knowledge of its past realization. In the signal processing literature $1 - P$ is a measure of *spectral flatness*, taking value 1 for a white noise process. Given that the spectrum is always positive and that the geometric average is no larger than the the arithmetic average, predictability is always in the range (0,1).

As for interpolability, Battaglia and Bhansali (1987) defined the index of linear determinism:

$$A = 1 - \frac{\text{Var}(x_t|\mathcal{F}_{\backslash t})}{\text{Var}(x_t)} = 1 - \frac{v_{-1}}{v_1}.$$

$A - 1$ measures the proportion of the variance that cannot be explained from knowledge of the past and the future realisations of the process.

4 Estimation of the variance profile

The simplest nonparametric estimator of the variance profile is based on the following bias corrected power mean of the periodogram:

$$\hat{v}_p = \left\{ \frac{1}{N} \sum_{j=1}^{N} (2\pi I(\omega_j))^p \, (\Gamma(p+1))^{-1} \right\}^{\frac{1}{p}}, \tag{9}$$

where $N = [(n-1)/2]$, $[\cdot]$ denotes the integer part of the argument, and

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

is the periodogram, evaluated at the Fourier frequencies $\omega_j = \frac{2\pi j}{n} \in (0, \pi)$, $1 < j < [n/2]$. Notice that, for simplicity of exposition, we have ruled out from estimation the frequencies 0 and π. The latter can be included without substantially modifying the estimator, see the discussion in Hannan and Nicholls (1977).

The factor $(\Gamma(p+1))^{-\frac{1}{p}}$ serves as a bias correction term, that we shall discuss in details later in this section. The price to be paid by correcting for the bias is that the asymptotic distribution of (9) exists, and it is normal, only for $p > -\frac{1}{2}$, which obviously excludes the relevant case of $p = -1$, when v_p gives the interpolation error variance. The reason is that the random variables $(2\pi I(\omega_j))^p$, used to estimate $(2\pi f(\omega_j))^p$, are distributed as independent Weibull (when p is positive) or Frechét (p negative) random variables with parameters $\alpha = \frac{1}{p}, \beta = (2\pi f(\omega_j))^p, \alpha, \beta > 0$ and the first two moments of the latter are finite only for $p > -\frac{1}{2}$ (see the Appendix). This essentially follows from the properties of the periodogram, that in large samples is equal to a scaled chisquare random variable (Koopmans, 1974, Chapter 8),

$$I(\omega_j) = \begin{cases} \frac{1}{2} f(\omega_j) \chi_2^2, & 0 < \omega_j < \pi \\ f(\omega_j) \chi_1^2, & \omega_j = 0, \pi, \end{cases}$$

where χ_m^2 denotes a chisquare random variable with m degrees of freedom.

The case when $p \to 0$ corresponds to the Davis and Jones (1968) estimator for the prediction error variance

$$\hat{\sigma}^2 = \exp\left[\frac{1}{N}\sum_{j=1}^{N}\log I(\omega_j) + \gamma\right], \tag{10}$$

where the log-additive bias correction term γ is the Euler gamma, i.e. minus the expectation of a log chi-square random variable. The authors showed that $\log\hat{\sigma}^2$ is asymptotically normal,

$$\log\hat{\sigma}^2 \sim \mathrm{N}\left(\log\sigma^2, \frac{2\pi^2}{6n}\right)$$

and recommended using a lognormal distribution for $\hat{\sigma}^2$ when n is not too large.

Hannan and Nicholls (1977) proposed replacing the raw periodogram ordinates by the non-overlapping averages of m consecutive ordinates,

$$\hat{\sigma}^2(m) = m\exp\left[\frac{1}{M}\sum_{j=0}^{M-1}\log\left\{\frac{1}{m}\sum_{k=1}^{m}2\pi I(\omega_{jm+k})\right\} - \psi(m)\right], \tag{11}$$

where $M = [(n-1)/(2m)]$ and $\psi(m)$ is the digamma function. The estimator (9) is obtained in the case $m = 1$. The large sample distributions of (11) and its log transform are, respectively,

$$\hat{\sigma}^2(m) \sim \mathrm{N}\left(\sigma^2, \frac{2\sigma^4 m\psi'(m)}{n}\right), \quad \log\hat{\sigma}^2(m) \sim \mathrm{N}\left(\log\sigma^2, \frac{2m\psi'(m)}{n}\right)$$

and the estimator results in a smaller mean square estimation error; increasing m reduces the variance but inflates the bias.

This suggests the following estimator, that for $m > 1$ can be computed for any $p > -\frac{m}{2}$, thereby overcoming the drawback of the estimator (9),

$$\hat{v}_p(m) = m\left[\frac{1}{M}\sum_{j=0}^{M-1}\left(\frac{1}{m}\sum_{k=1}^{m}2\pi I(\omega_{jm+k})\right)^p \frac{\Gamma(m)}{\Gamma(m+p)}\right]^{\frac{1}{p}}. \tag{12}$$

The multiplicative bias correction term is determined based on the properties of a power transform of a gamma random variable (Johnson and Kotz, 1972; see also the Appendix) and on its scaling properties. Note that, if $p \to 0$, then

$$\lim_{p\to 0}\left(\frac{\Gamma(m+p)}{\Gamma(m)}\right)^{\frac{1}{p}} = \exp\left\{-\sum_{k=1}^{m-1}\frac{1}{k} + \gamma\right\} = \exp\{-\psi(m)\} \tag{13}$$

and the estimator (12) tends to (11) (to (10) when further $m = 1$).

The asymptotic properties of the estimator (12) along with the relations with estimators (11) and (10) are stated in the following theorem.

Theorem *Let x_t be generated by a stationary Gaussian process with absolutely continuous spectral density function $f(\omega)$, whose powers $f(\omega)^p$ are integrable and uniformly bounded. Then, for $p > -\frac{m}{2}$,*

(i) $\hat{v}_p(m)$ *is consistent for* v_p,

(ii) $\sqrt{n}\{\hat{v}_p(m) - v_p\} \rightarrow_d \mathrm{N}(0, \mathrm{V}_p)$, *where* $\mathrm{V}_p = 2m\left(\frac{v_p}{p}\right)^2\left(\frac{v_{2p}}{v_p}\right)^{2p}\left(\frac{\Gamma(m+2p)\Gamma(m)}{\Gamma^2(m+p)} - 1\right)$, *and*

(iii) $\mathrm{V}_0 = 2m\sigma^4\psi'(m)$.

The proof, provided in the Appendix, is based on the properties of power transforms of basic Gamma random variables (Johnson and Kotz, 1972) and uses a central limit theorem for linear combinations of independent and identically distributed random variables by Gleser (1965), which relates to Eicker (1963) and Kolmogorov and Gnedenko (1954) and essentially establishes a Lindeberg-Feller type condition that is easy to check.

The third statement deals with case when $p \to 0$, when the asymptotic variance of the variance profile estimator is equal to the asymptotic variance of the prediction error variance estimator (11). Indeed, the Appendix provides, as a side product, an alternative proof of the asymptotic normality of the Hannan and Nicholls (1977) estimator, which was based on the asymptotic equivalence of moments.

5 The variance profile of AR, MA and long memory processes

In this section we illustrate the characterisation of certain classes of stationary processes via the variance profile.

5.1 Variance profile for AR and MA processes

The variance profile for a linear process in the autoregressive (AR) moving average (MA) form is straightforward to obtain, as a polynomial function of the process parameters.

$$\phi(B)x_t = \theta(B)\xi_t, \quad \xi_t \sim \mathrm{WN}(0, \sigma^2)$$
$$v_p = \sigma^2\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{|\theta(e^{-\imath\omega})|^{2p}}{|\phi(e^{-\imath\omega})|^{2p}}d\omega\right\}^{\frac{1}{p}}$$

Analytical formulae are easy to obtain in the case of AR(1) and MA(1) processes. Consider, for instance, the MA(1) process

$$x_t = (1 - \theta B)\xi_t, \quad \xi_t \sim \mathrm{WN}(0, \sigma^2),$$

for which we define the associated fractional power transformed process

$$u_{pt} = (1 - \theta B)^p\xi_t = \sum_{k=0}^{r}\binom{p}{k}(-\theta B)^k\xi_t,$$

where $\binom{p}{k} = \Gamma(p+1)/(\Gamma(p-k+1)\Gamma(k+1))$, and $r = p$, if p is a positive integer, $r = \infty$, otherwise, provided that the process is invertible, in that the weaker (with respect to invertibility) condition

$|\theta| \leq 1$ is required for the Binomial expansion to exist in the case of a generic exponent. The variance profile is

$$v_p = \left\{ \sum_{k=0}^{r} \binom{p}{k}^2 \theta^{2k} \right\}^{\frac{1}{p}} \sigma^2.$$

For the stationary AR(1) process,

$$(1 - \phi B)x_t = \xi_t, \quad \xi_t \sim \text{WN}(0, \sigma^2),$$

with $|\phi| < 1$, the associated fractional power transformed process is

$$u_{pt} = (1 - \phi B)^{-p} \xi_t = \sum_{k=0}^{r} \binom{-p}{k} (\phi B)^k \xi_t,$$

where r is defined as before. The summation is convergent since we have assumed that the process is stationary, i.e. $|\phi| < 1$ (it would be convergent also if the process had a unit root). The variance profile is then given by

$$v_p = \left\{ \sum_{k=0}^{r} \binom{-p}{k}^2 \phi^{2k} \right\}^{\frac{1}{p}} \sigma^2. \tag{14}$$

For AR(1) and MA(1) processes, the variance profile does not depend on the sign of the parameter ϕ or θ and tends to an horizontal straight line when $|\phi|, |\theta| \to 0$. On the other hand, for absolute values of ϕ and θ increasing towards unity, the curves described by v_p for an AR and a MA process become different. Specifically, the plot of v_p for MA(1) processes has an inflexion point in $p = 0$, where the variance profile curve changes its concavity. This does not happen to the variance profile graph of an AR(1) process which shows the same concavity for all the values of $p \in [-1, 1]$. Indeed, the variance profile of an AR(1) process shows an inflexion point in $p = 1$ (see section 5.2). Figure 1 evidences the difference between the variance profile of an autoregressive and a moving average process.

An interesting feature is that $p\,[\ln v_p - \ln v_0]$ for an AR(1) process with parameter ϕ is the same as $p\,[\ln v_{-p} - \ln v_0]$ for an MA(1) with parameter $\theta = \phi$.

5.2 Cycle models

A popular cyclical model is the circular model proposed by Harvey (1989) and West and Harrison (1989, 1997), see also Luati and Proietti (2010), which is an ARMA(2,1) process with complex conjugates AR roots and pseudo-cyclical behavior.

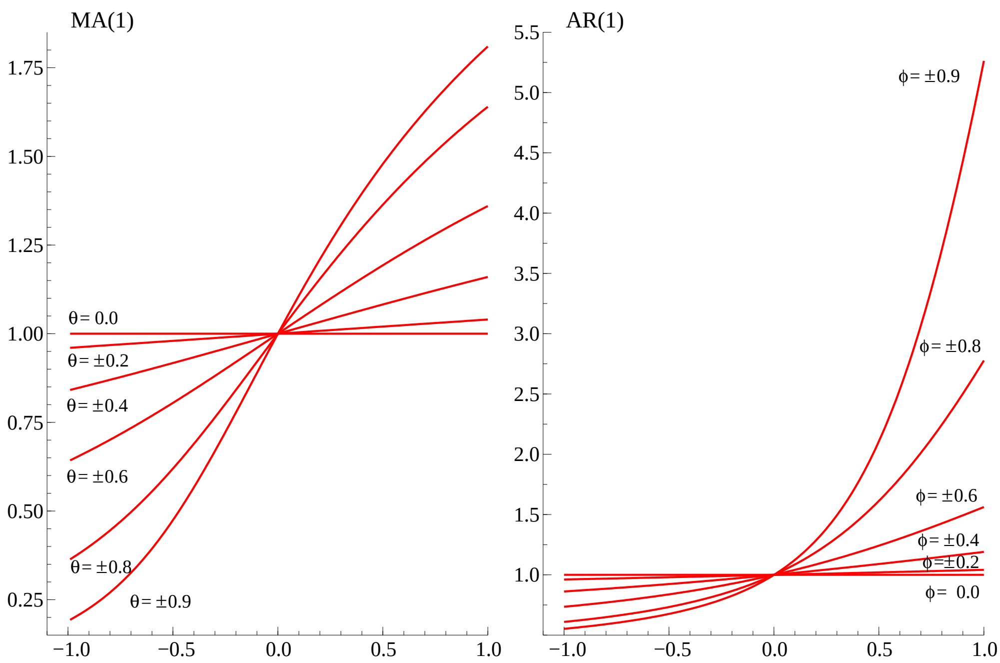
In the sequel we shall refer to the representation provided by Haywood and Tunnicliffe-Wilson (2000):

$$(1 - 2\rho \cos \varpi B + \rho^2 B^2) y_t = \frac{\sqrt{G}}{2}(1 + B)\kappa_t + \frac{\sqrt{H}}{2}(1 - B)\kappa_t^* \tag{15}$$

where $\varpi \in [0, \pi]$ is the cycle frequency, ρ is a damping factor, taking values in $[0, 1)$, κ_t and κ_t^* are two uncorrelated white noise disturbances with variance σ_κ^2, and

$$G = \sin^2\left(\frac{\varpi}{2}\right)(1 + \rho)^2 + \cos^2\left(\frac{\varpi}{2}\right)(1 - \rho)^2, H = \sin^2\left(\frac{\varpi}{2}\right)(1 - \rho)^2 + \cos^2\left(\frac{\varpi}{2}\right)(1 + \rho)^2,$$

Figure 1: Variance profiles for MA(1) and AR(1) processes with unit p.e.v.



When $\varpi = 0$, y_t is the AR(1) process $(1 - \rho B)y_t = \xi_t \sim \text{WN}(0, \sigma^2_\kappa)$; when $\varpi = \pi$, $(1 + \rho B)y_t = \xi_t$. Finally, for $\varpi = \pi/2$, $(1 + \rho^2 B^2)y_t = \sqrt{1 + \rho^2}\xi_t$.

By integrating the Fourier transform of both sides of (15) we obtain

$$\text{Var}(y_t) = \frac{\sigma^2_\kappa}{1 - \rho^2}$$

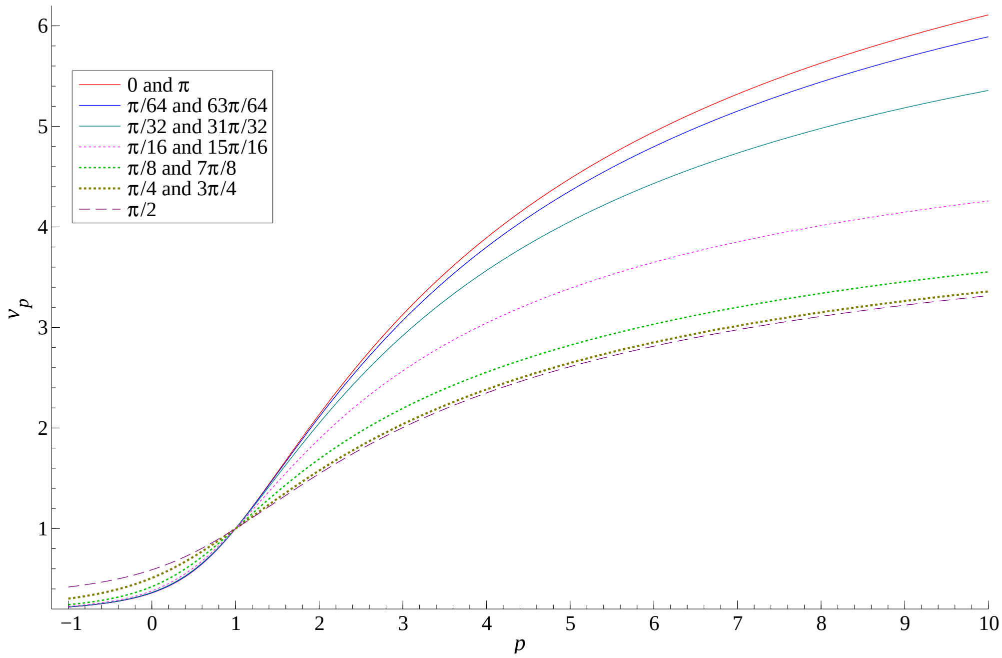
independently of the cycle frequency. Thus, the cycle models that differ only for the cycle frequency are characterised by the same variance; however, the prediction error variance and the other v_p values, $p \neq 1$ will vary with ϖ. Figure 2 illustrates this fact with reference to the case when $\rho = 0.8$ and $\sigma^2_\kappa = 1 - \rho^2$, so that $\text{Var}(y_t) = 1$. The variance profiles have an inflection point at $p = 1$ and for $p \to \infty$ converge to the maximum of the spectrum.

A seasonal component is modelled by summing trigonometric cycles defined at the fundamental frequency and at the harmonic frequencies using the same scale parameter σ^2_κ and the same ρ (e.g. $\rho \to 1$ in nonstationary seasonal models, see Hannan, Terrell and Tuckwell, 1970, and Harvey, 1989). In this case the individual cycles will be characterised by different predictability and interpolability; moreover, the maximum of the spectrum also varies.

To obtain cycle processes defined at different frequencies ϖ, but characterised by the same v_p, ρ and σ^2 have to vary according to the expression

$$\frac{d\rho}{\rho} = -\frac{1}{2}(1 - \rho^2)\frac{d\sigma^2_\kappa}{\sigma^2_\kappa}$$

Figure 2: Variance profiles for cyclical models with $\rho = 0.8$ and $\sigma^2_\kappa = (1 - \rho^2)$



For instance, the process $(1 + 0.91^2)y_t = (1 + \rho^2)^{0.5}\kappa_t$ has the same v_p as $(1 \pm 0.8)y_t = \kappa_t$.

5.3 Variance profile for long memory processes

Let us consider the fractionally integrated noise (FN) process

$$x_t = (1 - B)^{-d}\xi_t, \xi_t \sim \text{WN}(0, \sigma^2), \tag{16}$$

which is stationary for $d < 1/2$ and invertible for $d > -1$ (see Palma, 2007, Theorem 3.4 and Remark 3.1). In this range, x_t has Wold representation

$$x_t = \sum_{j=0}^{\infty} \frac{\Gamma(j + d)}{\Gamma(j + 1)\Gamma(d)}\xi_{t-j},$$

autocovariance function

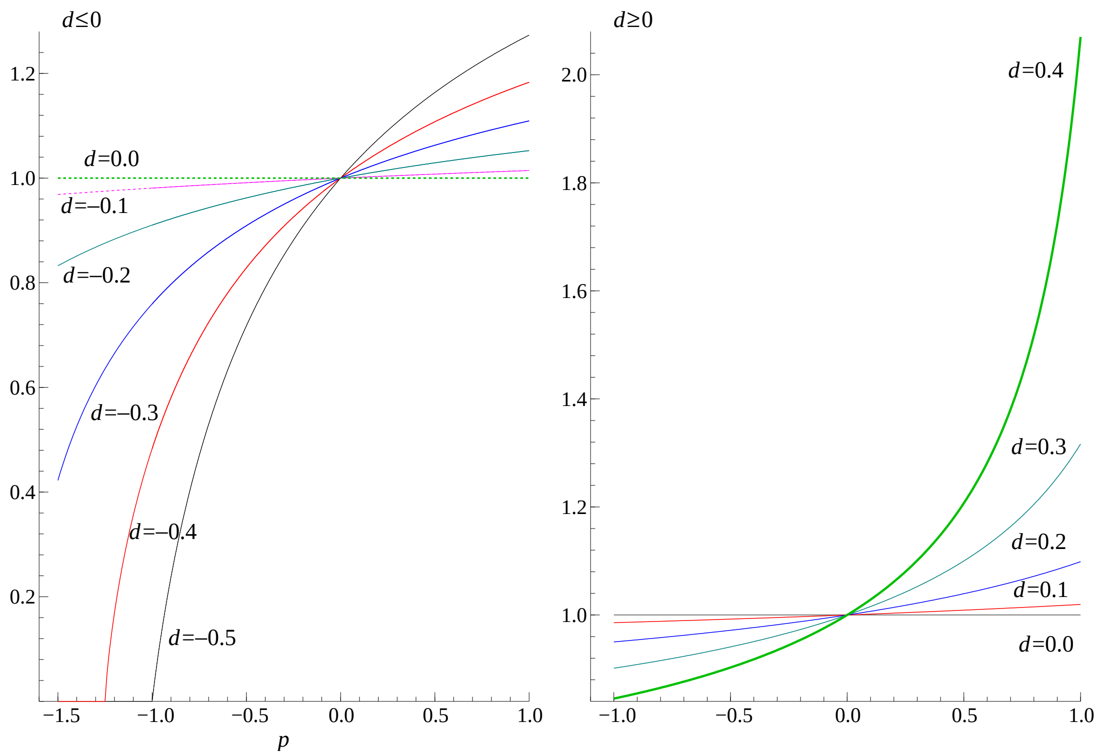
$$\gamma(h) = \sigma^2 \frac{\Gamma(1 - 2d)}{\Gamma(1 - d)\Gamma(d)} \frac{\Gamma(h + d)}{\Gamma(1 + h - d)}$$

and spectrum $f(\omega) = (2\pi)^{-1}\sigma^2 \left[2\sin(\lambda/2)\right]^{-2d}$.

The variance profile is

$$v_p = \begin{cases} \left[\frac{\Gamma(1-2pd)}{\Gamma^2(1-pd)}\right]^{1/p} \sigma^2, & dp < 0.5 \\ \infty & dp \geq 0.5, \quad d, p > 0, \\ 0 & dp \geq 0.5, \quad d, p < 0, \end{cases} \tag{17}$$

Figure 3: Variance profiles for fractional noise process with memory parameter d.



When $d \leq -0.5$ and $p = -1$ we obtain the result discussed in Walden (1994), which specialises Grenander and Rosenblatt (1957, p. 84), i.e. the interpolation error variance of a non-invertible FN process is zero. For instance, let $d = -1$ in (16), so that $x_t = \xi_t - \xi_{t-1}$. It follows immediately that $x_t = \sum_{j=1}^{\infty}(x_{t+j} - x_{t-j})$, so that x_t can be perfectly interpolated from the infinite past and future. In this case, analogous to the case of a deterministic process which occur when $\int \log 2\pi f(\omega) = -\infty$, the integral of $[2\pi f(\omega)]^{-1}$ does not exist.

Figure 3 displays the variance profiles for a FN process with varying d values. For $d \in (-0.5, 0.5)$ and $p \in (-1, 1)$, v_p exists and it is different from zero. It ought to be noticed that for negative values of d the variance profile is negatively convex, whereas for $d > 0$ the convexity is positive.

When $d > 0$, the distinctive feature of the variance profile, as compared to a short memory process with high persistence (e.g. an AR(1) with $\phi = 0.9$) is that $v_p \to \infty$ as $p \to (2 - d)$, whereas for the latter is converges to the finite maximum of the spectral density.

6 Estimation of the long memory parameter based on the variance profile

A method of moments estimator of the long–memory parameter d of a fractionally integrated process based on the variance profile can be constructed as the minimiser of the weighted Euclidean distance

Table 1: Estimation of the long memory parameter: true value is $d = 0.4$

$n = 500$

	Minimum distance estimator				G-PH estimator			
	$m = 3$	$m = 7$	$m = 11$	$m = 15$	$R = [n/16]$	$R = [n/8]$	$R = [n/4]$	$R = n$
Bias	-0.0601	-0.0550	-0.0560	-0.0562	0.0071	0.0060	0.0057	0.0033
Std. err.	0.0481	0.0421	0.0463	0.0467	0.1373	0.0912	0.0624	0.0472
MSE	0.0059	0.0048	0.0053	0.0053	0.0189	0.0083	0.0040	0.0022

$n = 1000$

	Minimum distance estimator				G-PH estimator			
	$m = 3$	$m = 7$	$m = 11$	$m = 15$	$R = [n/16]$	$R = [n/8]$	$R = [n/4]$	n
Bias	-0.0431	-0.0408	-0.0418	-0.0417	0.0053	0.0038	0.0028	0.0020
Std. err.	0.0266	0.0270	0.0295	0.0319	0.0947	0.0671	0.0457	0.0339
MSE	0.0026	0.0024	0.0026	0.0027	0.0090	0.0045	0.0021	0.0012

between the sample and the theoretical variance profile in (17):

$$\int_a^b k(p)(\hat{v}_p(m) - v_p)^2 dp.$$

In practice, in the persistent case ($d > 0$), we evaluate both (12) and (17) on a regular grid of p values from $a > -m/2$ to $b = 1$. As v_p depends also on σ^2, we replace the latter by $\hat{\sigma^2}(m) = \hat{v}_0(m)$; this yields the theoretical profile for an FN process characterised by the same p.e.v. estimated on the time series. The weights $k(p)$ may be uniform or inversely related to the asymptotic variance V_p.

A Monte Carlo experiment using 5000 replications has been performed to assess the properties of the proposed estimator ($k(p) = 1$), when the true memory parameter is $d = 0.4$. For comparison of the bias, standard error and the mean square error we also report the same quantities for the widely applied Geweke and Porter-Hudak (1983) estimator

$$\tilde{d} = \frac{\sum_{j=1}^{R} [\ln I(\omega_j)(w_j - \bar{w})]}{\sum_{j=1}^{R}(w_j - \bar{w})^2},$$

based on the least squares regression of $\ln I(\omega_j)$ on a constant and $w_j = -2\ln(2\sin(\omega_j/2)), j = 1, \ldots, R$, where $\bar{w} = R^{-1}\sum_{j=1}^{R} w_j$.

The simulation evidence, presented in table 1, shows that for $n = 500$ and $n = 1000$ our proposed estimator performs as efficiently as the GPH using $n/8$ periodogram ordinates. The value of m that optimises the performance, among the four considered, is $m = 7$.

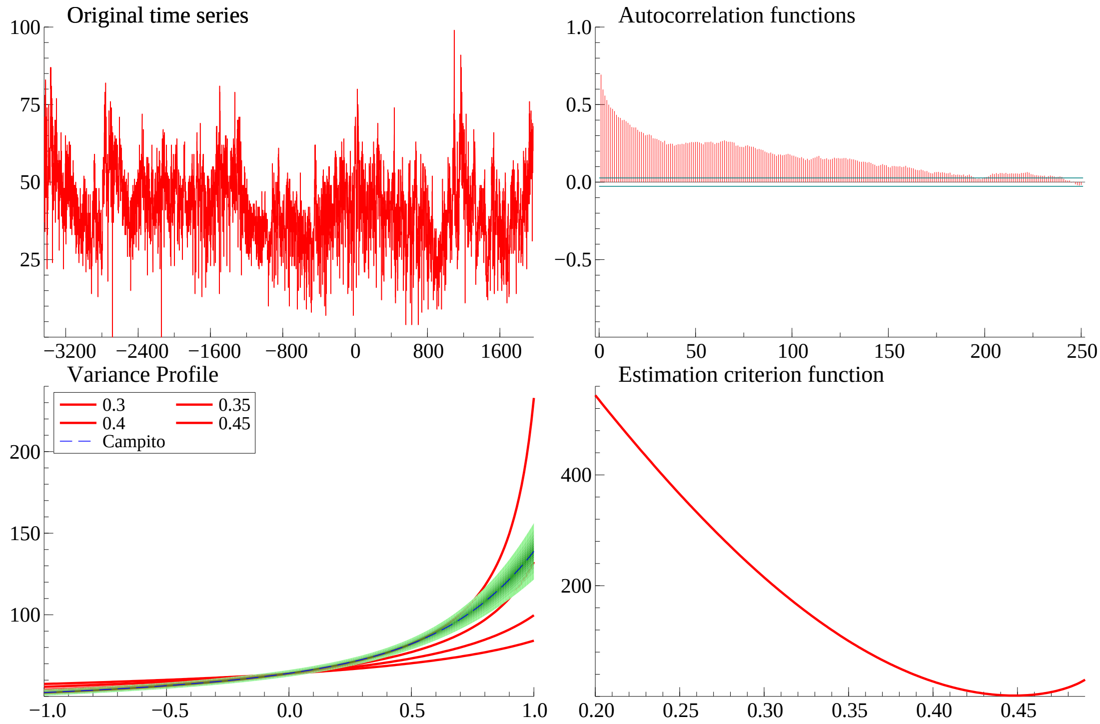


Figure 4: Mount Campito tree rings data: series (top left panel); sample autocorrelation function (top right), estimate of the variance profile using $m = 23$ (bottom left) and estimation criterion $(\hat{v}_p(m) - v_p)'(\hat{v}_p(m) - v_p)$, where the variance profile is evaluated in an equally spaced grid of values in the range (-6, 1).

7 Empirical Illustrations

7.1 Mount Campito tree rings data

The Mount Campito data is a popular time series consisting of 5405 annual values of bristlecone pine tree ring widths, spanning the period from the year 3426 BC to 1969 AD. The series is plotted in the upper left panel of figure 4; the sample autocorrelations are persistently positive and decay very slowly (see upper right panel).

The estimated variance profile is that of a long memory process with high d. It is displayed in the bottom left panel along with the 95% interval estimates, computed as $\hat{v}_p(m) \pm 1.96\sqrt{\hat{\mathrm{V}}_p/n}$ using $m = 23$. The long memory parameter is estimated equal to 0.448, a value in accordance with the literature see e.g. Baillie (1996, page 45). The estimation criterion function is plotted in the last panel.

7.2 Power transformation of absolute returns

Let r_t denote an asset return. Ding, Granger and Engle (1993) addressed the issue of determining the value of the Box-Cox (1964) power transformation parameter, λ, for which the autocorrelation property of the transformed series

$$x_t(\lambda) = |r_t|^\lambda$$

is strongest. Focusing on the Standard & Poor stock market daily closing price index over the period 3/1/1928-30/8/1991, they argued that the long memory property is strongest when $\lambda = 1$.

The analysis of two time series of returns according to the variance profile provides a broad confirmation of these findings. We focus on the daily returns computed on the Nasdaq and Standard & Poor stock market daily closing price index, available for the sample period 3/1/1989 - 7/3/2011 ($n = 10110$). As we may record zero returns we adopt the shifted-mean power transformation (see Atkinson, 1985)

$$x_t(\lambda) = (|r_t| + c)^\lambda,$$

where $c = 0.001$ (the choice of c turns out to be unimportant).

An issue arises as to whether normalised Box-Cox transform or the standardised one should be considered. The former is obtained by dividing $x_t(\lambda)$ by $\sqrt[n]{J}$, where $J = \prod_t \left|\frac{\partial x_t(\lambda)}{\partial x_t}\right|$ is the Jacobian of the transformation, (Atkinson, 1985). We prefer the second solution, as we would like to determine the transformation for which the series has the smallest normalised variance profile. In other words, we will constraint $v_1 = 1$ for all the λ values.

Setting $m = 17$ we estimate the variance profile for the standardised transformed series for values of λ in the interval $(-0.5, 2.3)$. The results for the SP500 series are presented in figure 6.

The variance profile of the standardised $x_t(\lambda)$ can be used to determine the value of the transformation parameter for which the long memory property is strongest. Figure 5 plots the value of the value of d, estimated according to section 4, against λ. It turns out that for both series the maximum d is achieved for λ around 1.25. However, the variance profile does not differ significantly for that associated to $\lambda = 1$, which does not contradict Ding, Granger and Engle (1993).

This fact is illustrated by figure 6, which refers to the SP500 series, displayed in the top right panel. The plot also shows that the normalised variance profile is a minimum for λ in the vicinity of 1.25. The last display shows the interval estimates of v_p for $\lambda = 1, 1.25$ and 2. It can be seen that the variance profile for the absolute returns does not differ from that for $\lambda = 1.25$, whereas the squared returns ($\lambda = 2$) differ significantly. The implication is that the squared returns are less predictable and interpolable than the absolute returns. Another conclusion is that the volatility of Nasdaq returns is more predictable than SP500's.

7.3 The Great moderation

The term Great Moderation (GM) refers to a substantive reduction in the volatility of macroeconomic fluctuations that took place around the mid 1980's up to the the inception of the last recession (around 2008). See, among others, McConnell and Perez-Quiros (2000) and Stock and Watson (2002). The causes of this well documented phenomenon have been the matter of a an interesting debate, with two alternative explanations being considered: a reduction in the size of economic shocks (which could be measured by the one-step ahead prediction error variance), and the change in the transmission mechanism by which shocks are propagated (which is measured by the change in the coefficients of the Wold representation).

Figure 5: Nasdaq and SP500 daily absolute returns: estimates of the long memory parameter d based on the variance profile as a function of the transformation parameter λ.

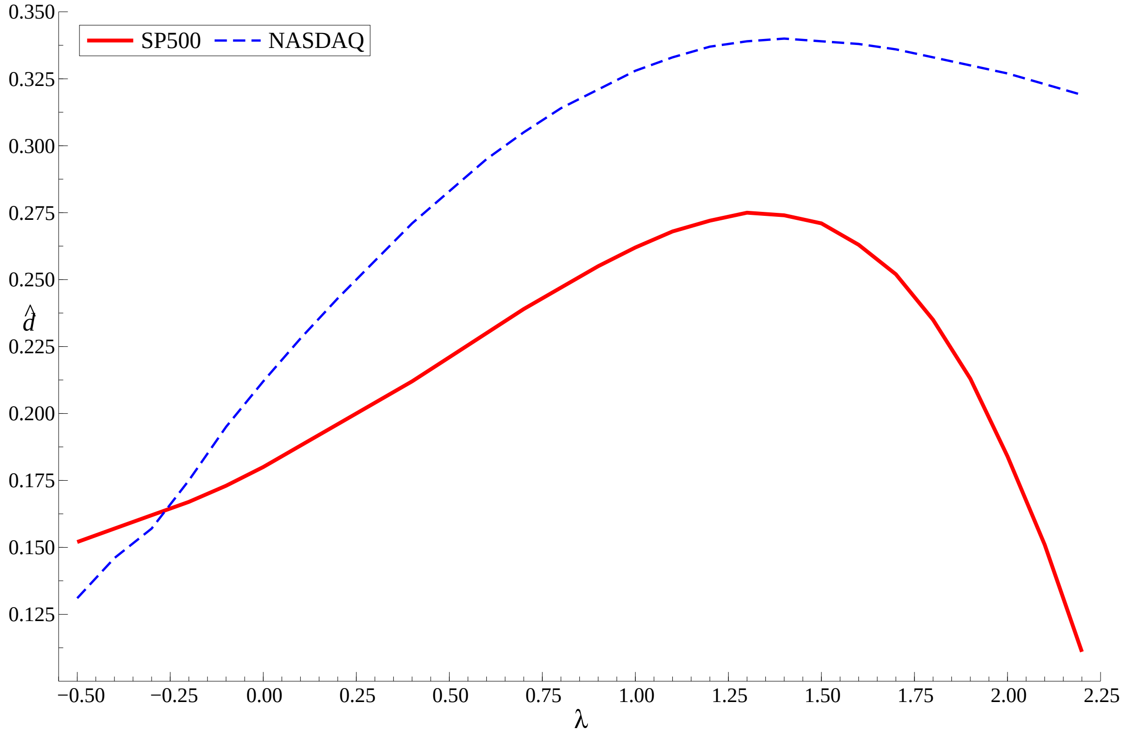
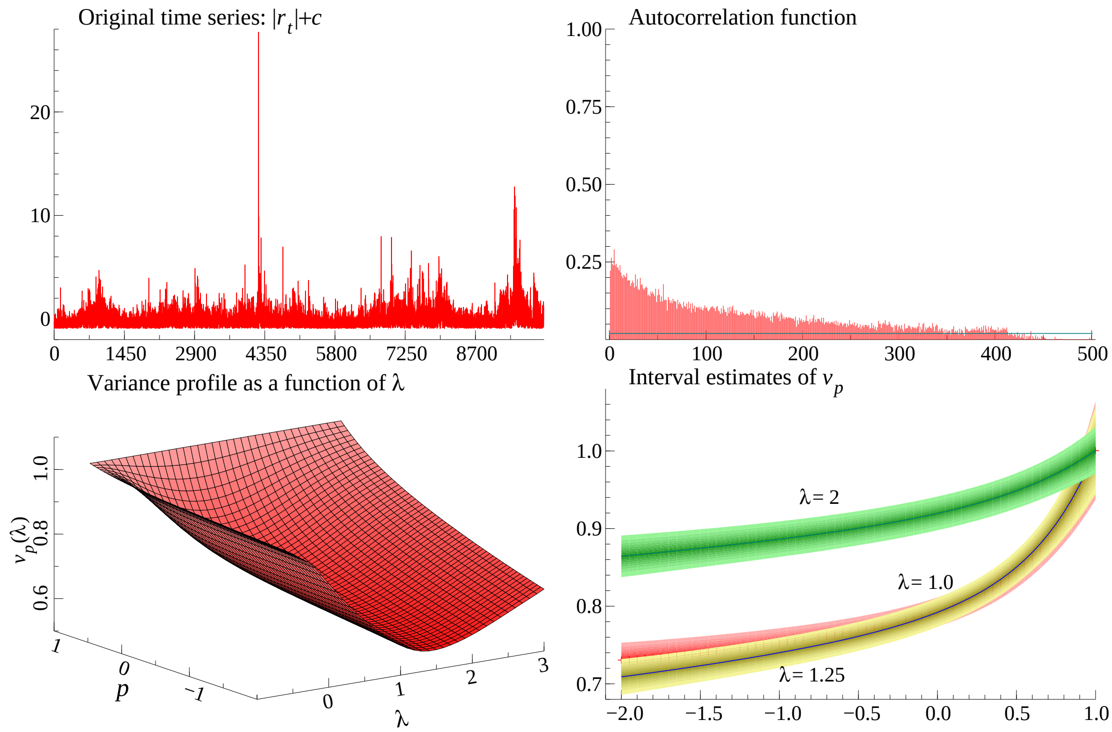


Figure 6: Standard and Poor 500 daily absolute returns (standardised): variance profile as a function of the Box-Cox transformation parameter: series (top left panel); sample autocorrelation function (top right), estimate of the variance profile using $m = 17$ (bottom left) and comparison of the interval estimates for $\lambda = 1, 1.25$ and $\lambda = 2$.



The variance profile can provide further insight on this issue. We focus on the U.S. monthly index of industrial production, made available by the Federal Reserve Board, both in seasonally adjusted and unadjusted form. We set off analysing the series of yearly growth rates for the period 1949.1-2008.6, which is split into two subseries, the first covering the period predating the GM (1949.1-1983.12) and the second covering the GM period (1984.1-2008.6). The series are plotted in figure 7; the volatility reduction is indeed visible and the patterns of the autocorrelations are also different - the behaviour is less cyclical in the GM period.

The estimated variance profile (using $m = 7$) for the two subperiods reveals that both the variance and the prediction error variance are significantly reduced in the GM period. For $p \to -1$ $\hat{v}_p$ gets very close to zero for both subseries. This is a likely consequence of the fact that seasonality in the original series is very stable, so that the yearly growth rates are likely to be non-invertible at the seasonal frequencies as a result of the application of the filter $(1 - B^{12})$.

When we come to the monthly growth rates (computed on the seasonally adjusted series), see the bottom panel of figure 7, we also find a significant change in the variance profile, which flattens downs to an almost horizontal pattern.

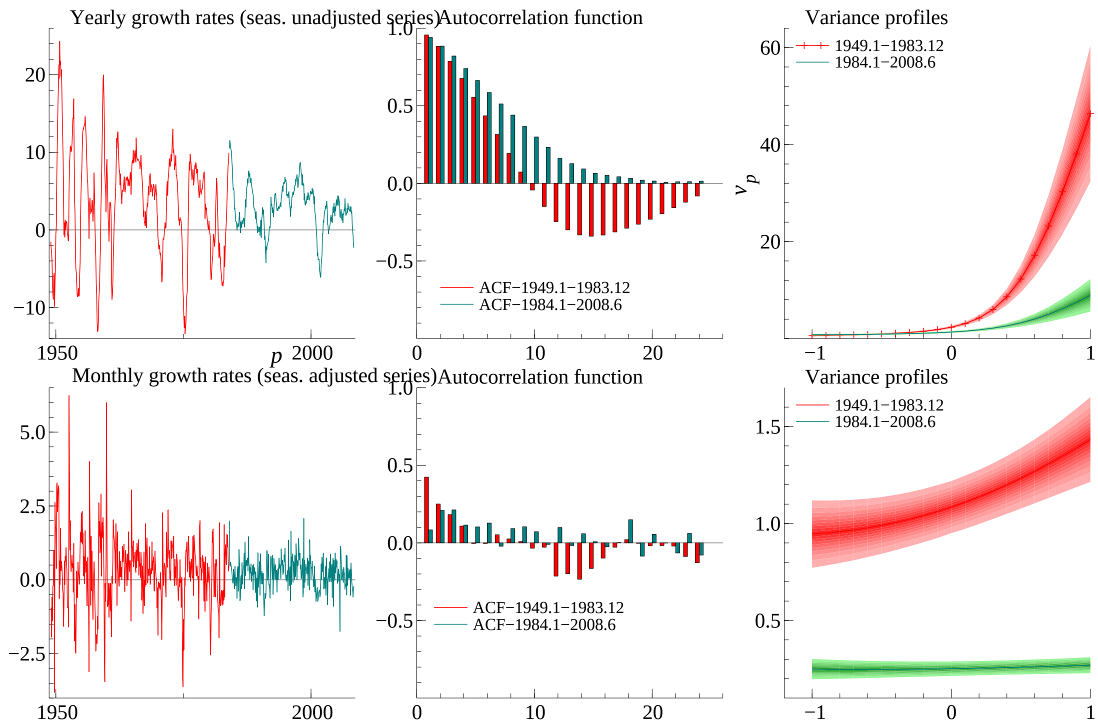
It should be noticed, however, that even though the GM is associated to a significant drop in the prediction error variance (v_0), the relative predictability, $1 - v_0/v_1$, decreased, as well as the interpolability, as measured by the index of linear determinism $1 - v_{-1}/v_1$.

8 Conclusions

The paper has introduced the variance profile and has proposed an estimator based on the smoothed periodogram, which generalises the Hannan and Nicholls (1977) estimator of the prediction error variance. The variance profile estimator was shown to be asymptotically normal and consistent.

We leave to future research the estimation of the variance profile using tapered and multi-tapered periodograms (see Pukkila and Nyquist, 1985; Walden 1995, 2000; and, more recently, Kohli and Pourahmadi, 2011) and the comparison with alternative parametric estimators based on autoregressive model fitting, as in Cleveland (1972) and Bhansali (1993). A further alternative is the estimation of the variance profile building on the Bayesian estimation of the spectral density of a time series (Choudhuri, Ghosal and Roy, 2004).

Figure 7: U.S. Index of industrial production, yearly (seasonally unadjsted) and monthly growth rates (seas. adj.) for the two subperiods 1949.1-1983.12 (pre) and 1984.1-2008.6 (Great Moderation). Comparison of autocorrelation function, variance profile and relative index $1 - v_p/v_1$.



Appendix

We provide the proof of the consistency and asymptotic normality of the estimator

$$\hat{v}_p(m) = m \left[\frac{1}{M} \sum_{j=0}^{M-1} \left(\frac{1}{m} \sum_{k=1}^{m} 2\pi I(\omega_{jm+k}) \right)^p \frac{\Gamma(m)}{\Gamma(m+p)} \right]^{\frac{1}{p}}.$$

We start from the case when $p \neq 0$; the case when $p \to 0$ will be considered afterwards.

For m odd, the quantity $\frac{1}{m}\sum_{k=1}^{m} 2\pi I(\omega_{jm+k})$ can be interpreted as a Daniell type estimator for $2\pi f(\omega_{jm+\frac{m+1}{2}})$. Hence, assuming M and m large enough for asymptotics and $\frac{m}{M}$ small enough for $f(\omega)$ to be constant over frequency intervals of length $\frac{2\pi m}{M}$, for fixed m, and for $1 \leq k \leq m$,

$$\sum_{k=1}^{m} \frac{I(\omega_{jm+k})}{\frac{1}{2} f(\omega_{jm+\frac{m+1}{2}})} \sim \chi^2_{2m}$$

(see Koopmans, 1974, pp. 269-270) and therefore

$$\sum_{k=1}^{m} 2\pi I(\omega_{jm+k}) = 2\pi f(\omega_{jm+\frac{m+1}{2}}) X_j$$

where the X_j are independent and identically distributed random variables $X_j \sim \frac{1}{2}\chi^2_{2m}$ or equivalently, $X_j \sim G(m,1)$, a basic Gamma random variable with shape parameter equal to m. Thus,

$$\left(\sum_{k=1}^{m} 2\pi I(\omega_{jm+k}) \right)^p = \left(2\pi f(\omega_{jm+\frac{m+1}{2}}) \right)^p X_j^p. \tag{18}$$

By direct integration,

$$\mathrm{E}\left(X_j^p \right) = \frac{\Gamma(m+p)}{\Gamma(m)} \tag{19}$$

and by the usual formula for the variance of a random variable one gets,

$$\mathrm{Var}(X_j^p) = \frac{\Gamma(m+2p)}{\Gamma(m)} - \frac{\Gamma^2(m+p)}{\Gamma^2(m)}, \tag{20}$$

which exist for $p > -\frac{m}{2}$. Hence, the random variable Z_j defined as

$$Z_j = \frac{X_j^p - \frac{\Gamma(m+p)}{\Gamma(m)}}{\sqrt{\frac{\Gamma(m+2p)}{\Gamma(m)} - \frac{\Gamma^2(m+p)}{\Gamma^2(m)}}}, \tag{21}$$

has zero mean and unit variance.

Under the assumption of a uniformly bounded power of the spectral density function and by approximating the integral with its Riemannian sum,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} (2\pi f(\omega))^{2p} \, d\omega = \lim_{M\to\infty} \frac{1}{M} \sum_{j=0}^{M-1} \left(2\pi f(\omega_{jm+\frac{m+1}{2}}) \right)^{2p} \tag{22}$$

the quantity

$$Q_M = \frac{1}{M}\sum_{j=0}^{M-1}\left(2\pi f(\omega_{jm+\frac{m+1}{2}})\right)^{2p}$$

exists and has a limit, $\lim_{M\to\infty} Q_M = v_{2p}^{2p}$.

Let now

$$b_j = \frac{\left(2\pi f(\omega_{jm+\frac{m+1}{2}})\right)^{p}}{\sqrt{MQ_M}}, \tag{23}$$

which satisfies

$$\sum_{j=0}^{M-1} b_j^2 = 1. \tag{24}$$

Moreover, since the p-th power of the spectral density function is uniformly bounded and since Q_M converges to a positive term, we have that

$$\max_{0\le j\le M-1} |b_j| \to 0.$$

Hence the assumptions for the central limit theorem for linear combinations of sequences of random variables (Gleser, 1965, Theorem 3.1, which relates to Eicker, 1963, and Gnedenko and Kolmogorov, 1954) are satisfied and

$$\sum_{j=0}^{M-1} b_j Z_j \to_d \mathrm{N}(0,1).$$

It follows by (21) and (24) that

$$\sum_{j=0}^{M-1} b_j X_j^p \to_d \mathrm{N}\left(\sum_{j=0}^{M-1} b_j \frac{\Gamma(m+p)}{\Gamma(m)}, \left(\frac{\Gamma(m+2p)}{\Gamma(m)} - \frac{\Gamma^2(m+p)}{\Gamma^2(m)}\right)\right)$$

and, as a function of our estimator, using (23),

$$\{\hat{v}_p(m)\}^p = \frac{1}{M}\sqrt{MQ_M}\sum_{j=0}^{M-1} b_j X_j^p \frac{\Gamma(m)}{\Gamma(m+p)} \to_d \mathrm{N}\left(\frac{1}{M}\sum_{j=0}^{M-1}\left(2\pi f(\omega_{jm+\frac{m+1}{2}})\right)^{p}, \Omega_M\right) \tag{25}$$

where

$$\Omega_M = \frac{1}{M} Q_M \left(\frac{\Gamma(m+2p)}{\Gamma(m)} - \frac{\Gamma^2(m+p)}{\Gamma^2(m)}\right)\left(\frac{\Gamma(m)}{\Gamma(m+p)}\right)^2.$$

By taking the limits

$$\sqrt{M}(\{\hat{v}_p(m)\}^p - v_p^p) \to_d \mathrm{N}\left(0, v_{2p}^{2p}\left(\frac{\Gamma(m+2p)\Gamma(m)}{\Gamma^2(m+p)} - 1\right)\right) \tag{26}$$

and applying the delta method we finally get the asymptotic distribution

$$\sqrt{n}(\hat{v}_p(m) - v_p) \to_d \mathrm{N}\left(0, 2m\left(\frac{v_p}{p}\right)^2\left(\frac{v_{2p}}{v_p}\right)^{2p}\left(\frac{\Gamma(m+2p)\Gamma(m)}{\Gamma^2(m+p)} - 1\right)\right). \tag{27}$$

We now prove the consistency of $\hat{v}_p(m)$ for v_p, that is a consequence of three facts: the Chebychev weak law of large numbers, applied to the sequence of random variables $Y_j = \sqrt{MQ_M} b_j X_j \frac{\Gamma(m)}{\Gamma(m+p)}$ in $\hat{v}_p(m)^p$, see equation (25); the convergence of the Riemannian sum to the integral, see equation (22); the Slutsky theorem for the probability limit, which allows us to state that since $\hat{v}_p(m)^p$, is consistent for v_p^p then $\hat{v}_p(m)$ is a consistent estimator for v_p, given that the power function is continuous.

Let us now consider $p \to 0$. In this case, the estimator (12) equals the prediction error variance estimator (11), see equation (13); moreover, in this context, the case $p \to 0$ correspond to the case when the logarithm of X_j is taken, rather then its power, i.e. when $p \to 0$, X_j^p is to be read as $\log X_j$. Hence, $\mathrm{E}\left(\exp\{t \log X_j\}\right)$, given in equation (19), is the moment generating function of $\log X_j$ and gives $\mathrm{E}\left(\log X_j\right) = \psi(m)$ and $\mathrm{Var}\left(\log X_j\right) = \psi'(m)$. Therefore, when $p \to 0$, equation (18) becomes (some parentheses are omitted for sake of notation)

$$\log \sum_{k=1}^{m} 2\pi I(\omega_{jm+k}) = \log 2\pi f(\omega_{jm+\frac{m+1}{2}}) + \log X_j$$

with

$$\mathrm{E} \log \left(\sum_{k=1}^{m} 2\pi I(\omega_{jm+k}) \right) = \log 2\pi f(\omega_{jm+\frac{m+1}{2}}) + \psi(m)$$

and

$$\mathrm{Var} \log \left(\sum_{k=1}^{m} 2\pi I(\omega_{jm+k}) \right) = \psi'(m),$$

respectively. What follows is that in the limit case, the bias correction via a multiplication (by the inverse expected value, see equation (25)), becomes a subtraction and the subtracted quantity does not modify the asymptotic variance of the estimator of the quantity $(\mathrm{E}(X_j^p))^{-2}$. Specifically, when $p \to 0$ $\hat{v}_p(m)^p$ takes the following form

$$\log \hat{\sigma}^2(m) = \frac{1}{M} \sum_{j=0}^{M-1} \Big(\log 2\pi f(\omega_{jm+\frac{m+1}{2}}) + \log X_j - \psi(m) \Big),$$

i.e. the sample means of M random variables each one having expected value $\log 2\pi f(\omega_{jm+\frac{m+1}{2}})$ and variance $\psi'(m)$. Since the variables are uniformly integrable (as implied by assuming that $\log f(\omega)$ is uniformly bounded for all ω) the central limit theorem applies and, since $\frac{1}{M}\sum_{j=0}^{M-1} \log f(\omega_{jm+\frac{m+1}{2}})$,

$$\sqrt{M}(\log \hat{\sigma}^2(m) - \log \sigma^2) \to_d \mathrm{N}(0, \psi'(m))$$

and replacing $M = (n-1)/(2m)$ and by the delta method,

$$\sqrt{n}(\hat{\sigma}^2(m) - \sigma^2) \to_d \mathrm{N}(0, 2m\sigma^4\psi'(m)).$$

The case when $m = 1$, is a particular case of (12). However, one could note that when $m = 1$ the estimator (12) becomes

$$\hat{v}_p = \left\{ \frac{1}{N} \sum_{j=1}^{N} \left[2\pi I(\omega_j)\right]^p \left[\Gamma(p+1)\right]^{-1} \right\}^{\frac{1}{p}}$$

and the random variables involved in its asymptotic distributions can be written as monotonic transforms of χ_2^2 random variables, as $Y_j = [2\pi I(\omega_j)]^p = \left[\pi f(\omega_j)\chi_2^2\right]^p$. It follows that for $p \neq 0$, by applying the density transform method one gets

$$f_{Yj}(y) = \frac{\frac{1}{|p|}}{[2\pi f(\omega_j)]^p}\left(\frac{y}{[2\pi f(\omega_j)]^p}\right)^{\frac{1}{p}-1}\exp\left\{-\left(\frac{y}{[2\pi f(\omega_j)]^p}\right)^{\frac{1}{p}}\right\}.$$

When p is positive and finite, then $f_{Yj}(y)$ is the density of a Weibull distribution with parameters (α, β) where $\alpha = \frac{1}{p}, \beta = [2\pi f(\omega_j)]^p$; on the other hand, when p is negative, then Y_j is distributed like a Frechét random variables with the same parameters. Note that when $p \to 0$ we find the Gumbel distribution, i.e. the distribution of the logarithm of an exponential random variable, that coincides with Davis and Jones (1968) distribution of the log-periodogram. For $p > -\frac{1}{2}$, the expected value and the variance of the Y_j are given by

$$\begin{aligned}\mathrm{E}(Y_j) &= [2\pi f(\omega_j)]^p\Gamma(p+1)\\ \mathrm{Var}(Y_j) &= [2\pi f(\omega_j)]^{2p}\left[\Gamma(2p+1) - \Gamma^2(p+1)\right]\end{aligned}$$

from which follows that the random variables $Z_j = Y_j\Gamma(p+1)^{-1}$ have mean and variance given by

$$\begin{aligned}\mathrm{E}\left(Z_j\right) &= [2\pi f(\omega_j)]^p\\ \mathrm{Var}(Z_j) &= [2\pi f(\omega_j)]^{2p}\left[\Gamma(2p+1)\Gamma^{-2}(p+1) - 1\right].\end{aligned}$$

and since they are uniformly bounded the Lindeberg-Feller central limit theorem applies and we get (26) and (27) with $m = 1$.

Note that for $p > 0$ we find the result of Corollary 1 in Taniguchi (1980), which requires positivity of the exponent for existence of the inverse Laplace transform upon which his estimator is based.

Finally, it is straightforward to verify that when $p \to 0$ and $m = 1$ we find the asymptotic distribution of the Davis and Jones (1968) estimator for the prediction error variance.

References

[1] Atkinson, A. C. (1985), *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Oxford University Press, Oxford, UK.

[2] Baillie, R.T. (1996), Long memory processes and fractional integration in econometrics, *Journal of Econometrics*, 73, 5, 1–59.

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

[4] 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.

[5] 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.

[6] Chatfield, C.S. (1972), Inverse Autocorrelations, *Journal of the Royal Statistical Society, Series A*, 142, 3, 363–377.

[7] Choudhuri, N., Ghosal, S., Roy, A. (2004), Bayesian Estimation of the Spectral Density of a Time Series, *Journal of the American Statistical Association*, 99, 468, 1050–1059.

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

[9] 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.

[10] Diebold, F.X. and Kilian, L. (2001), Measuring Predictability: Theory and Macroeconomic Applications, *Journal of Applied Econometrics*, 16, 657–669.

[11] Ding, Z., Granger, C.W.J. and Engle, R.F. (1993), A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, 1, 1, 83–106.

[12] Eicker F. (1963), Central Limit Theorems for Families of Sequences of Random Variables, *The Annals of Mathematical Statistics*, 34, 2, 439–446.

[13] Geweke, J. and Porter-Hudak, S. (1983), The Estimation and Application of Long-Memory Time Series Models, *Journal of Time Series Analysis*, 4, 221–238.

[14] 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.

[15] Gnedenko, B.V. and Kolmogorov, A.N. (1954), *Limit Distributions for Sums of Random Variables*, Addison-Wesley Publishing Company.

[16] Janacek, G. (1975), Estimation of the Minimum Mean Square Error of Prediction, *Biometrika*, 62, 1, 175–180.

[17] Johnson N.L. and Kotz, S. (1972), Power Transformations of Gamma Variables, *Biometrika*, 59, 1, 226–229.

[18] 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.

[19] Hannan, E.J., Terrell, R.D., and Tuckwell, N.E. (1970), The seasonal adjustment of economic time series, *International Economic Review*, 11, 24–52.

[20] Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

[21] Haywood, J. and Tunnicliffe-Wilson, G. (2000), An Improved State Space Representation for Cyclical Time Series, *Biometrika*, 87, 3, 724–726.

[22] Kensahara, Y., Pourahmadi, M. and Inoue, A. (2009), Duals of Random Vectors and Processes with Applications to Prediction Problems with Missing Values, *Statistics and Probability Letters*, 79, 1637–1646.

[23] Kohli, P. and Pourahmadi, M. (2011), Nonparametric estimation of the innovation variance and judging the fit of ARMA models, forthcoming in Bell, W.R., Holan, S.H. and McElroy T. (eds) *Economic Time Series: Modeling and Seasonality*, Chapman and Hall/CRC.

[24] Kolmogorov, A.N. (1941), Interpolation and Extrapolation of Stationary Random Sequences (in Russian), *Izv. Akad. Nauk SSSR Sere Mat.*, 5, 3–14.

[25] Kolmogorov, A.N. (1992), Interpolation and Extrapolation of Stationary Random Sequences, in *Selected works of A.N. Kolmogorov, Volume II, Probability and Mathematical Statistics*, 272–280.

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

[27] Luati, A. and Proietti, T. (2010), Hyper-spherical and Elliptical Stochastic Cycles, *Journal of Time Series Analysis*, 31, 169–181.

[28] Masani, P. (1960), The prediction theory of multivariate stochastic processes, III Unbounded spectral densities, *Acta Mathematica*, 104, 1-2, 141–162.

[29] McConnell, M.M. and Perez-Quiros, G. (2000), Output Fluctuations in the United States: What Has Changed since the Early 1980's?, *American Economic Review*, 90, 1464–1476.

[30] Nelson, C.R. (1976), The Interpretation of R^2 in Autoregressive-Moving Average Time Series Models, *The American Statistician*, 30, 175–180.

[31] Pukkila, T. and Nyquist, H. (1985), On frequency domain estimation of the innovation variance of a stationary univariate time series. *Biometrika*, 72, 317323.

[32] Salehi, H. (1979), Algorithms for Linear Interpolator and Interpolation Error for Minimal Stationary Stochastic Processes, *The Annals of Probability*, 7, 5, 840–846.

[33] Szegö, G. (1920), Beitrage zur Theorie der Toeplitzen Formen (Ersten Mit- teilung), *Mathematische Zeitschrift*, 6, 167-202.

[34] Szegö, G., Askey, R. (1982), *Gabor Szegö: Collected Papers, Vol. 1: 1915-1927*, Birkauser.

[35] Stock, J.H., and Watson, M.W. (2002), Has the Business Cycle Changed and Why? *NBER Macroeconomics Annual*, 17, 159–230, National Bureau of Economic Research.

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

[37] Tong, H. (1979), Final Prediction Error and Final Interpolation Error: A Paradox?, *IEEE Transactions on Information Theory*, It-25, 6, 758-759.

[38] Walden, A. T. (1995), Multitaper estimation of innovation variance of a stationary time series, *IEEE Transactions on Signal Processing*, 43, 181-187.

[39] Walden, A. T. (2000), A unified view of multitaper multivariate spectral estimation. *Biometrika*, 87, 767-788.

[40] West, M.and Harrison, J. (1989), *Bayesian Forecasting and Dynamic Models*, 1st edition, New York, Springer-Verlag.

[41] West, M.and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, 2nd edition, New York, Springer-Verlag.

[42] Wiener, N. (1949), *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, The Technology Press of MIT and John Wiley and Sons.