Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

The Variance Profile

Alessandra Luati [a], Tommaso Proietti [b c] & Marco Reale [d]

[a] Department of Statistics, University of Bologna, Italy

[b] University of Sydney Business School, NSW, Australia

[c] University of Rome "Tor Vergata,", Italy

[d] Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand

Accepted author version posted online: 14 May 2012. Version of record first published: 24 Jul 2012

To cite this article: Alessandra Luati, Tommaso Proietti & Marco Reale (2012): The Variance Profile, Journal of the American Statistical Association, 107:498, 607-621

To link to this article: http://dx.doi.org/10.1080/01621459.2012.682832

The Variance Profile

Alessandra LUATI, Tommaso PROIETTI, and Marco REALE

The variance profile is defined as the power mean of the spectral density function of a stationary stochastic process. It is a continuous and nondecreasing 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 characterization of a stochastic process; 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 article proposes a nonparametric 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.

KEY WORDS: Interpolation; Long memory; Nonparametric spectral estimation; Predictability.

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; Tong 1979; Whittle 1983, p. 68) 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 characterizing 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 nondecreasing function of p. For $p = -1$ (harmonic mean), it provides the interpolation error variance, that is, 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 measure of the *dynamic range* of a stochastic process (see Percival and Walden 1993). For noninteger p, the variance profile gives, up to a transformation, the unconditional variance of a power transform of the original process.

The variance profile provides a useful characterization of a stochastic process based on a functional of the spectral density and opens the way to the derivation of analytical results and new estimators of essential features. There are three main theoretical contributions in this article. First, by defining the variance profile in terms of the unconditional variance of a stochastic process characterized by a fractional power transformed Wold polynomial, we provide a direct and simple derivation of the Szegö-Kolmogorov formula and the interpolation error variance. Second, we propose a nonparametric estimator of the variance profile based on the power mean of the smoothed sample spectrum, generalizing the Davis and Jones (1968) and Hannan and Nicholls (1977) estimators for the prediction error variance. We prove the consistency and the asymptotic normality of the estimator, under mild assumptions on the spectral density function, and discuss its sampling properties. For $p = -1$, a novel estimator of the interpolation error variance is provided, which is an addition to the autoregressive and window estimators proposed by Battaglia and Bhansali (1987). Third, we illustrate that the variance profile provides an effective characterization of fractionally integrated processes; we introduce a feature matching (Xia and Tong 2011) minimum distance estimator of the memory parameter and discuss its performance.

From the empirical standpoint, we 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 article 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 prediction and interpolation error variance formulas. We move on to illustrate how the variance profile can be used to characterize stationary processes belonging to the class of autoregressive fractionally integrated moving average (ARFIMA) class encompassing AR, MA, as well as long-memory processes (Section 4). Section 5 proposes a nonparametric estimator of the variance profile and derives its asymptotic properties. A new minimum distance estimator of the long memory parameter is introduced. The results are illustrated in Section 6 with respect to three case studies dealing with a popular tree rings series characterized by long memory, the choice of the Box-Cox transformation parameter for series of absolute returns, and the change in the variance profile of macroeconomic time series that can be ascribed to the

Alessandra Luati is Associate Professor of Statistics, Department of Statistics, University of Bologna, Italy (E-mail: *alessandra.luati@unibo.it*). Tommaso Proietti is Associate Professor of Statistics, University of Sydney Business School, NSW, Australia and University of Rome "Tor Vergata," Italy (E-mail: *t.proietti@econ.usyd.edu.au*). Marco Reale is Senior Lecturer at the Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand (E-mail: *marco.reale@canterbury.ac.nz*). The authors thank Karim Abadir, Francesco Battaglia, Rajendra Bhansali, Esfandiar Maasoumi, Domenico Piccolo, Mohsen Pourahmadi, Howell Tong, the Editor, Professor Xuming He, an Associate Editor, and two Referees, for their helpful comments and suggestions that led to several improvements in this article.

Journal of the American Statistical Association
June 2012, Vol. 107, No. 498, Theory and Methods
DOI: 10.1080/01621459.2012.682832

so-called Great Moderation. In the conclusions (Section 7), we hint at other uses of the variance profile.

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 $\gamma_k = \int_{-\pi}^{\pi} e^{\iota\omega k} \mathrm{d}F(\omega)$, where $F(\omega)$ is the spectral distribution function of the process. The spectral representation of the process is $x_t = \int_{-\pi}^{\pi} e^{\iota\omega t} \mathrm{d}Z(\omega)$, where $\{Z(\omega)\}_{\omega\in[-\pi,\pi]}$ is an orthogonal increment stochastic process and $\mathrm{E}[\mathrm{d}Z(\omega)\overline{\mathrm{d}Z(\lambda)}] = \delta_{\omega,\lambda}\mathrm{d}F(\omega)$, with $\delta_{\omega,\lambda} = 1$ for $\omega = \lambda$ and zero otherwise (see, e.g., Brockwell and Davis 1991, pp. 138–139). We assume that the spectral density function of the process exists, $F(\omega) = \int_{-\pi}^{\omega} f(\lambda)\mathrm{d}\lambda$, and that the process is regular (Doob 1953, p. 564), that is $\int_{-\pi}^{\pi} \log f(\omega)\mathrm{d}\omega > -\infty$. We further assume that the powers $f(\omega)^p$ exist, are integrable with respect to $\mathrm{d}\omega$, and uniformly bounded for p in the real line.

The variance profile, denoted by v_p, is defined as

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

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 $[-\pi, \pi]$. In the limit case, when $p \to 0$, $\lim_{p\to 0} v_p = \exp\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log 2\pi f(\omega)\mathrm{d}\omega\}$, where the assumptions of a regular process with uniformly bounded powers of the spectrum allows exchanging the order between the limit operation and the integral sign. Under the assumption of integrability of the power spectrum, $\lim_{p\to\infty} v_p = \max\{2\pi f(\omega)\}$. The result is a direct consequence of the infinity norm and, actually, a convenient way to obtain results about v_p for all p is to define $v_p(f)$ for $p > 0$, $v_0(f) = \lim_{p\to 0} v_p(f)$ and $v_p(f) = v_{-p}(\frac{1}{f})$, like in the literature by Hardy, Littlewood, and Pólya (1934), to which we address the reader for detailed and alternative proofs of these and other interesting properties of power means. For instance, it is sufficient to set $\min f(\omega) = \frac{1}{\max \frac{1}{f(\omega)}}$ to get $\lim_{p\to-\infty} v_p = \min 2\pi f(\omega)$. Thus defined, the variance profile is a continuous and increasing function of p, since $v_p < v_q$ for $p < q$, except in the case of constant $f(\omega)$ (i.e., for a white noise process, whose variance profile is constant), which can be proved for positive exponents by the Hölder inequality and then extended to the negative index case.

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; Tong 1979; Whittle 1983, p. 68) 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 due to the remarkable formula by Szegö (1920; English translation, Szegö and Askey 1982), in the case of an absolutely continuous spectrum, and by 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, to Grenander and Rosenblatt (1957, p. 83) for a formal derivation in the frequency domain, to Battaglia and Bhansali (1987) and Pourahmadi (2001, sec. 8.5) for a time domain derivation, and to Kensahara, Pourahmadi, and Inoue (2009) who used a novel approach based on duals of random vectors.

It is relevant to redefine the variance profile in terms of the unconditional variance of an auxiliary process. Let $x_t = \psi(B)\xi_t$ denote the Wold representation of the process, with $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots$, $\sum_j |\psi_j| < \infty$, $\xi_t \sim \mathrm{WN}(0, \sigma^2)$, where B is the lag operator, $B^j x_t = x_{t-j}$, and define the stochastic process

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

with spectral density function $2\pi f_u(\omega) = [\psi(e^{\iota\omega})]^{2p}\sigma^2$, satisfying $2\pi f_u(\omega)(\sigma^2)^{p-1} = [2\pi f(\omega)]^p$. Existence of $\psi(B)^{-1}$ requires that all the roots of the characteristic equation $\psi(B) = 0$ are in modulus greater than 1. For arbitrary p, the power of $\psi(B)$ in Equation (2) is still a power series, $\psi(B)^p = \sum_{j=0}^{\infty} \varphi_j B^j$, with coefficients given by the recursive relation $\varphi_j = \frac{1}{j}\sum_{k=1}^{j}[k(p+1) - j]\psi_k\varphi_{j-k}$, $j > 0$, $\varphi_0 = 1$ (see Gould 1974). In most practical applications, a finite version or approximation of $\psi(B)$ can be considered, say a q dimensional polynomial $\psi_q(B)$, with q roots $-\zeta_1^{-1}, -\zeta_2^{-1}, \ldots, -\zeta_q^{-1}$ lying outside the unit circle to ensure invertibility. Hence, $\psi_q(B)^p = (1 + \zeta_1 B)^p(1 + \zeta_2 B)^p \ldots (1 + \zeta_q B)^p$, where each factor can be expanded using the binomial theorem holding for $p \in \mathbb{R}$ and $\zeta_i \in \mathbb{C}$, $(1 + \zeta_i B)^p = \sum_{k=0}^{\infty}\binom{p}{k}(\zeta_i B)^k$, where $\binom{p}{k} = \frac{p(p-1)(p-2)\cdots(p-k+1)}{k(k-1)(k-2)\cdots 1}$ with initial conditions $\binom{p}{0} = 1, \binom{p}{1} = p$, and where absolute convergence is implied by invertibility (see Graham, Knuth, and Patashnik 1994, chap. 5). 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 Equations (2) and (3) that, for $p = 1$, $u_{1t} = x_t$ and $v_1 = \mathrm{var}(x_t)$.

When $p = 0$, Equation (2) gives $u_{0t} = \xi_t$ and, consequently, $\mathrm{var}(u_{0t}) = \sigma^2$. It follows that $\mathrm{var}(u_{0t})\frac{1}{\sigma^2} = 1$ and $\lim_{p\to 0}\{\mathrm{var}(u_{0t})\frac{1}{\sigma^2}\}^{\frac{1}{p}}\sigma^2 = \sigma^2$. 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, $\lim_{p\to 0} v_p = \exp\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log 2\pi f(\omega)\mathrm{d}\omega\}$. The right-hand side of Equation (4) is the variance of the innovation process in the Wold representation of x_t, that is, the one-step ahead prediction error variance $\mathrm{var}(x_t|\mathcal{F}_{t-1}) = \mathrm{E}[x_t - \mathrm{E}(x_t|\mathcal{F}_{t-1})]^2 = \sigma^2$, where $\mathcal{F}_t = \mathfrak{S}\{x_s; s \le t\}$ is the sigma-algebra generated by the random variables x_s, $s \le t$. Hence, we have proved that

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

the Szegö-Kolmogorov formula.

We now consider the case $p = -1$, which uses the concept of inverse autocovariance, defined by Cleveland (1992) in the frequency domain and then considered by Chatfield (1979) in the time domain. When $p = -1$, $u_{-1t} = \psi(B^{-1})^{-1}\xi_t$ and $2\pi f_u(\omega) = \frac{\sigma^4}{2\pi} f_i(\omega)$ where $f_i(\omega) = \frac{1}{f(\omega)}$, satisfying $\gamma_{ik} = \int_{-\pi}^{\pi} e^{\imath\omega k} f_i(\omega)\mathrm{d}\omega$, where γ_{ik} is the inverse autocovariance function of x_t and equivalently the autocovariance function of the inverse process, u_{-1t}. It follows that $\mathrm{var}(u_{-1t}) = \frac{\sigma^4}{4\pi^2}\gamma_{i0}$, where γ_{i0} is the inverse variance of x_t, and, consequently, $v_{-1} = \frac{4\pi^2}{\gamma_{i0}}$. We now show that $v_{-1} = \frac{\sigma^4}{\mathrm{var}(u_{-1t})}$ is the interpolation error variance of x_t, $\mathrm{var}(x_t|\mathcal{F}_{\setminus t}) = \mathrm{E}[x_t - \mathrm{E}(x_t|\mathcal{F}_{\setminus t})]^2$, where $\mathcal{F}_{\setminus t} = \mathfrak{S}\{x_s; s \ne t\}$ is the past and future information set excluding the current x_t. Let us denote by $u^*_{-1t} = \frac{u_{-1t}}{\sigma^2}$ the inverse process u_{-1t} 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^*_{-1t}, with autocovariance function $\gamma^*_{ik} = \int_{-\pi}^{\pi} e^{\imath\omega k} f_i^*(\omega)\mathrm{d}\omega$ and corresponding autocorrelation ρ^*_{ik} can be represented, by Equation (2), as $u^*_{-1t} = \frac{1}{\sigma^2}\psi(B^{-1})^{-1}\psi(B)^{-1}x_t$. In fact, $\Gamma_i^*(B) = [\sigma^2\psi(B^{-1})\psi(B)]^{-1}$ is the autocovariance generating function of u_{pt} and therefore we can write $u^*_{-1t} = \sum_{k=-\infty}^{\infty}\gamma^*_{ik}x_{t-k}$, from which $\frac{u^*_{-1t}}{\gamma^*_{i0}} = x_t + \sum_{k=1}^{\infty}\rho^*_{ik}(x_{t-k} + x_{t+k})$. Given that, $\mathrm{E}(x_t|\mathcal{F}_{\setminus t}) = -\sum_{k=1}^{\infty}\rho^*_{ik}(x_{t-k} + x_{t+k})$ (see Masani 1960; Salehi 1979; and Battaglia and Bhansali 1987), it follows that $u^*_{-1t} = \gamma^*_{i0}[x_t - \mathrm{E}(x_t|\mathcal{F}_{\setminus t})]$. Turning to the original coordinate system, based on u_{-1t} and γ_{i0}, and taking the variance one obtains

$$v_{-1} = \mathrm{var}(x_t|\mathcal{F}_{\setminus t}) = \frac{4\pi^2}{\gamma_{i0}}.$$

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, taking values in the range $(0, 1)$. The above measure can be interpreted as a coefficient of determination, that is, 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. As for interpolability, Battaglia and Bhansali (1987) defined the index of linear determinism: $A = 1 - \frac{\mathrm{var}(x_t|\mathcal{F}_{\setminus t})}{\mathrm{var}(x_t)} = 1 - \frac{v_{-1}}{v_1}$. The quantity $A - 1$ measures the proportion of the variance that cannot be explained from knowledge of the past and the future realizations of the process.

4. THE VARIANCE PROFILE OF AR, MA, AND LONG MEMORY PROCESSES

We illustrate the characterization of certain classes of stationary processes via the variance profile. We shall consider the class of ARFIMA(r, d, q) models $\phi_r(B)(1 - B)^d x_t = \theta_q(B)\xi_t$, $\xi_t \sim \mathrm{WN}(0, \sigma^2)$, where $\phi_r(B)$ and $\theta_q(B)$ are polynomials of order r and q, respectively, with roots lying outside the unit circle, and $-1 < d < 0.5$. When $d = 0$, the process is ARMA(r, q), when $r = q = 0$ and $-1 < d < 0.5$, the process is fractional noise. The variance profile is

$$\begin{aligned} v_p &= \sigma^2\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}|1 - e^{-\imath\omega}|^{-2pd}\frac{|\theta_q(e^{-\imath\omega})|^{2p}}{|\phi_r(e^{-\imath\omega})|^{2p}}\mathrm{d}\omega\right\}^{\frac{1}{p}} \\ &= \sigma^2\left(\sum_{j=0}^{\infty}\varphi_j^2\right)^{\frac{1}{p}}, \end{aligned} \tag{5}$$

where the φ_j are obtained recursively from the Wold coefficients, as discussed in Section 2.

4.1 Variance Profile for AR and MA Processes

In the case of first-order processes, v_p can be derived directly by applying the binomial theorem. Consider, for instance, the MA(1) process $x_t = (1 - \theta B)\xi_t$, $\xi_t \sim \mathrm{WN}(0, \sigma^2)$, for which we define the associated power transformed process $u_{pt} = (1 - \theta B)^p\xi_t = \sum_{k=0}^{\infty}\binom{p}{k}(-\theta B)^k\xi_t$, where $\binom{p}{k} = \frac{p(p-1)(p-2)\cdots(p-k+1)}{k(k-1)(k-2)\cdots 1}$ and the invertibility condition $|\theta| < 1$ is required for the binomial expansion to be absolutely convergent in the case of a real exponent; if p is a positive integer, the summation terminates at p. The variance profile is

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

For the stationary AR(1) process, $(1 - \phi B)x_t = \xi_t$, $\xi_t \sim \mathrm{WN}(0, \sigma^2)$, with $|\phi| < 1$, the associated power transformed process is $u_{pt} = (1 - \phi B)^{-p}\xi_t = \sum_{k=0}^{\infty}\binom{-p}{k}(-\phi B)^k\xi_t = \sum_{k=0}^{\infty}\binom{p+k-1}{k}(\phi B)^k\xi_t$, where we have applied the basic identity $\binom{p}{k} = (-1)^k\binom{-p+k-1}{k}$ (see Graham, Knuth, and Patashnik 1994, p. 164). The variance profile is then given by

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

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 a horizontal straight line when $|\phi|, |\theta| \to 0$, that is, in the white noise case. On the other hand, for absolute values of ϕ and θ increasing toward unity, the curves described by v_p for an AR

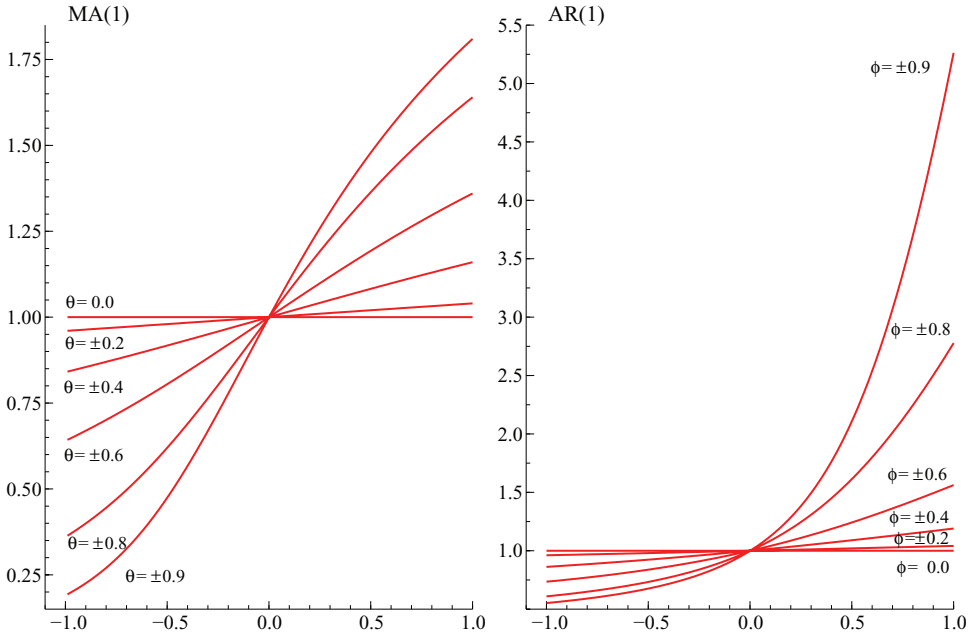

Figure 1. Variance profiles for MA(1) and AR(1) processes with unit prediction error variance. The online version of this figure is in color.

and a MA process become different. Specifically, the plot of v_p for MA(1) processes has an inflection point at $p = 0$, where the variance profile curve changes its concavity. This does not happen to the variance profile graph of an AR(1) process that shows the same concavity for all the values of $p \in [-1, 1]$. Figure 1 evidences the difference between the variance profile of an autoregressive and a moving average process.

The discrimination of an AR(1) process from an MA(1) one as a model for the disturbance term of a linear regression model has been an issue investigated since the seminal article by Nicholls, Pagan, and Terrell (1975; see King and McAleer 1987; Burke, Godfrey, and Tremayne 1990; Silvapulle and King 1991; Baltagi 2005, chap. 5 for further references). The variance profile can shed light on the issue. Defining $z_p = p[\ln v_p - \ln v_0]$, the differences $\frac{z_{-p}-z_p}{2p} = \ln v_0 - \frac{1}{2}(\ln v_p + \ln v_{-p})$, are zero for a white noise process, greater than zero for an AR(1) process, and negative for an MA(1) process.

4.2 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):

$$(1 - 2\rho\cos\varpi\, B + \rho^2 B^2)x_t = \frac{\sqrt{G}}{2}(1 + B)\kappa_t + \frac{\sqrt{H}}{2}(1 - B)\kappa_t^* \qquad (6)$$

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 $G = \sin^2(\frac{\varpi}{2})(1 + \rho)^2 + \cos^2(\frac{\varpi}{2})(1 - \rho)^2$, $H = \sin^2(\frac{\varpi}{2})(1 - \rho)^2 + \cos^2(\frac{\varpi}{2})(1 + \rho)^2$, When $\varpi = 0$, x_t is the AR(1) process $(1 - \rho B)x_t = \xi_t \sim$ WN$(0, \sigma_\kappa^2)$; when $\varpi = \pi$, $(1 + \rho B)x_t = \xi_t$. Finally, for $\varpi = \pi/2$, $(1 + \rho^2 B^2)x_t = \sqrt{1 + \rho^2}\xi_t$. By integrating the Fourier transform of both sides of Equation (6), we obtain $\text{var}(x_t) = \frac{\sigma_\kappa^2}{1-\rho^2}$ independently of the cycle frequency. Thus, the cycle models that differ only for the cycle frequency are characterized 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 $\sigma_\kappa^2 = 1 - \rho^2$, so that $\text{var}(x_t) = 1$. The variance profiles, obtained by Equation (5) with $l = 2$, $d = 0$, $q = 1$, have an inflection point at $p = 1$ and for $p \to \infty$ converge to the maximum of the spectrum.

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

To obtain cycle processes defined at different frequencies ϖ, but characterized by the same v_p, ρ and σ^2 have to vary according to the expression $\frac{\text{d}\rho}{\rho} = -\frac{1}{2}(1 - \rho^2)\frac{\text{d}\sigma_\kappa^2}{\sigma_\kappa^2}$. For instance, the process $(1 + 0.91^2 B^2)x_t = (1 + 0.91^2)^{0.5}\kappa_t$ has the same v_p as $(1 \pm 0.8)x_t = \kappa_t$.

4.3 Variance Profile for Long Memory Processes

Let us consider the fractionally integrated noise (FN) process $x_t = (1 - B)^{-d}\xi_t$, $\xi_t \sim$ WN$(0, \sigma^2)$, 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 $f(\omega) = (2\pi)^{-1}\sigma^2[2\sin(\omega/2)]^{-2d}$.

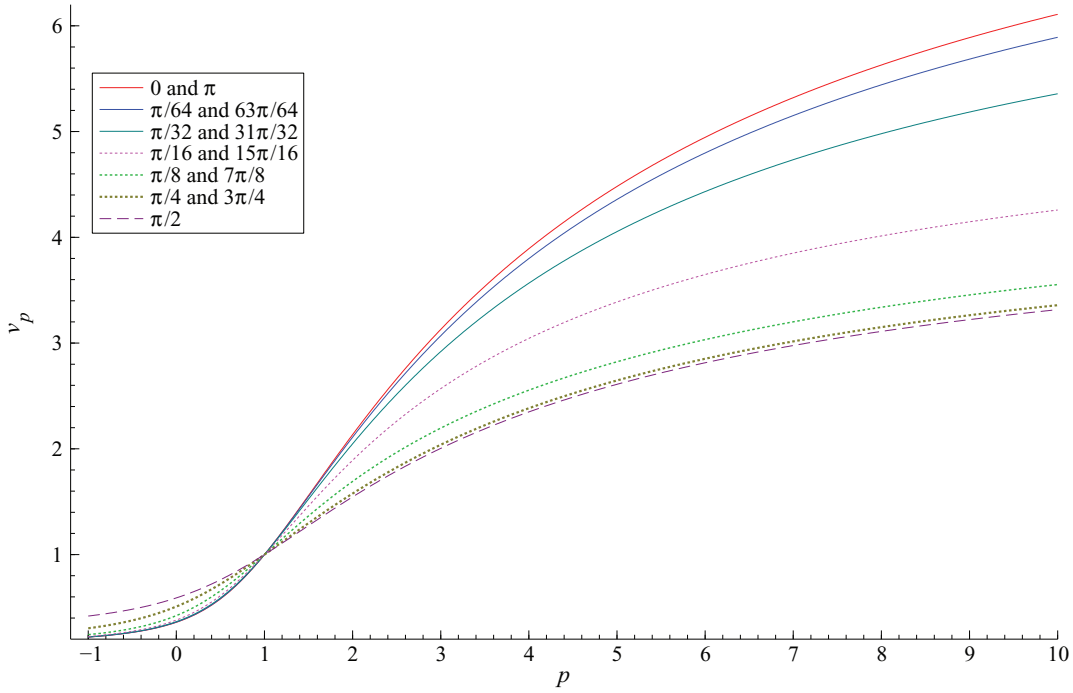

Figure 2. Variance profiles for cyclical models with $\rho = 0.8$ and $\sigma_\kappa^2 = (1 - \rho^2)$. The online version of this figure is in color.

The variance profile is

$$v_p = \begin{cases} \left[\dfrac{\Gamma(1-2pd)}{\Gamma^2(1-pd)}\right]^{1/p} \sigma^2, & dp < 0.5 \\ \infty & dp \geq 0.5, \quad d, p > 0, \\ 0 & dp \geq 0.5, \quad d, p < 0. \end{cases} \tag{7}$$

When $d \leq -0.5$ and $p = -1$, we obtain the result discussed in Walden (1994), which specializes Grenander and Rosenblatt (1957, p. 84), that is, the interpolation error variance of a noninvertible FN process is zero. For instance, if $d = -1$ then $x_t = \xi_t - \xi_{t-1}$. 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 that occurs when $\int_{-\pi}^{\pi} \log 2\pi f(\omega) \mathrm{d}\omega = -\infty$, 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

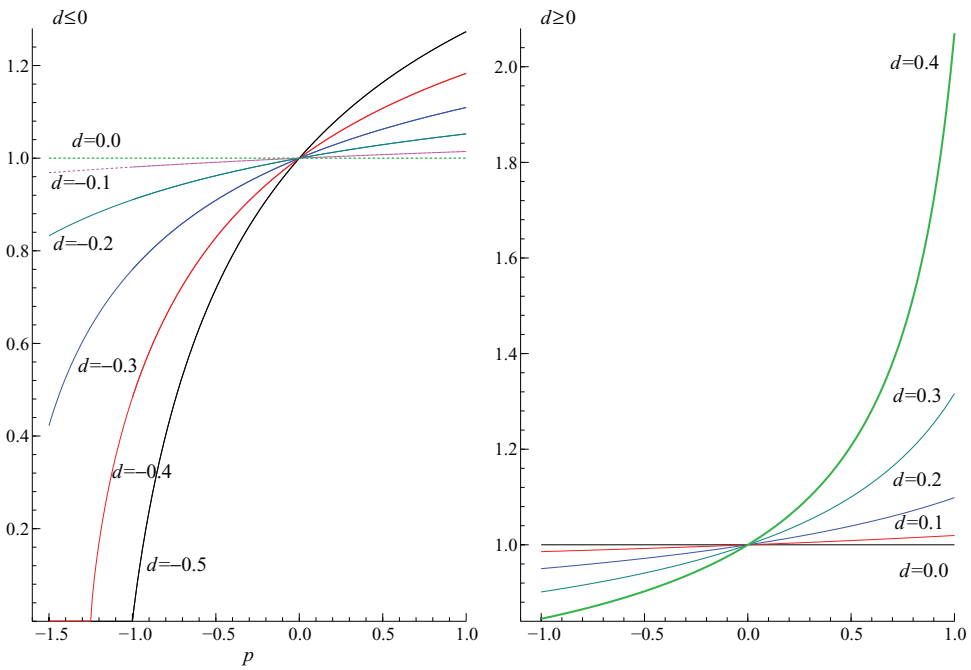

Figure 3. Variance profiles for fractional noise process with memory parameter d. The online version of this figure is in color.

short memory process with high persistence (e.g., an AR(1) with $\phi = 0.9$), is that $v_p \to \infty$ as $p \to (2d)^{-1}$, whereas for the latter it converges to the finite maximum of the spectral density.

5. 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:

$$\hat{v}_p = \left\{ \frac{1}{N} \sum_{j=1}^{N} (2\pi I(\omega_j))^p \, (\Gamma(p+1))^{-1} \right\}^{\frac{1}{p}}, \tag{8}$$

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 by Hannan and Nicholls 1977).

The factor $(\Gamma(p+1))^{-\frac{1}{p}}$ serves as a bias correction term, which 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 Equation (8) 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 Fréchet (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 chi-square random variable (Koopmans 1974, chap. 8),

$$I(\omega_j) = \begin{cases} \frac{1}{2} f(\omega_j) \chi^2_2, & 0 < \omega_j < \pi \\ f(\omega_j) \chi^2_1, & \omega_j = 0, \pi, \end{cases} \tag{9}$$

where χ^2_m denotes a chi-square 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

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

where the log-additive bias correction term γ is the Euler gamma, that is, 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 log-normal distribution for $\hat{\sigma}^2$ when n is not too large.

Hannan and Nicholls (1977) proposed replacing the raw periodogram ordinates by the nonoverlapping 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 in Equation (10) is obtained in the case $m = 1$. The large sample distributions of Equation (11) and its log transform are, respectively,

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

and the estimator results in a smaller mean square estimation error; increasing m reduces the variance but inflates the finite sample 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 in Equation (8),

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

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 in Equation (12) tends to that in Equation (11) (to Equation (10) when further $m = 1$).

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

Theorem 1. 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 $p > -\frac{m}{2}$,

(1) $\hat{v}_p(m)$ is consistent for v_p,
(2) $\sqrt{n}\{\hat{v}_p(m) - v_p\} \to_d \mathrm{N}(0, \mathrm{V}_p)$, where $\mathrm{V}_p = 2m(\frac{v_p}{p})^2 (\frac{v_{2p}}{v_p})^{2p} (\frac{\Gamma(m+2p)\Gamma(m)}{\Gamma^2(m+p)} - 1)$, and
(3) $\hat{v}_0(m) = \hat{\sigma}^2(m)$ and $\mathrm{V}_0 = 2m\sigma^4\psi'(m)$.

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 Gnedenko and Kolmogorov (1954) and essentially establishes a Lindeberg-Feller type condition that is easy to check.

The third statement of the theorem 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 in Equation (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.

In principle, to estimate v_p for $p \to -\infty$, the condition $p > -\frac{m}{2}$ requires $m \to \infty$. As we shall see in the following section, where the sampling properties of $\hat{v}_p(m)$ will be assessed, a reliable estimation of v_p using a large value of m requires a very large sample. In many cases, however, being $f(\omega) > 0$ and v_p nonnegative and increasing, it happens that v_p tends to zero for values of $p > -\frac{m}{2}$, with $m \ll \infty$; see for example Figures 1–3, where the variance profile tends to zero for values of p that are close to -1.

5.1 The Sampling Properties of $\hat{v}_p(m)$

The mean square error (MSE) of $\hat{v}_p(m)$ is a rather complicated function of p, m, and the spectral properties of x_t. While analytical formulas are difficult to obtain, the jackknife (Quenouille 1949; see Miller 1974; Efron and Tibshirani 1993, for reviews) provides an effective estimator that can serve the purpose of selecting the bandwidth parameter m, as it will be illustrated shortly.

The complexity of the dependence of the MSE is illustrated by Figure 4, which displays the logarithm of the MSE of the prediction error variance estimator $\hat{v}_0(m)$, when x_t is an AR(1) process with parameter $\phi = 0.8$, as a function of the sample size n and the bandwidth m. Similar "concertina" shapes are observed for different values of $p \neq 1$, the jumps occurring at the points of discontinuity of $M = [(n-1)/2m]$. When $p = 1$, the estimator is centered around the Riemannian approximation of v_1, $(mM)^{-1}\sum_j \sum_k 2\pi f(\omega_{jm+k})$, with $\text{var}(\hat{v}_p(1)) \approx (mM)^{-1}[(mM)^{-1}\sum_j \sum_k (2\pi f(\omega_{jm+k}))^2]$, where the factor in parenthesis converges to v_2^2; hence, for a fixed sample size, apart from minor effects that make mM deviate from $[(n-1)/2]$, the MSE is independent of m.

The fact that the MSE has multiple minima makes the selection of the optimal bandwidth hard. In practice, the variance and the component of the bias that depends on m can be estimated using the jackknife method. The jackknife estimate of the variance is

$$\hat{V}_p(m) = \frac{n^* - 1}{n^*} \sum_{r=1}^{n^*} \left(\hat{v}_{p,r}(m) - \hat{v}_{p,\cdot}(m)\right)^2, \; \hat{v}_{p,\cdot}(m) = \frac{1}{n^*} \sum_r \hat{v}_{p,r}(m),$$

where $\hat{v}_{p,r}(m)$ is the variance profile estimator in Equation (12) when the periodogram ordinate $I(\omega_r)$ is omitted, $r = 1, \ldots, n^* = [(n-1)/2]$; the above estimate is valid under the assumption that the periodogram ordinates are independent (see Equation 9). Also, writing the bias, $\text{B}[\hat{v}_p(m)] = \text{E}[\hat{v}_p(m)] - v_p$, as

$$\begin{aligned} \text{B}[\hat{v}_p(m)] &= \text{E}[\hat{v}_p(m) - \hat{v}_p(1)] + \text{E}[\hat{v}_p(1)] - v_p \\ &= \text{E}[\hat{v}_p(m) - \hat{v}_p(1)] + \text{B}[\hat{v}_p(1)], \end{aligned} \tag{14}$$

it can be seen that the bias depends on m solely via the first term, which can be estimated from averaging out the leave-one-out cross-validatory differences $\hat{v}_{p,r}(m) - \hat{v}_{p,r}(1)$. The second term is the bias of the raw estimator in Equation (8) and depends (for n fixed) on p and $f(\omega)$.

We ran a Monte Carlo experiment consisting of 1000 replications of an AR(1) process with parameters $\phi = 0.8$, $\sigma^2 = 1$, of length $n = 90, 200, 500$. The case $n = 90$ is used for comparison with the simulation results by Hannan and Nicholls (1977). For each replication, we estimate the prediction error variance by $\hat{v}_0(m)$, $m = 1, \ldots, 12$, and evaluate the jackknife estimate of the variance and the bias term $\text{E}[\hat{v}_0(m) - \hat{v}_0(1)]$. The output of the experiment is the Monte Carlo (MC) estimate of the MSE,

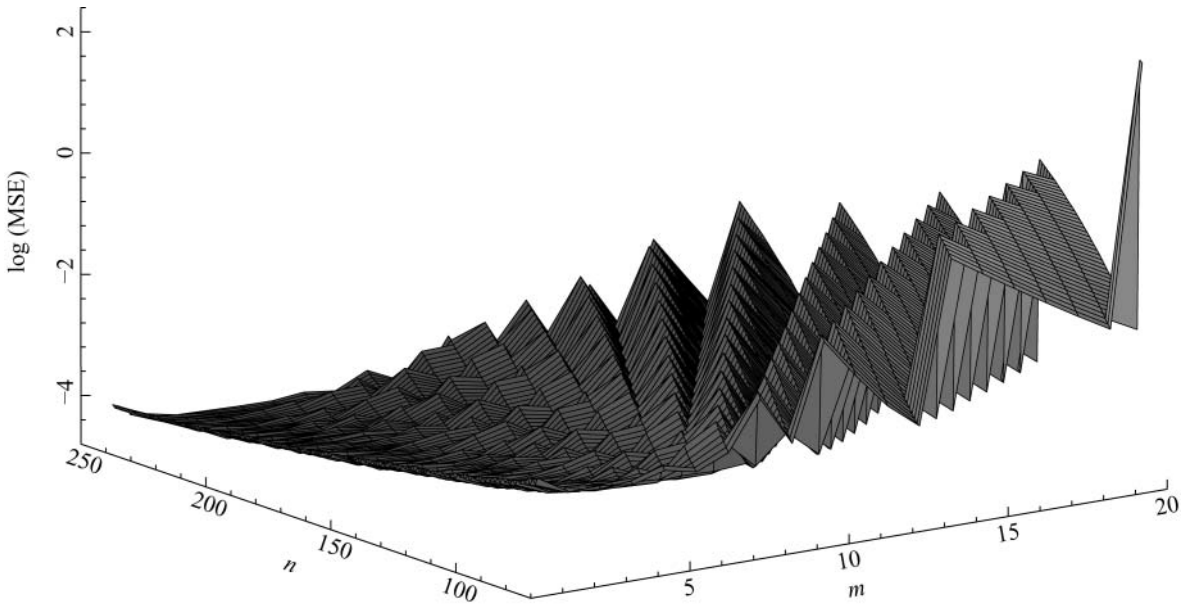

Figure 4. Logarithm of the MSE of the prediction error variance estimator $\hat{v}_0(m)$ for $n = 80, \ldots, 250$ and $m = 1, \ldots, 20$. The MSE is estimated by Monte Carlo simulation using 5000 replications.

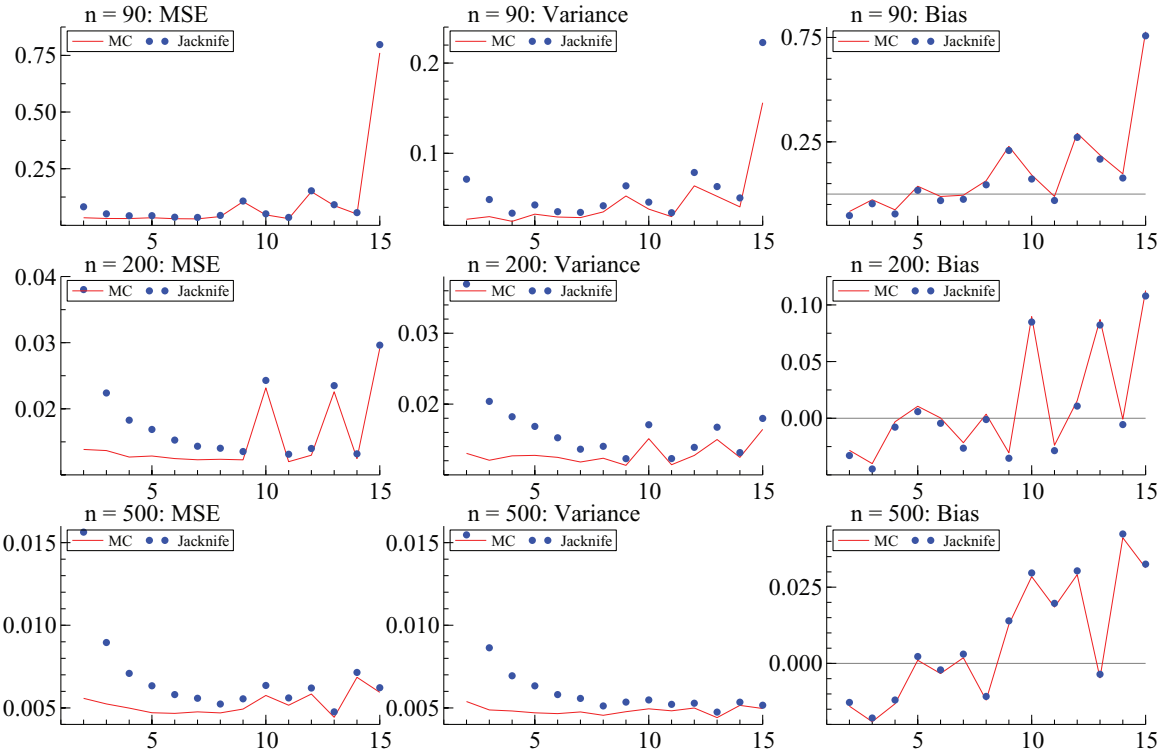

Figure 5. Plot of MSE, variance and bias of $\hat{v}_0(m)$ as a function of m for series of length $n = 90, 200, 500$. The online version of this figure is in color.

variance, and bias of the estimator; moreover, we can average across the 1000 replications the jackknife estimates of bias and variance to assess their reliability.

Figure 5 displays and compares the MC and the jackknife estimates of the MSE, variance, and bias. Each row corresponds to a different sample size. The plots confirm that the bias increases with m and decreases with n. The variance decreases with m, but with discontinuities that occur at the points of discontinuity of $\psi'(m)/M$. The jackknife estimates are upward biased for m small. The jackknife method provides a very accurate estimate of the bias, as it can be seen from the right plots. Although it estimates only the first term of Equation (14), it can be shown that for p around 0 the contribution of the term $\mathrm{B}[\hat{v}_p(1)]$ to the overall bias is small.

The overall conclusion is that for values of p for which $\mathrm{B}[\hat{v}_p(1)]$ is negligible, the MSE estimated by the jackknife method can provide reliable guidance over the selection of m. For instance, for $n = 90$ the choice $m = 4$ is suitable, whereas for $n = 500$, $m = 13$ appears to be the best choice.

5.2 Estimation of the Long Memory Parameter Based on the Variance Profile

A minimum distance estimator (MDE) of the long memory parameter d of a fractionally integrated process based on the variance profile can be constructed as the minimizer of a weighted distance between the sample and the theoretical variance profile in Equation (7):

$$\hat{d} = \min_{d \in D} \int_a^b k_p(\hat{v}_p(m) - v_p)^2 \mathrm{d}p, \tag{15}$$

where D and (a, b) are such that $-1 < d < 0.5$ and $dp < 0.5$. The weights k_p may be uniform or inversely related to the asymptotic variance V_p, in which case it should be noticed that for a FN process V_p exists provided that $pd < 0.25$, since V_p requires the existence of v_{2p}. If the range of integration (a, b) excludes all the points $pd < 0.25$ (e.g., $b < 0.625$ for $d = 0.4$), the minimum distance estimator $\hat{d}$ is consistent for d and $\sqrt{n}\{\hat{d} - d\} \to_d \mathrm{N}(0, \mathrm{V}_d)$, where

$$\mathrm{V}_d = \frac{\int_a^b \int_a^b v_p v_s [\psi(1-pd) - \psi(1-2pd)][\psi(1-sd) - \psi(1-2sd)] k_p k_s \gamma_{ps} \mathrm{d}p \mathrm{d}s}{\left\{\int_a^b v_p^2 [\psi(1-pd) - \psi(1-2pd)]^2 k_p \mathrm{d}p\right\}^2}$$

and γ_{ps} is the covariance between $\hat{v}_p(m)$ and $\hat{v}_s(m)$, as can be shown using the same arguments of Tieslau, Schmidt, and Baillie (1996), who derived a minimum distance estimator for d based on the sample and population autocorrelation function. In general, $\mathrm{V}_d \geq \frac{6}{\pi^2}$, the variance of the maximum likelihood (ML) estimator of d, which does not depend on d. On the other hand, for fixed m, V_d depends on d and also on p. It follows that for any d it is possible to choose a range of p where V_d is as close as possible to the asymptotic lower bound implied by Fisher information for d.

The estimator lends itself quite naturally to the interpretation as a feature matching estimator (see Xia and Tong 2011), in that in estimating d the focus is not in fitting a specific feature such as the prediction error variance (which would amount to optimizing the one-step ahead predictive performance of the fitted model), but in matching the sample and the population variance profile across a range of values of p that encompasses the prediction error variance, the interpolation error variance, and other process characteristics defined at different values of p (the power means of the spectral density).

A Monte Carlo experiment using 5000 replications from a FN process with long memory parameter $d = 0.4$ and $\sigma^2 = 1$ has been performed to assess the properties of the proposed estimator. In practice, we set $k_p = 1$ for all p, and we evaluate both Equations (7) and (12) on a regular grid of 100 values of p from $a > -m/2$ to b. The experiment also aims at establishing the optimal values of a and b.

For assessing the sampling properties of the estimator we report the bias, standard error, and the mean square error and for a comparative assessment we report the same quantities for the widely applied Geweke and Porter-Hudak (1983, G-PH henceforth) estimator

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

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 $\bar{w} = R^{-1}\sum_{j=1}^{R} w_j$. We include in the comparison the generalized linear regression (GLR) estimator proposed by Beran (1993), such that $\ln I(\omega_j)$ is regressed on a constant and w_j and the errors have a log-gamma distribution. Estimation is carried out by iteratively weighted least squares (see Cameron and Turner 1987). The GLR estimator is more accurate than G-PH as it embodies the correct distributional assumption on the nature of the disturbances. In fact, it is equivalent to Whittle's approximate maximum likelihood estimator and behaves in large samples as the exact maximum likelihood estimator (see Palma 2007, sec. 4.4).

5.2.1 Simulation Results. The results of the simulation experiment, reported in Table 1, indicate that the MDE estimator outperforms G-PH and can be made as efficient as the GLR estimator by a suitable choice of the interval (a, b). As m increases, further reductions in the MSE are available, as the variance of the estimator reduces and the bias is kept under control by varying (a, b), until the estimator in Equation (12) degenerates to a constant with respect to p and in Equation (15) estimates d equal to zero with zero variance and bias d^2 (for $n = 500$, this occurs for $m = 125$, in which case $M = [499/250] = 1$). The size of this interval increases with m. The lower extreme a decreases at a much slower rate than $-m/2$ and b increases very slowly. Also, as n increases from 500 to 1000, the length of the optimal interval reduces for a given m. On the contrary, if (a, b) is held fixed and m varies, the usual trade-off between bias and variance will be in operation. For $n = 1000$, the choice $(-2.2, 0.6)$ seems suitable, whereas for $n = 500$ one would consider $(-2.7, 0.6)$. Values outside this interval are discarded as the estimation error of v_p is larger. Interestingly, the upper integration limit b is smaller than 1: obviously, for a positive d, b needs to be smaller than $0.5/d$, which is the asymptote of v_p for a FN process. Similar considerations apply to other values of d and different sample sizes.

In conclusion, the proposed MDE estimator can be made as efficient as the maximum likelihood estimator based on the distributional result in Equation (9). Moreover, it does not suffer from the limitation affecting the minimum distance estimators based on the autocorrelations and the partial autocorrelations, that require $d < 0.25$.

6. EMPIRICAL ILLUSTRATIONS

6.1 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 6; 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 $m = 20$. The value of the bandwidth

Table 1. Estimation of d: bias, standard error, and MSE of the MDE estimator for different values of m, and G-PH and GLR estimators, based on 5000 replications of a FN process with $d = 0.4$. For the MDE, estimator a and b are the extremes of integration of the estimator in Equation (15)

	$n = 500$					$n = 1000$				
Estimator	a	b	Bias	Std. err.	MSE	a	b	Bias	Std. err.	MSE
MDE $m = 1$	−0.4	0.6	−0.06539	0.06743	0.00882	−0.4	0.6	−0.04138	0.04206	0.00348
MDE $m = 3$	−0.4	0.2	−0.02691	0.05539	0.00379	−0.4	0.4	−0.02957	0.03132	0.00186
MDE $m = 5$	−0.4	0.3	−0.03975	0.04287	0.00342	−0.4	0.2	−0.02116	0.03227	0.00149
MDE $m = 10$	−0.6	0.2	−0.02826	0.04330	0.00267	−0.5	0.2	−0.01685	0.02951	0.00115
MDE $m = 15$	−0.8	0.2	−0.02256	0.04306	0.00236	−0.6	0.2	−0.01427	0.02918	0.00105
MDE $m = 20$	−1.0	0.3	−0.01947	0.04303	0.00223	−0.7	0.2	−0.00844	0.02979	0.00096
MDE $m = 25$	−1.8	0.5	−0.01548	0.04237	0.00203	−1.1	0.4	−0.00825	0.02910	0.00092
MDE $m = 30$	−1.9	0.5	−0.01674	0.04142	0.00200	−1.6	0.5	−0.00855	0.02838	0.00088
MDE $m = 40$	−2.8	0.6	−0.01482	0.04039	0.00185	−2.3	0.6	−0.00805	0.02761	0.00083
MDE $m = 50$	−3.4	0.7	−0.01135	0.03775	0.00155	−2.0	0.6	−0.00244	0.02779	0.00078
G-PH, $R = [n/16]$			0.00711	0.13732	0.01891			0.00533	0.09472	0.00900
G-PH, $R = [n/8]$			0.00600	0.09117	0.00835			0.00381	0.06707	0.00451
G-PH, $R = [n/4]$			0.00572	0.06237	0.00392			0.00278	0.04574	0.00210
G-PH, $R = n$			0.00329	0.04725	0.00224			0.00202	0.03391	0.00115
GLR			0.00159	0.03749	0.00141			0.00107	0.02694	0.00073

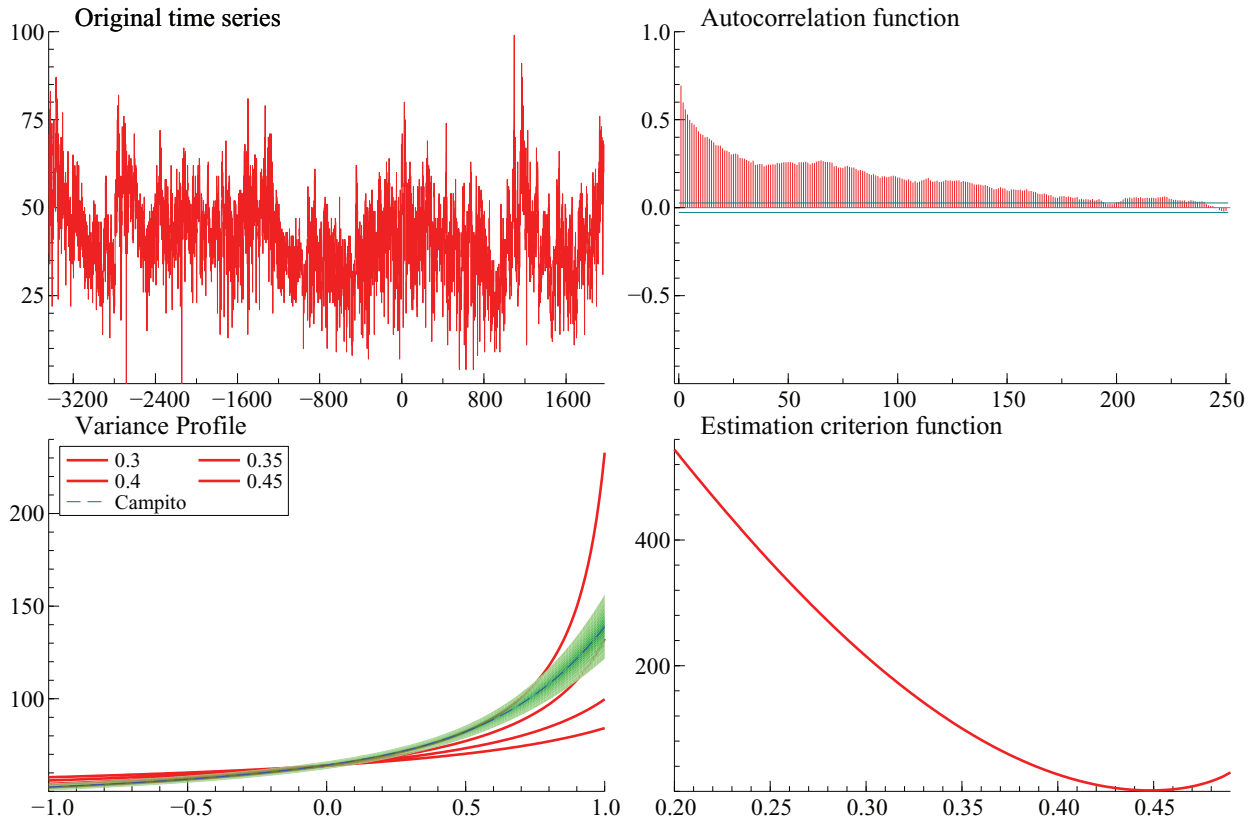

Figure 6. Mount Campito tree rings data: series (top left panel); sample autocorrelation function (top right), estimate of the variance profile using $m = 20$ (bottom left) and estimation criterion (15), where the variance profile is evaluated at an equally spaced grid of 100 values in the range $(-2.0, 0.6)$. The online version of this figure is in color.

parameter was chosen as the minimizer of the MSE estimated by the jackknife method, for p in the range (–0.5, 0.5). The long memory parameter is estimated equal to 0.453, a value in accordance with the literature (see, e.g., Baillie 1996, p. 45), where the ML estimate of d is 0.449. The estimation criterion function is plotted in the last panel.

6.2 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 and 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 January 3, 1928–August 30, 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 January 3, 1989–March 7, 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 the normalized Box-Cox transform or the standardized one should be considered. The former is obtained by dividing $x_t(\lambda)$ by $\sqrt[n]{J}$, where $J = \prod_t |\frac{\partial x_t(\lambda)}{\partial x_t}|$ 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 normalized variance profile for $p < 1$. In other words, we will constraint $v_1 = 1$ for all the λ values. Setting $m = 17$, we estimate the variance profile for the standardized transformed series for values of λ in the interval $(-0.5, 2.3)$. The results for the SP500 series are presented in Figure 8.

The variance profile of the standardized $x_t(\lambda)$ can be used to determine the value of the transformation parameter for which the long memory property is strongest. Figure 7 plots the value of d, estimated according to Section 5, against λ. It turns out that for both series the maximum d is achieved for λ around 1.25. The MLE shows a similar pattern. 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 8, which refers to the SP500 series, displayed in the top right panel. The plot also shows that the normalized 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 absolute returns. Another conclusion is that the volatility of Nasdaq returns is more predictable than SP500s.

6.3 The Great Moderation

The term Great Moderation (GM) refers to a substantive reduction in the volatility of macroeconomic fluctuations that

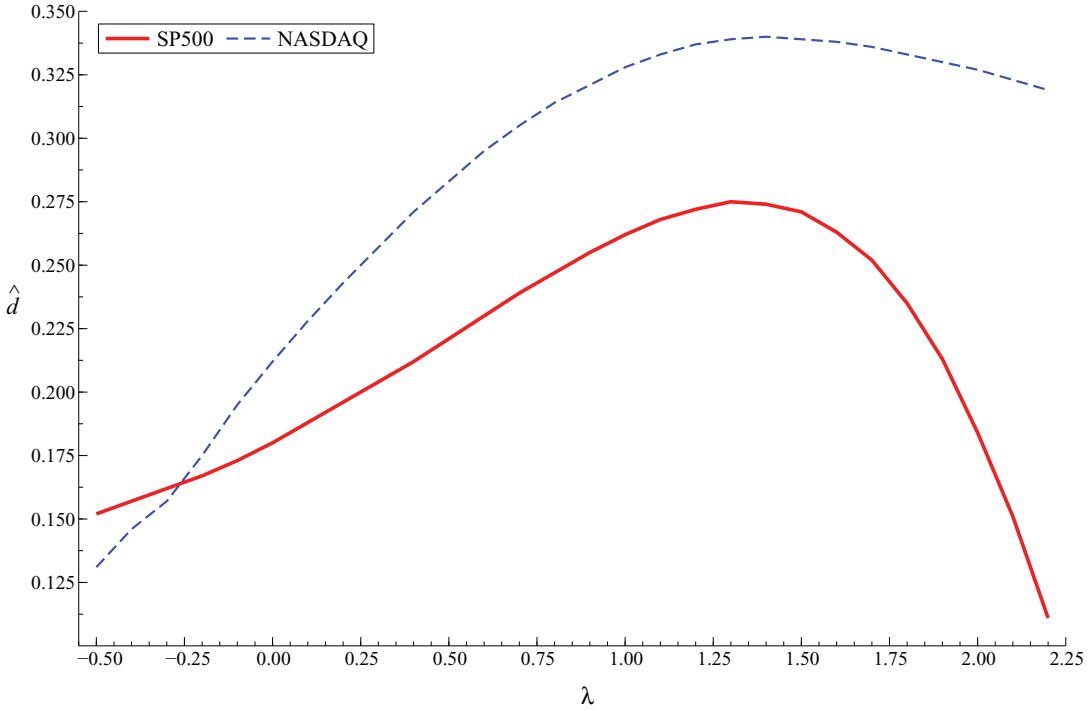

Figure 7. Nasdaq and SP500 daily absolute returns: estimates of the long memory parameter d based on the variance profile as a function of the transformation parameter λ. The online version of this figure is in color.

took place around the mid-1980s up to the inception of the last recession (around 2008; see, among others, McConnell and Perez-Quiros 2000). The causes of this well-documented phenomenon have been the matter of 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).

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 forms. We set off analyzing the series of yearly growth rates for the period January 1949–June 2008, which is split into two subseries, the first covering the

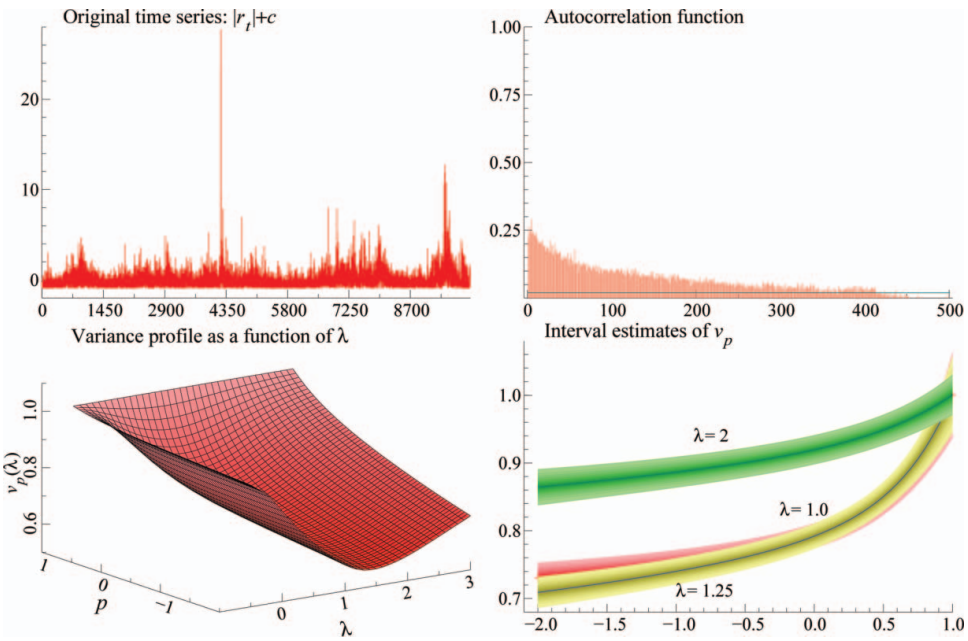

Figure 8. Standard and Poor 500 daily absolute returns (standardized): variance profile as a function of the Box-Cox transformation parameter: series (top left panel); sample autocorrelation function (top right), estimate of the variance profile using $m = 17$ (bottom left) and comparison of the interval estimates for $\lambda = 1$, 1.25 and $\lambda = 2$. The online version of this figure is in color.

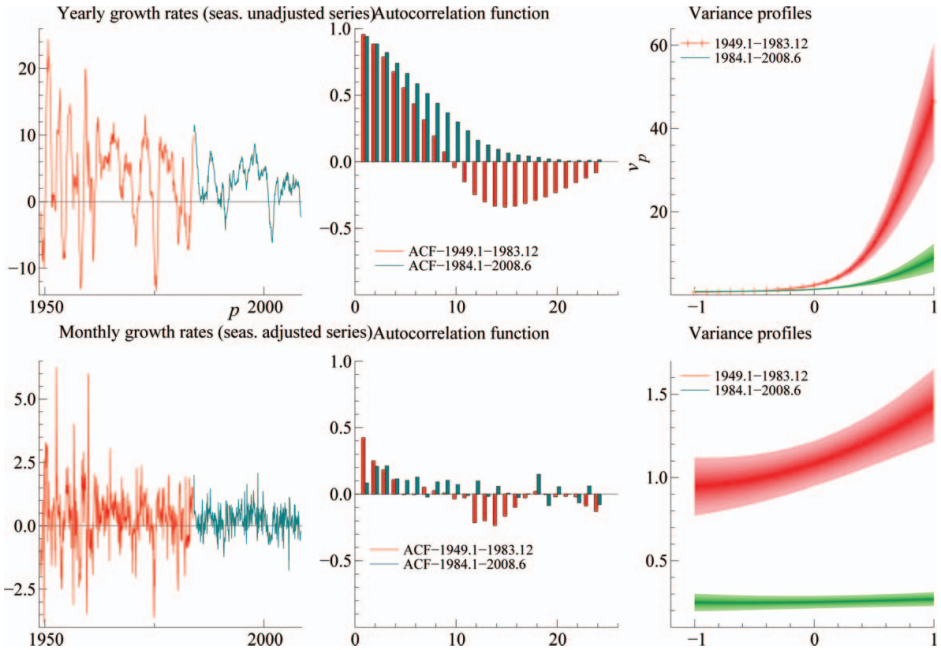

Figure 9. U.S. Index of industrial production, yearly (seasonally unadjusted) and monthly growth rates (seasonally adjusted) for the two subperiods 1949.1–1983.12 (pre) and 1984.1–2008.6 (Great Moderation). Comparison of autocorrelation functions and variance profiles. The online version of this figure is in color.

period predating the GM (January 1949–December 1983) and the second covering the GM period (January 1984–June 2008). The series are plotted in Figure 9; the volatility reduction is indeed visible and the patterns of the autocorrelations are also different—the behavior 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 noninvertible 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 9), we also find a significant change in the variance profile, which flattens down 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 $1 - v_{-1}/v_1$.

7. CONCLUSIONS

The variance profile, v_p, provides a characterization of a stochastic process that is useful for feature assessment and feature matching. The article has focused on the derivation of a minimum distance estimator of the long memory parameter. Another application along this direction is time series discrimination and clustering: our preliminary experience (not reported for brevity) on the discrimination of AR and MA processes suggests that focusing on the distance between the variance profiles provides a very effective strategy (more effective than using the distance between the autocorrelation functions in the MA vs. AR case).

In feature assessment/matching, it is important to consider p as a continuous parameter; restricting attention only to few selected values, namely $p = 0, \pm 1$, results in a much poorer feature assessment/matching. From the theoretical standpoint, considering the variance profile as a continuous function of p enables the direct derivation of fundamental time series results.

Also, having a nonparametric estimator that is a continuous function of p, as it is based on the power transformation of the pooled periodogram, is relevant for the estimation of ratios or contrasts based on different values of v_p (i.e., relative measures of predictability and interpolability). Other possible applications of the variance profile, or its transformation $z_p = p(\log v_p - \log v_0)$, are in the field of goodness-of-fit tests. Davis and Jones (1968) discussed the test of the null $H_0 : z_1 = 0$ as a test for white noise; the test can be extended to other values of p (e.g., $H_0 : \log v_1 - \log v_{-1} = 0$ and $H_0 : z_{-1} = 0$ are also tests of the white noise hypothesis) and a combination of such tests across different values of p may prove more efficient. The test of $H_0 : 2(\log v_2 - \log v_1) = 0$, under which the sum of the squared autocorrelation at all lags is 1, is related to the goodness-of-fit test of Milhoj (1981; see also Beran 1992; Hong 1996).

We leave to future research the estimation of the variance profile using tapered and multitapered periodograms (see Pukkila and Nyquist 1985; Walden 1995, 2000; and, more recently, Kohli and Pourahmadi 2012) and the comparison with alternative parametric estimators based on autoregressive model fitting, as in Cleveland (1972) and Bhansali (1993). Finally, a multivariate generalization of the variance profile can be carried on in the spirit of Mohanty and Pourahmadi (1999).

APPENDIX

We provide the proof of the consistency and asymptotic normality of the estimator in Equation (12). 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 $1 \le k \le m$, $\sum_{k=1}^{m} \frac{I(\omega_{jm+k})}{\frac{1}{2} f(\omega_{jm+\frac{m+1}{2}})} \sim \chi^2_{2m}$ (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,

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

By direct integration,

$$\mathrm{E}\left(X_j^p\right) = \frac{\Gamma(m+p)}{\Gamma(m)}, \tag{A.2}$$

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

$$\mathrm{var}\left(X_j^p\right) = \frac{\Gamma(m+2p)}{\Gamma(m)} - \frac{\Gamma^2(m+p)}{\Gamma^2(m)}, \tag{A.3}$$

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

$$Z_j = \frac{X_j^p - \frac{\Gamma(m+p)}{\Gamma(m)}}{\sqrt{\frac{\Gamma(m+2p)}{\Gamma(m)} - \frac{\Gamma^2(m+p)}{\Gamma^2(m)}}}, \tag{A.4}$$

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,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} (2\pi f(\omega))^{2p}\,\mathrm{d}\omega = \lim_{M\to\infty} \frac{1}{M}\sum_{j=0}^{M-1}\left(2\pi f\left(\omega_{jm+\frac{m+1}{2}}\right)\right)^{2p} \tag{A.5}$$

the quantity $Q_M = \frac{1}{M}\sum_{j=0}^{M-1}(2\pi f(\omega_{jm+\frac{m+1}{2}}))^{2p}$ exists and has a limit, $\lim_{M\to\infty} Q_M = v_{2p}^{2p}$.

Let now

$$b_j = \frac{\left(2\pi f\left(\omega_{jm+\frac{m+1}{2}}\right)\right)^p}{\sqrt{MQ_M}}, \tag{A.6}$$

which satisfies $\sum_{j=0}^{M-1} b_j^2 = 1$. Moreover, since the pth 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 Equation (A.4) and $\sum_{j=0}^{M-1} b_j^2 = 1$ 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 Equation (A.6),

$$\begin{aligned}\{\hat{v}_p(m)\}^p &= \frac{1}{M}\sqrt{MQ_M}\sum_{j=0}^{M-1} b_j X_j^p \frac{\Gamma(m)}{\Gamma(m+p)} \\ &\to_d \mathrm{N}\left(\frac{1}{M}\sum_{j=0}^{M-1}\left(2\pi f\left(\omega_{jm+\frac{m+1}{2}}\right)\right)^p, \Omega_M\right)\end{aligned} \tag{A.7}$$

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

$$\sqrt{M}\left(\{\hat{v}_p(m)\}^p - v_p^p\right) \to_d \mathrm{N}\left(0, v_{2p}^{2p}\left(\frac{\Gamma(m+2p)\Gamma(m)}{\Gamma^2(m+p)} - 1\right)\right) \tag{A.8}$$

and applying the delta method we finally get the asymptotic distribution

$$\begin{aligned}&\sqrt{n}(\hat{v}_p(m) - v_p) \\ &\to_d \mathrm{N}\left(0, 2m\left(\frac{v_p}{p}\right)^2\left(\frac{v_{2p}}{v_p}\right)^{2p}\left(\frac{\Gamma(m+2p)\Gamma(m)}{\Gamma^2(m+p)} - 1\right)\right).\end{aligned} \tag{A.9}$$

We now prove the consistency of $\hat{v}_p(m)$ for v_p, that is a consequence of three facts: (1) the Chebychev weak law of large numbers, applied to the sequence of random variables

$$\sqrt{MQ_M} b_j X_j \frac{\Gamma(m)}{\Gamma(m+p)}$$

in $\hat{v}_p(m)^p$ (see Equation A.7); (2) the convergence of the Riemannian sum to the integral (see Equation A.5); and (3) 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 of v_p, given that the power function is continuous.

Let us now consider $p \to 0$. In this case, the estimator in Equation (12) equals the prediction error variance estimator in Equation (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 than its power, that is when $p \to 0$, X_j^p is to be read as $\log X_j$. Hence, $\mathrm{E}(\exp\{t \log X_j\})$, given in Equation (A.2), is the moment generating function of $\log X_j$ and gives $\mathrm{E}(\log X_j) = \psi(m)$ and $\mathrm{var}(\log X_j) = \psi'(m)$. Therefore, when $p \to 0$, Equation (A.1) 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 $\mathrm{E}\log(\sum_{k=1}^{m} 2\pi I(\omega_{jm+k})) = \log 2\pi f(\omega_{jm+\frac{m+1}{2}}) + \psi(m)$ and $\mathrm{var}\log(\sum_{k=1}^{m} 2\pi I(\omega_{jm+k})) = \psi'(m)$, respectively. What follows is that in the limit case, the bias correction via a multiplication (by the inverse expected value, see Equation A.7) 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, $\log\hat{\sigma}^2(m) = \frac{1}{M}\sum_{j=0}^{M-1}(\log 2\pi f(\omega_{jm+\frac{m+1}{2}}) + \log X_j - \psi(m))$, that is, 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 $m = 1$ is a particular case of Equation (12). However, one could note that when $m = 1$, the estimator in Equation (12) becomes $\hat{v}_p = \{\frac{1}{N}\sum_{j=1}^{N}[2\pi I(\omega_j)]^p[\Gamma(p+1)]^{-1}\}^{\frac{1}{p}}$ and the random variables involved in its asymptotic distributions can be written as monotonic transforms of χ_2^2 random variables, as $Y_j = [2\pi I(\omega_j)]^p = [\pi f(\omega_j)\chi_2^2]^p$. It follows that for $p \neq 0$, by applying the density transform method, one gets

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

When p is positive and finite, then $f_{Y_j}(y)$ 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 Fréchet random variables with the same parameters. Note that when $p \to 0$ we find the Gumbel distribution, that is, 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 $\mathrm{E}(Y_j) = [2\pi f(\omega_j)]^p\Gamma(p+1)$ and $\mathrm{var}(Y_j) = [2\pi f(\omega_j)]^{2p}[\Gamma(2p+1) - \Gamma^2(p+1)]$, respectively.

It follows that the random variables $Z_j = Y_j\Gamma(p+1)^{-1}$ have mean and variance given by $\mathrm{E}(Z_j) = [2\pi f(\omega_j)]^p$ and $\mathrm{var}(Z_j) = [2\pi f(\omega_j)]^{2p}[\Gamma(2p+1)\Gamma^{-2}(p+1) - 1]$, respectively, and since they are uniformly bounded, the Lindeberg-Feller central limit theorem applies and we get Equations (A.8) and (A.9) 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.

[Received April 2011. Revised March 2012.]

REFERENCES

Ash, R. B., and Gardner M. F. (1975), *Topics in Stochastic Processes*, New York: Academic Press. [608]

Atkinson, A. C. (1985), *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Oxford: Oxford University Press. [616]

Baillie, R. T. (1996), "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics*, 73 (5), 1–59. [615]

Baltagi, B. H. (2005), *Econometric Analysis of Panel Data*, New York: Wiley. [610]

Battaglia, F., and Bhansali, R. (1987), "Estimation of the Interpolation Error Variance and an Index of Linear Determinism," *Biometrika*, 74 (4), 771–779. [607,608,609]

Beran, J. (1993), "Fitting Long-Memory Models by Generalized Linear Regression," *Biometrika*, 80 (4), 817–822. [615]

——— (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. [618]

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. [618]

Box, G. E. P., and Cox, D. R. (1964), "An Analysis of Transformations (with discussion)," *Journal of the Royal Statistical Society,* Series B, 26, 211–246. [616]

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods*, New York: Springer-Verlag. [608]

Burke, S., Godfrey, L. G., and Tremayne, A. R. (1990), "Testing AR(1) Against MA(1) Disturbances in the Linear Regression Model: An Alternative Procedure," *Review of Economic Studies*, 57 (1), 135–145. [610]

Cameron, M. A., and Turner, T. R. (1987), "Fitting Models to Spectra Using Regression Packages," *Journal of the Royal Statistical Society,* Series C, 36 (1), 47–57. [615]

Chatfield, C. S. (1979), "Inverse Autocorrelations," *Journal of the Royal Statistical Society,* Series A, 142 (3), 363–377. [609]

Cleveland, W. S. (1972), "The Inverse Autocorrelations of a Time Series and Their Applications," *Technometrics*, 14 (2), 277–293. [609,618]

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. [607,612,618,619,620]

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. [616]

Doob, J. L. (1953), *Stochastic Processes*, New York: Wiley. [608]

Efron, B., and Tibshirani R. J., (1993), *An Introduction to the Bootstrap*, Boca Raton, FL: Chapman & Hall. [613]

Eicker, F. (1963), "Central Limit Theorems for Families of Sequences of Random Variables," *The Annals of Mathematical Statistics*, 34 (2), 439–446. [612,619]

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. [615]

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. [612,619]

Gnedenko, B. V., and Kolmogorov, A. N. (1954), *Limit Distributions for Sums of Random Variables*, Cambridge, MA: Addison-Wesley. [612,619]

Gould, H. W. (1974), "Coefficient Identities for Powers of Taylor and Dirichlet Series," *American Mathematical Monthly*, 81 (1), 3–14. [608]

Graham, R. L., Knuth, D. E., and Patashnik, O. (1994), *Concrete Mathematics*, Cambridge, MA: Addison-Wesley. [608,609]

Grenander, U., and Rosenblatt, M. (1957), *Statistical Analysis of Stationary Time Series*, New York: Wiley. [608,611]

Hannan, E. J. (1970), *Multiple Time Series*, New York: Wiley. [607,608]

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. [612,613]

Hannan, E. J., Terrell, R. D., and Tuckwell, N. E. (1970), "The Seasonal Adjustment of Economic Time Series," *International Economic Review*, 11, 24–52. [610]

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1934), *Inequalities*, Cambridge: Cambridge University Press. [608]

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press. [610]

Haywood, J., and Tunnicliffe-Wilson, G. (2000), "An Improved State Space Representation for Cyclical Time Series," *Biometrika*, 87 (3), 724–726. [610]

Hong, Y. (1996), "Consistent Testing for Serial Correlation of Unknown Form," *Econometrica*, 64 (4), 837–864. [618]

Johnson N. L., and Kotz, S. (1972), "Power Transformations of Gamma Variables," *Biometrika*, 59 (1), 226–229. [612]

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. [608]

King, M. L., and McAleer, M. (1987), "Further Results on Testing AR (1) Against MA (1) Disturbances in the Linear Regression Model," *Review of Economic Studies*, LIV, 649–663. [610]

Kohli, P., and Pourahmadi, M. (2012), "Nonparametric Estimation of the Innovation Variance and Judging the Fit of ARMA Models," in *Economic Time Series: Modeling and Seasonality*, eds. W. R. Bell, S. H. Holan and T. McElroy, London: Chapman & Hall, chap. 19. [618]

Kolmogorov, A. N. (1941), "Interpolation and Extrapolation of Stationary Random Sequences" (in Russian), *Izvestiya Akademii Nauk SSSR. Seriya Matematicheskaya*, 5, 3–14. [608]

——— (1992), "Interpolation and Extrapolation of Stationary Random Sequences," in *Selected Works of A.N. Kolmogorov, Volume II: Probability and Mathematical Statistics*, ed. A.N. Shiryayev, Dordrecht, The Netherlands: Kluwer Academic, pp. 272–280. [608]

Koopmans, L. H. (1974), *The Spectral Analysis of Time Series*, London: Academic Press. [612,618]

Luati, A., and Proietti, T. (2010), "Hyper-Spherical and Elliptical Stochastic Cycles," *Journal of Time Series Analysis*, 31, 169–181. [610]

Masani, P. (1960), "The Prediction Theory of Multivariate Stochastic Processes, III: Unbounded Spectral Densities," *Acta Mathematica*, 104 (1–2), 141–162. [609]

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. [616]

Milhoj, A. (1981), "A Test of Fit in Time Series Models," *Biometrika*, 68 (1), 177–187. [618]

Miller, R. G. (1974), "The Jackknife: A Review," *Biometrika*, 61 (1), 1–15. [613]

Mohanty, R., and Pourahmadi, M. (1999), "Estimation of the Generalized Prediction Error Variance of a Multiple Time Series," *Journal of the American Statistical Association*, 91 (433), 294–299. [618]

Nelson, C. R. (1976), "The Interpretation of R^2 in Autoregressive-Moving Average Time Series Models," *American Statistician*, 30, 175–180. [609]

Nicholls, D. F., Pagan, A. R., and Terrell R. D., (1975), "The Estimation and Use of Models With Moving Average Disturbance Terms: A Survey," *International Economic Review*, 16 (1), 113–134. [610]

Palma, W. (2007), *Long Memory Time Series: Theory and Methods*, Hoboken, NJ: Wiley. [610,615]

Percival, D. B., and Walden, A. T. (1993), *Spectral Analysis for Physical Applications*, Cambridge: Cambridge University Press. [607]

Pourahmadi, M. (2001), *Foundations of Time Series Analysis and Prediction Theory*, New York: Wiley. [608]

Priestley, M. B. (1981), *Spectral Analysis and Time Series, Volume 2: Multivariate Series, Prediction and Control*, New York: Academic Press. [608]

Pukkila, T., and Nyquist, H. (1985), "On Frequency Domain Estimation of the Innovation Variance of a Stationary Univariate Time Series," *Biometrika*, 72, 317–323. [618]

Quenouille, M. H. (1949), "Approximate Tests of Correlation in Time Series," *Journal of the Royal Statistical Society,* Series B, 11, 68–84. [613]

Salehi, H. (1979), "Algorithms for Linear Interpolator and Interpolation Error for Minimal Stationary Stochastic Processes," *The Annals of Probability*, 7 (5), 840–846. [609]

Silvapulle, P., and King, M. L. (1991), "Testing Moving Average Against Autoregressive Disturbances in the Linear-Regression Model," *Journal of Business and Economic Statistics*, 9 (3), 329–335. [610]

Szegö, G. (1920), "Beitrage zur Theorie der Toeplitzen Formen (Ersten Mitteilung)," *Mathematische Zeitschrift*, 6, 167–202. [608]

Szegö, G., and Askey, R. (1982), *Gabor Szegö: Collected Papers, Vol. 1: 1915–1927*, Basel: Birkauser. [608]

Taniguchi, M. (1980), "On Estimation of the Integrals of Certain Functions of Spectral Density," *Journal of Applied Probability*, 17 (1), 73–83. [619]

Tieslau, M. A., Schmidt, P., and Baillie, R. T. (1996), "A Minimum Distance Estimator for Long-Memory Processes," *Journal of Econometrics*, 71, 249–264. [614]

Tong, H. (1979), "Final Prediction Error and Final Interpolation Error: A Paradox?," *IEEE Transactions on Information Theory*, 25 (6), 758–759. [607,608]

Walden, A. T. (1994), "Interpretation of Geophysical Borehole Data via Interpolation of Fractionally Differenced White-Noise," *Journal of the Royal Statistical Society,* Series C, 43, 335–345. [611]

——— (1995), "Multitaper Estimation of Innovation Variance of a Stationary Time Series," *IEEE Transactions on Signal Processing*, 43, 181–187. [618]

——— (2000), "A Unified View of Multitaper Multivariate Spectral Estimation," *Biometrika*, 87, 767–788. [618]

West, M., and Harrison, J. (1989), *Bayesian Forecasting and Dynamic Models* (1st ed.), New York: Springer-Verlag. [610]

——— (1997), *Bayesian Forecasting and Dynamic Models* (2nd ed.), New York: Springer-Verlag. [610]

Whittle, P. (1983), *Prediction and Regulation by Linear Least Squares* (2nd ed.), Minneapolis, MN: University of Minnesota Press. [607,608]

Wiener, N. (1949), *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, Cambridge, MA: The MIT Press. [608]

Xia, Y., and Tong, H. (2011), "Feature Matching in Time Series Modeling," *Statistical Science*, 26 (1), 21–46. [607,614]