LONG-MEMORY TIME SERIES

P.M. Robinson*

London School of Economics

August 2, 2018

Abstract

Issues in the development of long memory time series analysis are discussed. Articles included in the volume are introduced, and related to the wider literature.

Revised version published in *Time Series With Long Memory* (P. M. Robinson, ed.), Oxford University Press, Oxford (2003), 1-32.

*Research supported by a Leverhulme Trust Personal Research Professorship and ESRC Grant R000238212. I thank Fabrizio Iacone for careful checking of bibliographic details and for locating typographical errors.

1. INTRODUCTION

Long memory has usually been described in terms of autocovariance or spectral density structure, in case of covariance stationary time series. Let x_t, $t = 0, \pm 1, ...$, be a time series indexed by time, t. It is covariance stationary, meaning that $E(x_t) = \mu$ and $Cov(x_t, x_{t+j}) = \gamma(j)$ do not depend on t. If x_t has absolutely continuous spectral distribution function, then it has a spectral density, given formally by

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{-ij\lambda}, \quad -\pi \leq \lambda \leq \pi; \tag{1.1}$$

$f(\lambda)$ is a non-negative, even function, periodic of period 2π when extended beyond the "Nyqvist" range $[-\pi, \pi]$. It is then common to say that x_t has long memory if

$$f(0) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) = \infty, \tag{1.2}$$

so that $f(\lambda)$ has a "pole" at frequency zero. The opposite situation of a zero at $\lambda = 0$,

$$f(0) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) = 0, \tag{1.3}$$

is sometimes referred to as "negative dependence" or "anti-persistence". The former term is natural because the second equality in (1.3) can only hold if the, positive, variance $\gamma(0)$ is balanced by predominantly negative autocovariances $\gamma(j)$, $j \neq 0$. We then say that x_t has "short memory" if

$$0 < f(0) < \infty. \tag{1.4}$$

These descriptions face the criticism that, consistent with (1.4), there is the possibility that $f(\lambda)$ has one or more poles or zeros at frequencies $\lambda \in (0, \pi]$, indicative of notable cyclic behaviour. We shall later refer to the modelling of such phenomena, but the bulk of interest has focussed on zero frequency. Going back to the 1960's, experience of nonparametric spectral estimation for many economic time series has suggested very marked peakedness around zero frequency, see Adelman (1965), to lend support to (1.2). However empirical evidence of long memory in various fields, such as astronomy, chemistry, agriculture and geophysics, dates from much earlier times, see for example Newcomb (1886), Student (1927), Fairfield Smith (1938), Jeffreys (1939), Hurst (1951). One aspect of interest was variation in the sample mean, $\bar{x} = n^{-1}\sum_{t=1}^{n} x_t$. If $f(\lambda)$ is continuous and positive at $\lambda = 0$, Féjèr's theorem indicates that

$$Var(\bar{x}) = \frac{1}{n}\sum_{j=1-n}^{n-1}\left(1 - \frac{|j|}{n}\right)\gamma(j) \tag{1.5}$$

$$\sim \frac{2\pi f(0)}{n}, \quad \text{as } n \to \infty, \tag{1.6}$$

where "$\sim$" indicates that the ratio of left and right sides tends to 1. But the empirical basis for this n^{-1} rate has been questioned, even by early experimenters, for example Fairfield Smith (1938) fitted a law $n^{-\alpha}$, $0 < \alpha < 1$ to spatial agricultural data. For future convenience, we change notation to $d = (1 - \alpha)/2$, later explaining why d is referred to as the "differencing" parameter. Fairfield Smith's law for the variance of the sample mean is thus n^{2d-1}, which from (1.5) is easily seen to arise if

$$\gamma(j) \sim c_1 j^{2d-1}, \quad \text{as } j \to \infty, \tag{1.7}$$

for $c_1 > 0$. Under additional conditions (see Yong, 1974), (1.7) is equivalent to a corresponding power law for $f(\lambda)$ near zero frequency,

$$f(\lambda) \sim c_2 \left|\lambda\right|^{-2d}, \quad \text{as } \lambda \to 0, \tag{1.8}$$

for $c_2 > 0$.

The behaviour of the sample mean under such circumstances, and the form and behaviour of the best linear unbiased estimate of the population mean, was discussed by Adenstedt (1974) (included as Chapter 2 of this volume). Adenstedt anticipated the practical usefulness of (1.8) in the long memory range $0 < d < \frac{1}{2}$, but also treated the anti-persistent case $-\frac{1}{2} < d < 0$, where (1.3) holds, as it does also for $d = -1$, which arises if a short memory process (see (1.4)) is first-differenced; Vitale (1973) had earlier discussed similar issues in this case. The sample mean tends to be highly inefficient under anti-persistence, but for long memory Samarov and Taqqu (1988) found it to have remarkably good efficiency.

Macroeconomic series can be regarded as aggregates across many micro-units, and explanations of how long memory behaviour might arise in macroeconomics has focussed on random-parameter short-memory modelling of micro-series. Consider the random-parameter autoregressive model of order 1 ($AR(1)$),

$$X_t(\omega) = A(\omega)X_{t-1}(\omega) + \varepsilon_t(\omega), \tag{1.9}$$

where ω indexes micro-units, the $\varepsilon_t(\omega)$ are independent and homoscedastic with zero mean across ω and t, and $A(\omega)$ is a random variable with support $(-1, 1)$ or $[0, 1)$. Then, conditional on ω, $X_t(\omega)$ is a stationary $AR(1)$ sequence. Robinson (1978a) showed that the "unconditional autovariance" which we again denote by $\gamma(j)$, is given by

$$\gamma(j) = Cov\left\{X_t(\omega), X_{t+j}(\omega)\right\} = \sum_{u=0}^{\infty} E\left\{A(\omega)^{j+2u}\right\}, \tag{1.10}$$

and that the "unconditional spectrum" $f(\lambda)$ (1.1) at $\lambda = 0$ is proportional to $E\left\{(1 - A(\omega))^{-2}\right\}$, and thus infinite, as in (1.2), if $A(\omega)$ has a density with a zero at 1 of order less than or equal to 1. One class with this property considered by Robinson (1978a) was the (possibly translated) Beta distribution, for which Granger (1980) explicitly derived the corresponding power law behaviour of the spectral density of cross-sectional aggregates $x_t = N^{-\frac{1}{2}} \sum_{i=1}^{N} X_t(\omega_i)$, where the ω_i are independent drawings: clearly $Cov(x_t, x_{t+u})$ is $\gamma(j)$, (1.10), due to the independence properties. Indeed, if $A(\omega)$ has a Beta $(c, 2 - 2d)$ distribution on $(0, 1)$, for $c > 0$, $0 < d < \frac{1}{2}$, $E\left\{A(\omega)^k\right\}$ decays like k^{2d-2}, so (1.10) decays like j^{2d-1}, as in (1.7). Intuitively, a sufficient density of individuals with close-to-unit-root behaviour produces the aggregate long memory.

A similar idea was earlier employed by Mandelbrot (1971) in computer generation of long memory time series, and for further developments, in relation to more general models than (1.9) see e.g. Goncalves and Gourieroux (1988), Lippi and Zaffaroni (1997).

The rest of the paper deals with various approaches to modelling long memory, for various kinds of data, and with relevant statistical inference. The following section provides background to estimation of parametric models. Semiparametric inference is discussed in Section 3. Section 4 describes some long memory stochastic volatility models. Section 5 concerns the extension of parametric and semiparametric inference to nonstationary series. In Section 6 we review regression models and cointegration.

2. PARAMETRIC MODELLING AND INFERENCE

Much interest in the possibility of long memory or anti-persistence focusses on the parameter d, which concisely describes long-run memory properties. In practice d is typically regarded as unknown, and so its estimation is of interest. Indeed, the discussion of the previous section indicates that an estimate of d is needed even to estimate the variance of the sample mean; this was pursued by Beran (1989), for example. In order to estimate d we need to consider the modelling of dependence in more detail.

The simplest possible realistic model for a covariance stationary series is a parametric one that expresses $\gamma(j)$ for all j, or $f(\lambda)$ for all λ, as a parametric function of just two parameters, d and an unknown scale factor. Perhaps the earliest such model is "fractional noise", which arises from considerations of self-similarity. A continuous time stochastic process $\{y(t); -\infty < t < \infty\}$ is self-similar with "self-similarity parameter" $H \in (0,1)$ if, for any $a > 0$, $\{y(at); -\infty < t < \infty\}$ has the same distribution as $\{a^H y(t); -\infty < t < \infty\}$. If the differences $x = y(t) - y(t-1)$, for integer t, are covariance stationary, we have

$$\gamma(j) = \frac{\gamma(0)}{2}\left\{|j+1|^{2H} - 2|j|^{2H} + |j-1|^{2H}\right\}. \tag{2.1}$$

As $j \to \infty$, (2.1) decays like j^{2H-2}, so on taking $H = d + \frac{1}{2}$ we have again the asymptotic law (1.7); $\gamma(0)$ is the unknown scale parameter in this model. The formula for the spectral density was derived by Sinai (1976); it is complicated, but satisfies (1.8).

"Fractional noise" was extensively studied by Mandelbrot and Van Ness (1968), Hipel and McLeod (1978) and others, but, perhaps because it extends less naturally to richer stationary series, and nonstationary series, and due to its unpleasant spectral form (see the later discussion of Whittle estimates) it has received less attention in recent years than another two-parameter model, the "fractional differencing" model proposed by Adenstedt (1974), in Chapter 2 of this volume,

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|1 - e^{i\lambda}\right|^{-2d}, \quad -\pi \le \lambda \le \pi. \tag{2.2}$$

For $d = 0$, (2.2) is just the spectral density of a white noise series (with variance σ^2), while for $0 < |d| < \frac{1}{2}$ both properties (1.7) and (1.8) hold, Adenstedt (1974) giving a formula for $\gamma(j)$ under (2.2), as well as other properties. Note that $d < \frac{1}{2}$ is necessary for integrability of $f(\lambda)$, that is for x_t to have finite variance, and this restriction is sometimes called the stationarity condition on d. Another mathematically important restriction is that of invertibility, $d > -\frac{1}{2}$. We shall discuss statistical inference on models such as (2.2).

Granger (1966) identified the "typical spectral shape of an economic variable" as not only having a pole or singularity at zero frequency, but then decaying monotonically. Both the "fractional differencing" and "fractional noise" models have this simple property. However, even if monotonicity holds, as it may, at least approximately, in case of deseasonalized economic series, the notion that the entire autocorrelation structure can be explained by a single parameter, d, is highly restrictive. While the value of d determines the long-run or low-frequency behaviour of $f(\lambda)$, greater flexibility in modelling short-run, high-frequency, behaviour may be desired. We referred to (2.2) as a "fractional differencing" model because it is the spectral density of x_t generated by

$$(1-L)^d x_t = e_t, \tag{2.3}$$

where $\{e_t\}$ is a sequence of uncorrelated variables with zero mean and variance σ^2, L represents the lag operator, $Lx_t = x_{t-1}$ and formally,

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} L^j. \tag{2.4}$$

With $d = 1$ (and an initial condition such as $x_0 = 0$), (2.3) would describe a random walk model. Box and Jenkins (1971) stressed the vector model

$$(1-L)^d a(L) x_t = b(L) e_t. \tag{2.5}$$

Here d is an integer, $a(L)$ and $b(L)$ are the polynomials

$$a(L) = 1 - \sum_{j=1}^{p} a_j L^j, \quad b(L) = 1 + \sum_{j=1}^{q} b_j L^j, \tag{2.6}$$

all of whose zeros are outside the unit circle, to ensure stationarity and invertibility, and $a(L)$ and $b(L)$ have no zero in common, to ensure unambiguity of the autoregressive (AR) order p and the moving average (MA) order q. Granger and Joyeux (1980) (Chapter 3 of this volume) considered instead fractional $d \in (-\frac{1}{2}, \frac{1}{2})$ in (2.5), giving a fractional autoregressive integrated moving average model of orders p, d, q (often abbreviated as $FARIMA(p,d,q)$ or $ARFIMA(p,d,q)$). It has spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left|1 - e^{i\lambda}\right|^{-2d} \left|\frac{b(e^{i\lambda})}{a(e^{i\lambda})}\right|^2, \quad -\pi \le \lambda \le \pi. \tag{2.7}$$

Much of the discussion of Granger and Joyeux (1980) concerned the simple $FARIMA(0,d,0)$ case (2.2) of Adenstedt (1974), but they also considered estimation of d, prediction, and computer generation of long memory series. Methods in the latter category include the aggregation approach of Mandelbrot (1971) and the Cholesky decomposition approach of Hipel and McLeod (1978); an elegant, more recent, one is due to Davies and Harte (1987), involving use of the fast Fourier transform.

Hosking (1981) provided further discussion of $FARIMA(p,d,q)$ processes, again much of it based on Adenstedt's (1974) model (2.2), but he also gave results for the general case (2.5), especially the $FARIMA(1,d,0)$. Further information on $FARIMA(p,d,q)$ models was given by Sowell (1992), Chung (1994), and others.

An early proposal for estimating d, or H, used the adjusted rescaled range (R/S) statistic

$$R/S = \frac{\max_{1\le j\le n} \sum_{t=1}^{j}(x_t - \bar{x}) - \min_{1\le j\le n} \sum_{t=1}^{j}(x_t - \bar{x})}{\left\{ \frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x})^2 \right\}^{\frac{1}{2}}} \tag{2.8}$$

of Hurst (1951), Mandelbrot and Wallis (1969). Asymptotic statistical behaviour of the R/S statistic was studied early on by Mandelbrot (1975), Mandelbrot and Taqqu (1979), and it was considered in an economic context by Mandelbrot (1972). However, while it behaves well with respect to long-tailed distributions, its limit distribution is nonstandard and difficult to use in statistical inference, while it has no known optimal efficiency properties with respect to any known family of distributions. The continued popularity of R/S, for example in the finance literature, may rest in part on an inadequate appreciation of rival procedures.

While long memory series do have distinctive features, there is no over-riding reason why traditional approaches to parametric estimation in time series should be abandoned in favour of rather special approaches like R/S. Indeed, if x_t is assumed Gaussian, the Gaussian maximum likelihood estimate (MLE) might be expected to have optimal asymptotic statistical properties, and unlile R/S, can be tailored to the particular parametric model assumed, be it (2.1), (2.2), (2.7), or whatever. The literature on the Gaussian MLE developed first with short memory processes in mind (see e.g. Whittle, 1951, Hannan, 1973), and it may be helpful to provide some background to this.

One important finding was that the Gaussian likelihood can be replaced by various approximations without affecting first order limit distributional behaviour. In particular, under suitable conditions, estimates maximizing such approximations, and called "Whittle estimates" are all $\sqrt{n}$-consistent and have the same limit normal distribution as the Gaussian MLE. Such approximations arise naturally in that treatment of pre-sample values is always an issue with time series models, but computational considerations are also an important factor. One particular Whittle estimate which seems usually particularly advantageous in the latter respect is the discrete-frequency form. Suppose the parametric spectral density has form $f(\lambda;\theta,\sigma^2) = (\sigma^2/2\pi)h(\lambda;\theta)$, where θ is an r-dimensional unknown parameter vector and σ^2 is a scalar as in (2.2). If σ^2 is regarded as varying freely from θ, and $\int_{-\pi}^{\pi} \log h(\lambda;\theta)d\lambda = 0$ for all admissible values of θ, then we have what might be called a "standard parameterization". For example, we have a standard parameterization in (2.2) with $\theta = d$, and is (2.7) with θ determining the a_j, $1 \leq j \leq p$ and b_j, $1 \leq j \leq q$. Define also the periodogram

$$I(\lambda) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} x_t e^{it\lambda}\right|^2 \tag{2.9}$$

and the Fourier frequencies $\lambda_j = 2\pi jn$. Denoting by θ_0 the true value of θ, then the discrete frequency Whittle estimate of θ_0 minimizes the following approximation to a constant minus the Gaussian log likelihood,

$$\sum_{j=1}^{n-1} \frac{I(\lambda_j)}{h(\lambda_j;\theta)}. \tag{2.10}$$

This estimate was stressed by Hannan (1973). From the viewpoint of the short memory models under discussion at that time, it had the advantages of using directly the form of h, which is readily written down in case of autoregressive moving average (ARMA) models, Bloomfeld's (1972) spectral model, and others; by contrast, autocovariances, partial autocovariances, AR coefficients and MA coefficients, which variously occur in other types of Whittle estimate, tend to be more complicated except in special cases, indeed for (2.7) the form of autocovariances, for example, can depend on the question of multiplicity of zeros of $a(L)$. Another advantage of (2.10) is that it makes direct use of the fast Fourier transform, which enables the periodograms $I(\lambda_j)$ to be rapidly computed even when n is very large. A third advantage is that mean-correction of x_t is dealt with simply by omission of the frequency $\lambda_0 = 0$.

Another important characteristic of Whittle estimates of θ_0, first established in case of short memory series, is that while they are only asymptotically efficient when x_t is Gaussian, their limit distribution (in case of "standard parameterizations") is unchanged by many departures from Gaussianity. Thus the same, relatively convenient, rules of statistical inference can be used without worrying too much about the question of Gaussianity. In particular, Hannan (1973) established this, for several Whittle forms in case x_t has a linear representation in homoscedastic stationary martingale differences having finite variance.

Hannan established first consistency under only ergodicity of x_t, so that long memory was actually included here. However, for his central limit result, with $\sqrt{n}$-convergence, which is crucial for developing statistical inference, his conditions excluded long memory, and clearly (2.10) appears easier to handle technically in the presence of a smooth h than of one with a singularity. Hannan's work was further developed in the short memory direction, to cover "nonstandard parameterizations" and multiple time series. One such treatment, of Robinson (1978b), has a central limit theorem that hints at how a modest degree of long memory might be covered. He reduced the problem to a central limit theorem for finitely many sample autocovariances, whose asymptotic normality had been shown by Hannan (1976) to rest crucially on square integrability of the spectral density; note that (2.2) and (2.3) are square integrable only for $d < \frac{1}{4}$. In fact for some forms of Whittle estimate, Yajima

(1985) established the central limit theorem, again with $\sqrt{n}$-rate, in case of model (2.2) with $0 < d < \frac{1}{4}$.

The major breakthrough in justifying Whittle estimation in long memory models was provided by Fox and Taqqu (1986) (Chapter 4 of this volume). Their objective function was not (2.10) but the continuous frequency form

$$\int_{-\pi}^{\pi} \frac{I(\lambda)}{h(\lambda;\theta)} d\lambda, \tag{2.11}$$

but Fox and Taqqu's insight applies to (2.10) also. The periodogram $I(\lambda)$ is an asymptotically unbiased estimate of the spectral density at continuity points and so $I(\lambda)$ can be expected to blow up as $\lambda \to 0$. However, since $h(\lambda;\theta)$ also blows up as $\lambda \to 0$ and appears in the denominator, some "compensation" can be expected. More precisely, limiting distributional behaviour depends on the "score" (the derivative in θ of (2.11)) at θ_0 being asymptotically normal; this, like (2.10), is a quadratic form in x_t, and Fox and Taqqu (1987) gave general conditions for such quadratic forms to be asymptotically normal, which then apply to Whittle estimates with long memory such that $0 < d < \frac{1}{2}$.

Fox and Taqqu (1986) assumed Gaussianity of x_t, as did Dahlhaus (1989), who also considered the actual Gaussian MLE and discrete-frequency Whittle (2.10), and established asymptotic efficiency. With reference to (2.11) Giraitis and Surgailis (1990) relaxed Gaussianity to a linear process in independent and identically distributed (iid) innovations, thus providing a partial extension of Hannan's (1973) work to long memory. Heyde and Gay (1993), Hosoya (1996) considered multivariate extensions. Overall, the bulk of this asymptotic theory has not directly concerned the discrete frequency form (2.10), and has focussed mainly on the continuous frequency form (2.11), though the former benefits from the neat form of the spectral density in case of the popular $FARIMA(p,d,q)$ class (2.7); on evaluating the integral in (2.11), we have a quadratic form involving the Fourier coefficients of $h(\lambda;\theta)^{-1}$, which are generally rather complicated for long memory models. Also, in (2.11) and the Gaussian MLE, correction for an unknown mean must be explicitly carried out, not dealt with merely by dropping zero frequency; Monte Carlo results of Cheung and Diebold (1994) compare this approach with sample mean-correction in Gaussian MLE.

We briefly refer to other estimates that have been considered. Whittle estimation of the models (2.1), (2.2) and (2.7) requires numerical optimization, but Kashyap and Eom (1988) proposed a closed-form estimate of d in (2.2) by a log periodogram regression (across λ_j, $j = 1, ..., n-1$). This idea does not extend nicely to $FARIMA(p, d, q)$ models with $p > 0$ or $q > 0$, but it does to

$$f(\lambda) = \frac{1}{2\pi} \left|1 - e^{i\lambda}\right|^{-2d} \exp\left\{\sum_{k=1}^{p-1} \beta_k \cos((k-1)\lambda)\right\}, \quad -\pi \leq \lambda \leq \pi \tag{2.12}$$

(see Robinson (1994a)) which combines (2.2) with Bloomfeld's (1972) short memory exponential model; Moulines and Soulier (1998) have recently provided asymptotic theory for log peridogram regression estimation of (2.12). They assumed Gaussianity, and ironically, for technical reasons, this is harder to avoid when a nonlinear function of the periodogram, such as the log, is involved, than in Whittle estimation, which was originally motivated by Gaussianity. Whittle estimation is also feasible with (2.12), indeed Robinson (1994a) noted that it can be reparameterized as

$$f(\lambda) = \frac{\exp}{2\pi} \left\{\sum_{k=1}^{p-1} \theta_k \cos\{(k-1)\lambda\} - 2d \sum_{k=p-1}^{\infty} \frac{\cos(k\lambda)}{k}\right\}, \tag{2.13}$$

taking $\theta_1 = \beta_1$, $\theta_k = \beta_k - 2/(k-1)$, $2 \leq k \leq p-1$, from which it can be deduced that the limiting covariance matrix of Whittle estimates is desirably diagonal.

Generalized method of moments (GMM) has become extremely popular in econometrics, and its use has been proposed in estimating long memory models, either in the time domain or the frequency domain. However, GMM objective functions seem in general to be less computationally attractive than (2.10), require stronger regularity conditions in asymptotic theory, and do not deal so nicely with an unknown mean.

Also, unless a suitable weighting is employed they will be less efficient than Whittle estimates in the Gaussian case, have a relatively cumbersome limiting covariance matrix, and are not even asymptotically normal under $d > \frac{1}{4}$. It must be acknowledged, however, that finite sample properties are very important in practice, and while there seems no intuitive reason why GMM estimates, for example, should be superior in this respect, it also cannot be asserted that any particular form of Whittle, such as (2.10), is always better in finite samples, or that some other, non-Whittle, estimate might not be preferable in some cases. Indeed, $\sqrt{n}$-consistency and asymptotic normality of Whittle estimates cannot even be taken for granted, having been shown not to hold over some or all of the range $d \in (0, \frac{1}{2})$ for certain nonlinear functions x_t of a underlying Gaussian long memory process (see Fox and Taqqu (1985), Giraitis and Taqqu (1999)).

Nonstandard limit distributional behaviour for Whittle estimates can also arise even under Gaussianity, in certain models. As observed in Section 1, a spectral pole (or zero) could arise at a non-zero frequency, to explain a form of cyclic behaviour. In particular Gray, Zhang and Woodward (1989) proposed the "Gegenbauer" model

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left|1 - 2e^{i\lambda}\cos\omega + e^{2i\lambda}\right|^{-2d} \left|\frac{b(e^{i\lambda})}{a(e^{i\lambda})}\right|^2, \quad -\pi \leq \lambda \leq \pi, \tag{2.14}$$

for $\omega \in (0, \pi]$. To compare with (2.2), $f(\lambda)$ has a pole at frequency ω if $d > 0$. If ω is known then our previous discussion of estimation and asymptotic theory continues to apply, see Hosoya (1996, 1997). If ω is unknown, then Whittle procedures can be adapted, but it seems that such estimates of ω (but not of the other parameters) will be n-consistent with a nonstandard limit distribution. This was claimed by Chung (1996a,b) albeit with inadequate proof, while Giraitis, Hidalgo and Robinson (2001) established n-consistency for an estimate of ω that even lacks a limit distribution.

3. SEMIPARAMETRIC MODELLING AND INFERENCE

In the use of parametric $FARIMA(p, d, q)$ models, correct specification of p and q is important. Under-specification of p or q leads to inconsistent estimation of AR and MA coefficients, but also of d, as does over-specification of both, due to a loss of identifiability. Order-determination procedures developed for short memory models, such as AIC, can be adapted to FARIMA models (indeed see Beran, Bhansali and Ocker (1998) for the $FARIMA(p, d, 0)$ case) but there is no guarantee that the underlying model belongs to the finite-parameter class proposed. It seems especially unfortunate that an attempt to seriously model short-run features can lead to inconsistent estimation of long-run properties, if the latter happen to be the aspect of most interest.

The asymptotic behaviour (1.7) and (1.8) indicates that short-run modelling is almost irrelevant at very low frequencies and very long lags, where d dominates. It thus appears that estimates of d can be based on information arising from only one or other of these domains, and that such estimates should have validity across a wide range of short memory behaviour. As this robustness requires estimates to essentially be based on only a vanishingly small fraction of the data as sample size increases, one expects slower rates of convergence than for estimates based on a correct finite-parameter model. However, in very long series, such as arise in finance, the degrees of freedom available may be sufficient to provide adequate precision. These estimates are usually referred to as "semiparametric", though their slow convergence rates make them more akin to "nonparametric" estimates in other areas of statistics, indeed some are closely related to the smoothed nonparametric spectrum estimates familiar from short memory time series analysis.

Before presenting some individual semiparametric estimates, it is worth stressing that not just point estimation is of interest, but also interval estimation and hypothesis testing. Perhaps the test of most interest to practitioners is a test of long memory, or rather, a test of short memory $d = 0$ against long memory alternatives $d > 0$, or anti-persistent alternatives $d < 0$, or both, $d \neq 0$. What is then needed is a statistic with a distribution that can be satisfactorily approximated, and computed, under $d = 0$, and that has good power. In a parametric context, tests of $d = 0$ - perhaps of Wald, Lagrange multiplier or likelihood-ratio type - can be based on Whittle functions such as (2.10) and the $FARIMA(p, d, q)$ family. (Strictly speaking, much of the limit distribution theory for Whittle estimation primarily concerned with stationary long memory, $0 < d < \frac{1}{2}$, does not cover $d = 0$, or $d < 0$, but other earlier short memory theory, such as Hannan's (1973), can provide null limit theory for testing $d = 0$.) The test statistic is based on assumed p and q, but the null limit distribution considered on this basis is generally invalid if p and q are misspecified, as discussed earlier; this can lead, for example, to mistaking unnaccounted-for short memory behaviour for long memory, and rejecting the null too often. Lo (1991) (Chapter 5 of this volume) observed the invalidity of tests for $d = 0$ based on asymptotic theory of Mandelbrot (1975) for the R/S statistic (2.8) in the presence of unanticipated short memory autocorrelation. He proposed a corrected statistic (using smoothed nonparametric spectral estimation at frequency zero) and developed its limit distribution under $d = 0$ in the presence of a wide range of short memory dependence (described by mixing conditions), and tested stock returns for long memory.

Lo's paper is perhaps especially notable as an early, rigorous treatment of asymptotic theory in a semiparametric context. However, the null limit theory for his modified R/S statistic is nonstandard. In principle, any number of statistics has sensitivity to long memory. Some have the character of "method-of-moments" estimates, minimizing a "distance" between population and sample properties. In the frequency domain, Robinson (1994b) proposed an "averaged periodogram" estimate of d, employing what would be a consistent estimate of $f(0)$ under $d = 0$. He established consistency, requiring finiteness of only second moments and allowing for the presence of an unknown slowly varying factor $L(\lambda)$ in $f(\lambda)$, so that (1.8) is relaxed to

$$f(\lambda) \sim L(\lambda) |\lambda|^{-2d}, \quad \text{as } \lambda \to 0. \tag{3.1}$$

Delgado and Robinson (1996) proposed data-dependent choices of the bandwidth number (analogous to the one discussed later in relation to log periodogram estimation, for example) that is required in the estimation, and Lobato and Robinson (1996) established limit distribution theory, which is complicated: the estimate is asymptotically normal for $0 \leq d < \frac{1}{4}$, but non-normal for $d \geq \frac{1}{4}$. Lobato (1997) extended Robinson's (1994b) consistency result to the averaged cross-periodogram of bivariate series. A number of other semiparametric estimates of d share this latter property, which is due to $f(\lambda)$ not being square-integrable for $d \geq \frac{1}{4}$, for example the time-domain "variance-type" estimate of Teverovsky and Taqqu (1997).

We might then turn to the traditional statistical practice of regression. In the time domain, the asymptotic rule (1.7) suggests two approaches, nonlinearly regressing sample autocovariances on cj^{2d-1}, and ordinary linear regression (OLS) of logged sample autocovariances on $\log j$ and an intercept. These were proposed by Robinson (1994a), the second proposal then being studied by Hall, Koul and Turlach (1997). However, the limit distributional properties of these estimates are as complicated as those for the averaged periodogram estimate, intuitively because OLS is a very *ad hoc* procedure in this setting, the implied "disturbances" in the "regression model" being far from uncorrelated or homoscedastic.

Nice results can only be expected from OLS if the disturbances are suitably "whitened". At least for short memory series, the (Toeplitz) covariance matrix of $x_1, ..., x_n$ is approximately diagonalized by a unitary transformation, such that normalized periodograms $u_j = \log \{I(\lambda_j)/f(\lambda_j)\}$ (cf (2.10)), sufficiently resemble a zero-mean, uncorrelated, homoscedastic sequence. For long memory series, (1.8) suggests consideration of

$$\log I(\lambda_j) \simeq \log c - 2d \log \lambda_j + u_j, \tag{3.2}$$

for a positive constant c and λ_j close to zero. This idea was pursued by Geweke and Porter-Hudak (1983) (Chapter 6 of this volume) though they instead employed a narrow band version of the "fractional differencing" model (2.2), specifically replacing $\log \lambda_j$ by $\log \left|1 - e^{i\lambda_j}\right|$. They performed OLS regression over $j = 1, ..., m$, where m, a bandwidth or smoothing number, is much less than n but is regarded as increasing slowly with n in asymptotic theory. Geweke and Porter-Hudak's approach was anticipated by a remark of Granger and Joyeux (1980).

Geweke and Porter-Hudak argued, in effect, that as $n \to \infty$ their estimate $\tilde{d}$ satisfies

$$m^{\frac{1}{2}}(\tilde{d} - d) \to_d N\left(0, \frac{\pi^2}{24}\right), \tag{3.3}$$

giving rise to extremely simple inferential procedures. However the heuristics underlying their argument are slightly defective, and they, and some subsequent authors, did not come close to providing a rigorous proof of (3.3). A difficulty with their heuristics is that for long memory (and anti-persistent) series the u_j are not actually asymptotically uncorrelated or homoscedastic for fixed j with $n \to \infty$, as shown by Kűnsch (1986), and elaborated upon by Hurvich and Beltrao (1993), Robinson (1995a). As Robinson (1995a) showed, this in itself invalidates Geweke and Porter-Hudak's (1983) argument. Even for increasing j, the approximation of the u_j by an uncorrelated, homoscedastic sequence is not very good, and this, and the nonlinearly-involved periodogram, makes a proof of (3.3) non-trivial.

Robinson (1995a) established (3.3), explicitly in case of the approximation (3.2) rather than Geweke and Porter-Hudak's version, though indicating that the same result holds there. Robinson's result applies to the range $|d| < \frac{1}{2}$, providing simple interval estimates as well as a simple test of short memory, $d = 0$. His treatment actually covered multiple time series, possibly involving differing memory parameters, and tests for equality of these, and more efficient estimates using the equality were also given. Robinson's treatment assumed Gaussianity, but Velasco (2000) gave an extension to linear processes x_t. Both authors employed Künsch's (1986) suggestion of trimming out the lowest λ_j to avoid the anomalous behaviour of periodograms there, but Hurvich, Deo and Brodsky (1998) showed that this was unnecessary for (3.3) to hold, under suitable conditions. They also addressed the important issue of choice of the bandwidth, m, providing optimal asymptotic minimum mean-squared error theory. When $f(\lambda)\lambda^{2d}$ is twice differentiable at $\lambda = 0$, the optimal bandwidth is of order $n^{4/5}$, but the multiplying constant depends on unknown population quantities. Hurvich and Deo (1999) proposed a consistent estimate of this constant, and hence a feasible, data-dependent choice of m. Previously, Hurvich and Beltrao (1994) had related mean squared error to integrated mean squared error in spectral density estimation, and thence proposed cross-validation procedures for choosing both m and the trimming constant.

"Log-periodogram estimation" has been greatly used empirically, deservedly so in view of its nice asymptotic properties and strong intuitive appeal. However, in view of the limited information it employs there is a concern about precision, and it is worth asking at least whether the information can be used more efficiently. Robinson (1995a) showed that indeed the asymptotic variance in (3.3) can be reduced by "pooling" adjacent periodograms, prior to logging. A proposal of Künsch (1987), however, leads to an alternative frequency-domain estimate that does even better. He suggested a narrow-band discrete-frequency Whittle estimate (cf (2.10)). Essentially this involves Whittle estimation of the "model" $f(\lambda) = C\lambda^{-2d}$ over frequencies $\lambda = \lambda_j$, $j = 1, ..., m$, where m plays a similar role as in log periodogram estimation. Then C can be eliminated by a side calculation (much as the innovation variance is eliminated in getting (2.10)), and d is estimated by $\hat{d}$ which minimizes

$$\log\left\{\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2d}I(\lambda_j)\right\}-\frac{2d}{m}\sum_{j=1}^{m}\log\lambda_j. \tag{3.4}$$

There is no closed-form solution but (3.4) is easy to handle numerically. Robinson (1995b) (Chapter 7 of this volume) established that

$$m^{\frac{1}{2}}(\hat{d}-d)\rightarrow_d N(0,\frac{1}{4}). \tag{3.5}$$

Using the same m sequence, $\hat{d}$ is then more efficient than the log periodogram estimate $\tilde{d}$ (cf (3.3)), while the pooled log periodogram estimate of Robinson (1995a) has asymptotic variance that converges to $\frac{1}{4}$ from above as the degree of pooling increases. The estimate $\hat{d}$ is only implicitly defined, but it is nevertheless easy to locate, and the linear involvement of the periodogram in (3.4) makes it possible to establish (3.5) under simpler and milder conditions than needed for (3.3), Robinson employing a linear process for x_t in martingale difference innovations. This, and the coverage of all $d\in(-\frac{1}{2},\frac{1}{2})$, may have implications also for further development of the asymptotic theory of parametric Whittle estimates discussed in the previous section. Another feature of the asymptotic theory of Robinson (1995a), and that of Robinson (1995b), is the purely local nature of the assumptions on $f(\lambda)$; and the way in which the theory fits in with earlier work on smoothed nonparametric spectral estimation for short memory series; (1.8) is relaxed to

$$f(\lambda)=C\left|\lambda\right|^{-2d}\left(1+O\left|\lambda\right|^{\beta}\right),\quad \text{as } \lambda\rightarrow 0, \tag{3.6}$$

where $\beta\in(0,2]$ is analogous to the local smoothness parameter involved in the spectral estimation work, and no smoothness, or even boundedness, is imposed on f away from zero frequency. The parameter β also enters into rules for optimal choice of m; see Henry and Robinson (1996). A nice feature of the "Gaussian semiparametric" or "local Whittle" estimate $\hat{d}$ is that it extends naturally to multivariate series; see

Lobato (1999). If only a test for short memory is desired, Lobato and Robinson (1998) provided a Lagrange multiplier one based on (3.4) that avoids estimation of d.

Work has proceeded on refinements to the semiparametric estimates $\tilde{d}$ and $\hat{d}$, and their asymptotic theory. Monte Carlo simulations have indicated bias in $\tilde{d}$, and Hurvich and Beltrao (1994), Hurvich and Deo (1999) have proposed bias-reduced estimates. Andrews and Guggenberger (2000), Robinson and Henry (2000) have developed estimates that can further reduce the bias, and have smaller asymptotic minimum mean squared error, using respectively an extended regression and higher-order kernels, Robinson and Henry (2000) at the same time introducing a unified M-estimate class that includes $\tilde{d}$ and $\hat{d}$ as special cases. Bias reduction, and a rule for bandwidth choice, also results from Giraitis and Robinson's (2000) development of an Edgeworth expansion for a modified version of $\hat{d}$. Moulines and Soulier (1999, 2000) and Hurvich and Brodsky (2001) considered a broad-band version of $\tilde{d}$ originally proposed by Janacek (1982), effectively extending the regression in (3.2) over all Fourier frequencies after including cosinusoidal terms, corresponding to the model (2.12) with p, now a bandwidth number, increasing slowly with n. They showed that if $f(\lambda)\lambda^{2d}$ is analytic over all frequencies, an asymptotic mean squared error of order $\log n/n$ can thereby be obtained, which is not achievable by the refinements to $\tilde{d}$ and $\hat{d}$ we have discussed, though the latter require only local-to-zero assumptions on $f(\lambda)$. An alternative semiparametric estimate with nice properties was proposed by Parzen (1986) and studied by Hidalgo (2001).

4. STOCHASTIC VOLATILITY MODELS

We have so far presented "long memory" as purely a second-order property of a time series, referring to autocovariances or spectral structure. These do not completely describe non-Gaussian processes, where "memory" might usefully take on a rather different meaning. In particular, passing a process through a nonlinear filter can change asymptotic autocovariance structure. As Rosenblatt (1961) indicated, if x_t is a stationary long memory Gaussian process satisfying (1.7), then x_t^2 has autocovariance decaying like j^{4d-2}, so has "long memory" only when $\frac{1}{4} \leq d < \frac{1}{2}$, and even here, since $4d - 2 < 2d - 1$, x_t^2 has "less memory" than x_t.

Many financial time series suggest a reverse kind of behaviour. Asset returns, or logged asset returns, frequently exhibit little autocorrelation, as is consistent with the efficient markets hypothesis, whereas their squares are noticeably correlated. Engle (1982) proposed to model this phenomenon by the autoregressive conditionally heteroscedastic model of order p ($ARCH(p)$), such that

$$x_t = \varepsilon_t \sigma_t, \tag{4.1}$$

where σ_t is the square root of

$$\sigma_t^2 \overset{def}{=} E\left(x_t^2 \,|x_{t-1}, x_{t-2}, ...\right) = \alpha_0 + \sum_{j=1}^{p} \alpha_j x_{t-j}^2, \tag{4.2}$$

where $\alpha_0 > 0$, $\alpha_j \geq 0$, $1 \leq j \leq p$, and ε_t is a sequence of iid random variables (possibly Gaussian). Then, under suitable conditions on the α_j, it follows that the x_t are martingale differences (and thus uncorrelated), whereas the x_t^2 have an $AR(p)$ representation, in terms of martingale difference (but not conditionally heteroscedastic) innovations. Engle's model was extended by Bollerslev (1986) to the generalized autoregressive conditionally heteroscedastic model of index p, q ($GARCH(p,q)$) which implies that the x_t^2 have an $ARMA(\max(p,q), q)$ representation in a similar sense.

Both ARCH and GARCH models have found considerable use in finance. However, they imply that the autocorrelations of the squares x_t^2 either eventually cut off completely or decay exponentially, whereas empirical evidence of slower decay perhaps consistent with long memory, has accumulated, see e.g. Whistler (1990), Ding, Granger and Engle (1993). In fact Robinson (1991) (Chapter 8 of this volume) had already suggested ARCH-type models capable of explaining greater autocorrelation in squares, so that (4.2) is extended to

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{\infty} \alpha_j x_{t-j}^2, \tag{4.3}$$

or replaced by

$$\sigma_t^2 = \left(\alpha_0 + \sum_{j=1}^{\infty} \alpha_j x_{t-j} \right)^2. \tag{4.4}$$

For both models, and related situations, Robinson (1991) developed Lagrange multiplier or score tests of "no-ARCH" (which is consistent with $\alpha_j = 0,\ j \geq 1$) against general parameterizations in (4.3) and (4.4); such tests should be better at detecting autocorrelation in x_t^2 that falls off more slowly than ones based on the $ARCH(p)$, (4.2), say.

So far as (4.3) is concerned, we can formally rewrite it as

$$x_t^2 - \sum_{j=1}^{\infty} \alpha_j x_{t-j}^2 = \alpha_0 + \nu_t, \tag{4.5}$$

where the $\nu_t = x_t^2 - \sigma_t^2$ are martingale differences. In Section 5 of Robinson (1991) the possibility of using for α_j in (4.5) the AR weights from the $FARIMA(0,d,0)$ model (see (2.2)) was considered, taking $\alpha_0 = 0$, and Whistler (1990) applied this version of his test to test $d = 0$ in exchange rate series. The $FARIMA(0,d,0)$ case of (4.1) was further considered by Ding and Granger (1996), along with other possibilities, but sufficient conditions of Giraitis, Kokoszka and Leipus (2000) for existence of a covariance stationary solution of (4.5) rule out long memory, though they do permit strong autocorrelation in x_t^2 that very closely approaches it, and Giraitis and Robinson (2001) have established asymptotic properties of Whittle estimates based on squares for this model. For $FARIMA(p,d,q)$ AR weights α_j in (4.5), x_t^2 is not covariance stationary when $d > 0$, $\alpha_0 > 0$, and Baillie, Bollerslev and Mikkelsen (1996) called this FIGARCH, a model that has since been widely applied in finance. Gaussian ML has been used to estimate it, but there seems at the time of writing to be no rigorous asymptotic theory available for this, though work is proceeding in this direction. Indeed, until recently rigorous asymptotic theory had only been given for this approach to estimating $GARCH(1,1)$ and $ARCH(p)$ models within the $GARCH(p,q)$ class.

So far as model (4.4) is concerned, Giraitis, Robinson and Surgailis (2000) have shown that if the weights α_j decay like j^{d-1}, $0 < d < \frac{1}{2}$, then any integral power x_t^k, such as the square, has long memory autocorrelation, satisfying (1.7) irrespective of k. This model also has the advantage over (4.3) of avoiding the non-negativity constraints on the α_j, and an ability to explain leverage, but at present lacks an asymptotic theory for parametric estimation.

Another approach to modelling autocorrelation in squares, and other nonlinear functions, alongside possible lack of autocorrelation in x_t, expresses σ_t^2 in (4.2) directly in terms of past ε_t, rather than past x_t, leading to a nonlinear MA form. Nelson (1991) proposed the exponential GARCH (EGARCH) model, where in (4.2) we take

$$\ln \sigma_t^2 = \alpha_0 + \sum_{j=1}^{\infty} \alpha_j g(\varepsilon_{t-j}), \tag{4.6}$$

where g is a user-chosen nonlinear function, for example Nelson stressed $g(z) = \theta z + \gamma(|z| - E\,|z|)$, which is useful in describing a leverage effect. Nelson pointed out the potential for choosing the α_j to imply long memory in σ_t^2, but stressed short memory, ARMA, weights α_j. Robinson and Zaffaroni (1997) proposed nonlinear MA models, such as

$$x_t = \varepsilon_t \left(\alpha_0 + \sum_{j=1}^{\infty} \alpha_j \varepsilon_{t-j} \right), \tag{4.7}$$

where the ε_t are an iid sequence. They showed the ability to choose the α_j such that x_t^2 has long memory autocorrelation, and proposed use of Whittle estimation based on the x_t^2.

A closely related model to (4.7), proposed by Robinson and Zaffaroni (1998), replaces the first factor ε_t by η_t, where the η_t are iid and independent of the ε_t, again long memory potential was shown. This model is a special case of

$$x_t = \eta_t h(\varepsilon_{t-1}, \varepsilon_{t-2}, ...), \tag{4.8}$$

of which the short memory stochastic volatility model of Taylor (1986) is also a special case. Long memory versions of Taylor's model were studied by Breidt, Crato and de Lima (1998) (Chapter 9 of this volume), Harvey (1998), choosing

$$h(\varepsilon_{t-1}, \varepsilon_{t-2}, ...) = \exp\left(\alpha_0 + \sum_{j=1}^{\infty} \alpha_j \varepsilon_{t-j}\right), \tag{4.9}$$

where the α_j are MA weights in the $FARIMA(p, d, q)$. They considered Whittle estimation based on squares, Breidt et al. discussing its consistency, and applying the model to stock price data. Note that asymptotic theory for ML estimates of models such as (4.7), (4.8) and (4.9) is considerably more difficult to derive, indeed it is hard to write down the likelihood, given, say, Gaussian assumptions on ε_t and η_t. To ease mathematical tractability in view of the nonlinearity in (4.9), Gaussianity of ε_t was indeed stressed by Breidt et al. and Harvey. In that case, we can write the exponent of h in (4.9) as $\alpha_0 + z_t$, where z_t is a stationary Gaussian, possibly long memory, process, and likewise the second factor in (4.7). These models are all covered by modelling x_t as a general nonlinear function of a vector unobservable Gaussian process ξ_t. From an asymptotic expansion for the covariance of functions of multivariate normal vectors, Robinson (2001) indicated how long memory in nonlinear functions of x_t depends on the long memory in ξ_t and the nature of the nonlinearity involved, with application also to cyclic behaviour, cross-sectional and temporal aggregation, and multivariate models; see also Andersen and Bollerslev (1997). The allowance for quite general nonlinearity means that relatively little generality is lost by the Gaussianity assumption on ξ_t, while the scope for studying autocorrelation structure of functions such as $|x_t|$ can avoid the assumption of a finite fourth moment in x_t, which has been controversial.

5. NONSTATIONARY LONG MEMORY

Unit root models have been a major focus of econometrics during the past 15 years or so. Prior to this, modelling of economic time series typically involved a combination of short memory, $I(0)$, series satisfying (1.4), and ones that are nonstochastic, either in the sense of sequences such as dummy variables or polynomial time trends, or of conditioning on predetermined economic variables. Unit root modelling starts from the random walk model, i.e. (2.3) for $t \geq 1$ with $d = 1, e_t$ white noise and $x_0 = 0$, and then generalizes e_t to be a more general $I(0)$ process, modelled either parametrically or nonparametrically; x_t is then said to be an $I(1)$ process. Unit root models, often with the involvement also of nonstationary time trends, have been successfully used in macroeconometrics, frequently in connection with cointegration analysis.

A key preliminary step is the testing of the unit root hypothesis. Many such tests have been proposed, often directed against $I(0)$ alternatives, and using classical Wald, Lagrange multiple and likelihood-ratio procedures, see e.g. Dickey and Fuller (1979, 1981), Phillips (1987). In many other situations, these lead to a null χ^2 limit distribution, a non-central local χ^2 limit distribution, Pitman efficiency, and a considerable degree of scope for robustness to the precise implementation of the test statistics, for example to the estimate of the asymptotic variance matrix that is employed. The unit root tests against $I(0)$ alternatives lose such properties, for example the null limit distribution is nonstandard.

The nonstandard behaviour arises essentially because the unit root is nested unsmoothly in an AR system: in the $AR(1)$ case, the process is stationary with exponentially decaying autocovariance structure when the AR coefficient α lies between -1 and 1, has unit root nonstationarity at $\alpha = 1$, and is "explosive" for $|\alpha| > 1$. Moreover, the tests directed against AR alternatives seem not to have very good powers against fractional alternatives, as Monte Carlo investigation of Diebold and

Rudebusch (1991) suggests.

There is any number of models that can nest a unit root, and the fractional class turns out to have the "smooth" properties that lead classically to the standard, optimal asymptotic behaviour referred to earlier. Robinson (1994c) (Chapter 10 of this volume) considered the model

$$\varphi(L)x_t = u_t, \quad t \geq 1, \tag{5.1}$$

$$x_t = 0, \quad t \leq 0, \tag{5.2}$$

where u_t is an $I(0)$ process with parametric autocorrelation and

$$\varphi(L) = (1-L)^{d_1}(1+L)^{d_2}\prod_{j=3}^{h}(1-2\cos\omega_j L + L^2)^{d_j}, \tag{5.3}$$

in which the ω_j are given distinct real numbers in $(0,\pi)$, and the d_j, $1 \leq j \leq h$, are arbitrary real numbers. The initial condition (5.2) avoids an unbounded variance, the main interest being in nonstationary x_t. Robinson proposed tests for specified values of the d_j against, fractional, alternatives in the class (5.3). Thus, for example, in the simplest case the unit root hypothesis $d = 1$ can be tested, but against fractional alternatives $(1-L)^d$ for $d > 1$, $d < 1$ or $d \neq 1$. Other null d may be of interest, e.g. $d = \frac{1}{2}$, this being the boundary between stationarity and nonstationarity in the fractional domain. The region $d \in [\frac{1}{2}, 1)$ is referred to as mean-reverting, MA coefficients of x_t decaying, albeit more slowly than under stationary, $d < \frac{1}{2}$. The models (5.1)-(5.3) also cover seasonal and cyclical components (cf the Gegenbauer model (2.14)) as well as stationary and overdifferenced ones. Robinson showed that his Lagrange multiplier tests enjoy the classical large-sample properties of such tests, described above.

An intuitive explanation of this outcome is that, unlike in unit root tests against AR alternatives, the test statistics are based on the null differenced x_t, which are $I(0)$ under the null hypothesis. This suggests that estimates of memory parameters

d_j in (5.1)-(5.3) and of parameters describing u_t, such as Whittle estimates, will also continue to possess the kind of standard asymptotic properties - $\sqrt{n}$-consistency and asymptotic normality - under nonstationarity as we encountered in the stationary circumstances of Section 2. Indeed, Beran (1995), in case $\varphi(L) = (1-L)^d$ and u_t white noise, indicated this, though the initial consistency proof he provides, an essential preliminary to asymptotic distribution theory for his implicitly-defined estimate, appears to assume that the estimate lies in a neighbourhood of the true d, itself a consequence of consistency; over a suitably wide range of d-values, the objective function does not converge uniformly. Velasco and Robinson (2000) adopted a somewhat different approach, employing instead the model

$$
\begin{aligned}
(1-L)^s x_t &= v_t, \quad t \geq 1, && (5.4)\\
x_t &= 0, \quad t \leq 0, && (5.5)\\
(1-L)^{d-s} v_t &= u_t, \quad t = 0, \pm 1, ..., && (5.6)
\end{aligned}
$$

where s is the integer part of $d + \frac{1}{2}$ and u_t is a parametric $I(0)$ process such as white noise; note that v_t is a stationary $I(d-s)$ process, invertible also unless $d = -\frac{1}{2}$. The difference between the two definitions of nonstationary $I(d)$ processes, in (5.1) and (5.2) on the one hand and (5.4)-(5.6) on the other, was discussed by Marinucci and Robinson (1999); this entails, for example, convergence to different forms of fractional Brownian motion. Velasco and Robinson considered a version of discrete-frequency Whittle estimation (cf. (2.10)) but nonstationarity tends to bias periodogram ordinates, and to sufficiently reduce this they in general (for $d \geq \frac{3}{4}$ and with (5.6) modified so that v_t has an unknown mean) found it necessary to suitably "taper" the data (see e.g. Brillinger, 1975, Chapter 3) and then, in order to overcome the undesirable dependence this produces between neighbouring periodograms, to use only Fourier frequencies λ_j, such that j is a multiple of p: p is the "order" of the taper, such that $p \geq [d + \frac{1}{2}] + 1$ is required for asymptotic normality of the estimates, with

$\sqrt{n}$ rate of convergence; since d is unknown a large p can be chosen for safety's sake, but the asymptotic variance is inflated by a factor varying directly with p. The theory is invariant to an additive polynomial trend of degree up to p.

The use of tapering in a nonstationary fractional setting, and a similar model to (5.4)-(5.6), originated in Hurvich and Ray (1995) (Chapter 11 of this volume). In (5.4), they required $d = 1$, while in (5.6) they allowed $-\infty < d < \frac{1}{2}$, so that (unlike Beran (1995) and Velasco and Robinson (2000)) they covered nonstationarity only up to $d < 3/2$ (though this probably fits many applications) and on the other hand covered any degree of noninvertibility. Their concern, however, was not with asymptotic theory for parameter estimates. Hurvich and Ray found that asymptotic bias in $I(\lambda_j)$, for fixed j as $n \to \infty$, could be notably reduced by use of a cosine bell taper, leading them to recommend use of tapering (and omission of frequency λ_1) in the log periodogram estimation of d discussed in Section 3, in case nonstationarity is feared. Velasco (1999a) then established limit distribution theory, analogous to that described in Section 3, for log periodogram estimates in case $d \geq \frac{1}{2}$, in a semiparametric version of (5.4)-(5.6) (so u_t has nonparametric autocorrelation), using a general class of tapers; Velasco (1999b) established analogous results for local Whittle estimates, cf (3.4). Again, there is invariance to polynomial trends, but tapering, imposed for asymptotic normality when $d \geq \frac{3}{4}$ and for consistency when $d > 1$, entails skipping frequencies and/or an efficiency loss. However, Hurvich and Chen (2000) proposed a taper, applied to first differences when $d < 3/2$, that, with no skipping, loses less efficiency. On the other hand, Phillips (e.g. 1999, and various papers with co-workers) in the context of the model (5.1), (5.2) with $\phi(L) = (1 - L)^d$, found that untapered log periodogram and local Whittle estimates are inconsistent when $d > 1$, and not asymptotically normal when $d \geq \frac{3}{4}$. The position at present is thus that, despite its drawbacks, tapering is a wise precaution when nonstationarity is believed possible, while recent work has found it also useful in theoretical refinements even under stationarity, see e.g. Giraitis and Robinson (2000).

6. INFERENCE ON REGRESSION AND COINTEGRATION MODELS

Section 1 provided some preparatory discussion of regression, insofar as we considered estimation of the mean of a stationary long memory series. Rates of convergence and efficiency were discussed, but not limit distribution theory, and here some warning is in order. For short memory series x_t, the sample mean $\bar{x} = n^{-1}\sum_{t=1}^{n} x_t$ is asymptotically normal over a wide range of dependence conditions, including not only linear processes but also various kinds of "mixing" processes, the latter covering some forms of nonlinearity. However, Rosenblatt (1961) pointed out that if $x_t = u_t^2$, where u_t is a Gaussian long memory process with differencing parameter d (e.g. the $FARMA(0, d, 0)$ (2.21)), then $\bar{x}$ is not asymptotically normal when $d > \frac{1}{4}$. Taqqu (1975) described the limit distribution here as the "Rosenblatt distribution", and considerably developed Rosenblatt's work, modelling x_t as a quite general function of u_t. The limit distribution of $\bar{x}$ is then governed not only by d but by the lowest-degree non-vanishing term in the Hermite expansion of x_t (unlike for short memory u_t, when all terms effectively contribute). In particular, when a linear term is present, there is asymptotic normality.

Taqqu's approach has been employed and extended in connection with many statistics, but while the possible consequences of non-normality must thus be borne in mind it is not clear how to cope with this in designing useful rules of inference, and we instead focus on linear, but not necessarily Gaussian, processes, where central limit theory can be anticipated (and under conditions which are in some respects milder). Early contributions here are due to Eicker (1967) (Chapter 12 of this volume) and Ibragimov and Linnik (1971, pp.358-60).

Consider the model

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \alpha_j \varepsilon_{t-j}, \qquad \sum_{j=-\infty}^{\infty} \alpha_j^2 < \infty, \tag{6.1}$$

where the ε_t are iid with zero mean and finite variance. The square summability of the α_j is consistent with the long memory properties (1.7) and (1.8), for $d < \frac{1}{2}$. Then if also

$$\sum_{j=-\infty}^{\infty} (\alpha_{j-1} + ... + \alpha_{j-n})^2 \to \infty, \quad \text{as } n \to \infty, \tag{6.2}$$

we have

$$\{V(\bar{x})\}^{-\frac{1}{2}} (\bar{x} - \mu) \to_d N(0,1). \tag{6.3}$$

Condition (6.2) is merely equivalent to the essential divergence of the norming factor $V(\bar{x})$ in (6.3), and does not apply to "noninvertible" processes; under (1.7) $V(\bar{x})$ increases like n^{2d-1} for $|d| < \frac{1}{2}$. Both Eicker and Ibragimov and Linnik (1971) covered (6.1) but Eicker also treated multiple linear regression with quite general nonstochastic errors. Thus, for the model

$$y_t = \beta' z_t + x_t, \tag{6.4}$$

with y_t, and the $r \times 1$ nonstochastic vector z_t observed, the unobserved x_t again satisfies (6.1), (6.2) with $\mu = 0$. Eicker extended (6.3) to the ordinary least squares estimate (OLS) $\tilde{\beta}$ of the $r \times 1$ vector β. Under stronger conditions, Eicker established "autocorrelation-consistent" variance estimates of slope elements of β, of a different, convolution, form from the kernel spectral density type that has later become popular in econometrics. As a historical note, Section 3 of Eicker's paper discussed independent, heteroscedastic x_t in (6.4) and the "heteroskedasticity-consistent" variance estimates also later popularized in econometrics (though those earlier appeared in Eicker (1963)).

Eicker's conditions were later relaxed by Hannan (1979), but neither of them, nor Ibragimov and Linnik, explicitly discussed the impact of the asymptotic behaviour

(1.7) or (1.8) for x_t, let alone parametric models like (2.2). Under a model for x_t like (1.8), with $0 < d < \frac{1}{2}$, Yajima (1988, 1991) discussed both limit distribution theory and (extending work of Grenander 1954, Grenander and Rosenblatt, 1957, for short memory x_t) efficiency of OLS in (6.4), stressing particular regressors such as polynomial time trends and indicating how the asymptotic variance simplifies as a function of d. These can be implemented given a better-than-$\log n$-consistent (see Robinson, 1994b) estimate of d, so that semiparametric modelling of x_t (cf (1.7) or (1.8)) suffices. With a rather different treatment of nonstationary regressors, Dahlhaus (1995) presented estimates that achieve the same asymptotic efficiency as the generalized least squares (GLS) estimate; they involve d, and Dahlhaus then showed that the same efficiency can be achieved when d is estimated; d can take "noninvertible" values, but x_t is assumed Gaussian. Deo (1997) provided further developments along these lines, while Deo and Hurvich (1998) discussed the case of a linear time trend and $I(d)$ x_t, with $d < 3/2$, so that the errors can be nonstationary.

Time series regression in econometrics can involve stochastic regressors, and this can significantly affect the theory. Much of the nonstochastic regression theory stresses circumstances in which the limiting "spectral distribution function" of regressors z_t in "Grenander's conditions" has discrete jumps, or at least a jump at zero frequency where the spectral pole of x_t is located (as is true of polynomial time trends). For stationary stochastic z_t, the existence of a spectral density may on the other hand be plausible, and in the event of sufficiently strong long memory in both z_t and x_t, specifically if z_t is $I(c)$ with $c+d \geq \frac{1}{2}$, OLS can have a nonstandard limit distribution with convergence rate slower than the usual $\sqrt{n}$. However, Robinson and Hidalgo (1997) (Chapter 13 of this volume) proposed weighted (in both the time and frequency domains) estimates that are consistent and asymptotically normal, at rate $\sqrt{n}$. A special case of these are infeasible GLS estimates, allowing arbitrarily strong stationary long memory in both z_t and x_t, so that GLS has the advantage over OLS not only of efficiency but also, interestingly, of faster convergence rate and asymptotic normality; intuitively, the explanation is that in the frequency domain $f(\lambda)^{-1}$ is involved in GLS, and this has a zero, not a pole, under long memory, avoiding integrability problems. Robinson and Hidalgo also showed that the same asymptotic theory can hold when a parametric model for $f(\lambda)$ is estimated, and gave an extension to nonlinear regression. For the same setup, Hidalgo and Robinson (2001) extended Hannan's (1963) idea of adapting to nonparametric autocorrelation in x_t. In a single-regressor version of this model Choy and Taniguchi (2001) discussed various estimates under different combinations of long memory in x_t and z_t.

This stationary stochastic regressor theory assumed z_t and x_t are uncorrelated, for otherwise estimates will be asymptotically biased (indeed for technical reasons independence was actually assumed). If (6.4) represents a cointegrating relation between y_t and z_t, x_t has lower integration order than z_t, but there is no reason in general to suppose it is uncorrelated with z_t. The cointegration literature has usually assumed y_t and z_t are $I(1)$ while x_t is $I(0)$ (see e.g. Engle and Granger, 1987), and here OLS is still consistent due to the asymptotic dominance of x_t by z_t. This outcome extends to more general, fractional, nonstationary z_t, and x_t that are stationary or even less-nonstationary-than-z_t. However, it does not apply to stationary z_t, even when it has more long memory than x_t, due to simultaneous equations bias. The stationary case may be of interest in financial applications, and here Robinson (1994b) showed consistency of a narrow-band least squares (NBLS) estimate, in the frequency domain, where the number of (low) Fourier frequencies used increases more slowly than n.

Robinson and Marinucci (1997) (Chapter 14 of this volume) extended this result, making Robinson's z_t a vector and giving a rate of convergence, but mainly focussed on comparisons between OLS and NBLS in nonstationary circumstances. Fractional nonstationarity of form (5.1), (5.2) in z_t was considered, while x_t can be either nonstationary or stationary, possibly with long memory. They found that for some combinations of memory parameters NBLS is asymptotically equivalent to OLS (despite losing high-frequency information), but for other combinations NBLS is superior, simultaneous equations bias not preventing consistency of OLS but possibly slowing convergence. Limit distributions are nonstandard but for the most part can be characterized by applying functional limit theory of Akonom and Gourieroux (1987), Marinucci and Robinson (2000); see also Sowell (1990), Chan and Terrin (1995). Improved versions of Robinson and Marinucci's (1997) results are in Robinson and Marinucci (2001) for scalar z_t, while Robinson and Marinucci (2000) studied interaction between stochastic and nonstochastic components in z_t. The approach in these papers indicates the dominating role of low frequencies in cointegration, no parametric modelling of dependence structure being required. It thus has the usual advantages of "semiparametric" modelling, while both OLS and NBLS are computationally simple to implement. On the other hand, in cointegration involving $I(1)$ y_t and $I(0)$ x_t, OLS has been improved upon, e.g. by Phillips (1991a,b) by estimates that have a mixed normal limit distribution, with the effect that Wald statistics for testing β have classical χ^2 asymptotics. Jeganathan (1999), Robinson and Hualde (2000) have provided extensions of these results to parametric fractional models, gaining convergence rates than for some parameter combinations are faster than those of OLS and NBLS; some early empirical study of fractional cointegration is in Cheung and Lai (1993), for example.

REFERENCES

[1] ADELMAN, I. (1965), 'Long Cycles: Fact or Artefact?', *American Economic Review*, 55: 444-63.

[2] ADENSTEDT, R.K. (1974), 'On Large-sample Estimation of the Mean of a Stationary Random Sequence', *Annals of Statistics*, 2: 1095-107.

[3] AKONOM, J., and C. GOURIEROUX (1987), 'A Functional Central Limit Theorem for Fractional Processes', preprint.

[4] ANDERSEN, T.G., and T. BOLLERSLEV (1997), 'Heterogeneous Information Arrivals and Return Volatility Dynamics: Increasing the Long Run in High Frequency Returns', *Journal of Finance*, 52: 975-1005.

[5] ANDREWS, D.W.K., and K. GUGGENBERGER (2000), 'Bias-Reduced Log-periodogram Estimator for the Long Memory Parameter', preprint.

[6] BAILLIE, R.T., T. BOLLERSLEV and H.O. MIKKELSEN (1996), 'Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity; *Journal of Econometrics*, 74: 3-30.

[7] BERAN, J. (1989), 'A Test of Location for Data with Slowly Decaying Serial Correlations', *Biometrika*, 76: 261-9.

[8] ________ (1995), 'Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short- and Long-memory ARIMA Models', *Journal of the Royal Statistical Society, Series B*, 57: 659-72.

[9] ________, R. BHANSALI, and D. OCKER (1998), 'On Unified Model Selection for Stationary and Nonstationary Short- and Long-memory Autoregressive Processes', *Biometrika*, 85: 921-34.

[10] BLOOMFIELD, P. (1972), 'An Exponential Model for the Spectrum of a Scalar Time Series', *Biometrika*, 60: 217-26.

[11] BOLLERSLEV, T. (1986), 'Generalized Autoregressive Conditional Heteroskedasticity', *Journal of Econometrics*, 31: 307-27.

[12] BOX, G.E.P., and G.M. JENKINS (1971), *Time Series Analysis, Forecasting and Control*, San Francisco: Holden-Day.

[13] BREIDT, F.J., N. CRATO, and P. de LIMA (1998), 'The Detection and Estimation of Long Memory in Stochastic Volatility', *Journal of Econometrics*, 83: 325-34.

[14] BRILLINGER, D.R. (1975), *Time Series Data Analysis and Theory*, San Francisco: Holden Day.

[15] CHAN, N.H., and N. TERRIN (1995), 'Inference for Unstable Long-memory Processes with Applications to Fractional Unit Root Autoregressions', *Annals of Statistics*, 23: 1662-83.

[16] CHEUNG, Y.W., and K.S. LAI (1993), 'A Fractional Cointegration Analysis of Purchasing Power Parity', *Journal of Business and Economic Statistics*, 11: 103-12.

[17] __________ and F.X. DIEBOLD (1994), 'On Maximum Likelihood Estimation of the Differencing Parameter of Fractionally Integrated Noise with Unknown Mean', *Journal of Econometrics*, 62: 301-16.

[18] CHOY, K., and M. TANIGUCHI (2001), 'Stochastic Regression Model with Dependent Disturbances', *Journal of Time Series Analysis*, 22: 175-96.

[19] CHUNG, C.-F. (1994), 'A Note on Calculating the Autocovariance of the Fractionally Integrated ARMA Models, *Economics Letters*, 45: 293-7.

[20] _________ (1996a), 'Estimating a Generalized Long Memory Process', *Journal of Econometrics*, 73: 237-59.

[21] _________ (1996b), 'Generalized Fractionally Integrated Autoregressive Moving Average Process', *Journal of Time Series Analysis*, 17: 111-40.

[22] DAHLHAUS, R. (1989), 'Efficient Parameter Estimation for Self-similar Processes', *Annals of Statistics*, 17: 1749-66.

[23] ________ (1995), 'Efficient Location and Regression Estimation for Long Range Dependent Regression Models', *Annals of Statistics*, 23: 1029-47.

[24] DAVIES, R.B., and D.S. HARTE (1987), 'Tests for Hurst Effect', *Biometrika*, 74: 95-101.

[25] DELGADO, M.J., and P.M. ROBINSON (1996), 'Optimal Spectral Bandwidth for Long Memory', *Statistica Sinica*, 6: 97-112.

[26] DEO, R.S. (1997), 'Asymptotic Theory for Certain Regression Models with Long Memory Errors', *Journal of Time Series Analysis*, 18: 385-93.

[27] ______, and C.M. HURVICH (1998), 'Linear Trend with Fractionally Integrated Errors', *Journal of Time Series Analysis*, 19: 379-97.

[28] DICKEY, D.A., and W.A. FULLER (1979), 'Distribution of the Estimators for Autoregressive Time Series with a Unit Root', *Journal of the American Statistical Association*, 74: 427-31.

[29] ____________________ (1981), 'Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root', *Econometrica*, 49: 1057-72.

[30] DIEBOLD, F.X., and G.D. RUDEBUSCH (1991), 'On the Power of Dickey-Fuller Tests Against Fractional Alternatives', *Economic Letters*, 35: 155-60.

[31] DING, Z., and C.W.J. GRANGER (1996), 'Modelling Volatility Persistence of Speculative Returns: A New Approach', *Journal of Econometrics*, 73: 185-215.

[32] _____, C.W.J. GRANGER, and R.F. ENGLE (1993), 'A Long Memory Property of Stock Market Returns and a New Model', *Journal of Empirical Finance*, 1: 83-106.

[33] EICKER, F. (1963), 'Asymptotic Normality and Consistency of the Least Squares Estimator for Families of Linear Regressions', *Annals of Mathematical Statistics*, 34: 447-56.

[34] ______ (1967), 'Limit Theorems for Regressions with Unequal and Dependent Errors'. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol.1, Berkeley: University of California Press, pp.59-82.

[35] ENGLE, R.F. (1982), 'Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation', *Econometrica*, 50: 987-1007.

[36] _______, and C.W.J. GRANGER (1987), 'Cointegration and Error Correction: Representation, Estimation and Testing', *Econometrica*, 55: 251-76.

[37] FAIRFIELD SMITH, H. (1938), 'An Empirical Law Describing Heterogeneity in the Yields of Agricultural Crops', *Journal of Agricultural Science*, 28: 1-23.

[38] FOX, R., and M.S. TAQQU (1986), 'Large Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series', *Annals of Statistics*, 14: 517-32.

[39] ______________ (1987), 'Central Limit Theorems for Quadratic Forms in Random Variables Having Long-range Dependence', *Probability Theory and Related Fields*, 74: 213-40.

[40] GEWEKE, J., and S. PORTER-HUDAK (1983), 'The Estimation and Application of Long-Memory Time Series Models', *Journal of Time Series Analysis*, 4: 221-38.

[41] GIRAITIS, L., F.J. HIDALGO, and P.M. ROBINSON (2001), 'Gaussian Estimation of Parametric Spectral Density with Unknown Pole', *Annals of Statistics*, 29: 987-1023.

[42] _______, P. KOKOSZKA, and R. LEIPUS (2000), 'Stationary ARCH Models: Dependence Structure and Central Limit Theorem', *Econometric Theory*, 16:3-22.

[43] _______, and P.M. ROBINSON (2000), 'Edgeworth Expansions for Semiparametric Whittle Estimation of Long Memory', preprint.

[44] _______, and P.M. ROBINSON (2001), 'Whittle Estimation of ARCH Models', *Econometric Theory*, 17: 608-31.

[45] _______, P.M. ROBINSON, and D. SURGAILIS (2000), 'A Model for Long Memory Conditional Heteroscedasticity', *Annals of Applied Probability*, 10: 1002-24.

[46] _______, and D. SURGAILIS (1990), 'A Central Limit Theorem for Quadratic Forms in Strongly Dependent Random Variables and its Application to Asymptotical Normality of Whittle's Estimate', *Probability Theory and Related Fields*, 86: 87-104.

[47] _______, and M.S. TAQQU (1999), 'Whittle Estimator for Finite - Variance Non-Gaussian Time Series with Long Memory', *Annals of Statistics*, 27: 178-203.

[48] GONCALVES, E., and C. GOURIEROUX (1988), 'Agrégation de Processus Autorégressifs D'ordre 1', *Annales d'Économie et de Statistique*, 12: 127-49.

[49] GRANGER, C.W.J. (1966), 'The Typical Spectral Shape of an Economic Variable', *Econometrica*, 34: 150-61.

[50] __________ (1980), 'Long Memory Relationships and the Aggregation of Dynamic Models', *Journal of Econometrics*, 14: 227-38.

[51] __________, and R. JOYEUX (1980), 'An Introduction to Long-memory Time Series and Fractional Differencing', *Journal of Time Series Analysis*, 1: 15-29.

[52] GRAY, H.L., N.I. ZHANG, and W.A. WOODWARD (1989), 'On Generalized Fractional Processes', *Journal of Time Series Analysis*, 10: 233-57.

[53] GRENANDER, U. (1954), 'On the Estimation of Regression Coefficients in the Case of an Autocorrelated Disturbance', *Annals of Mathematical Statistics*, 25: 252-72.

[54] ___________, and M. ROSENBLATT (1957), *Statistical Analysis of Stationary Time Series*, New York: Wiley.

[55] HALL, P., H.L. KOUL, and B.A. TURLACH (1997), 'Note on Convergence Rates of Semiparametric Estimators of Dependence Index', *Annals of Statistics*, 25: 1725-39.

[56] HANNAN, E.J. (1963), 'Regression for Time Series', in M. Rosenblatt, ed., *Time Series Analysis*, New York: Wiley, pp.17-32.

[57] ________ (1973), 'The Asymptotic Theory of Linear Time Series Models', *Journal of Applied Probability*, 10: 130-45.

[58] ________ (1970), 'The Asymptotic Distribution of Serial Covariances', *Annals of Statistics*, 4: 396-9.

[59] ________ (1979), 'The Central Limit Theorem for Time Series Regression', *Stochastic Processes and their Applications*, 9: 281-9.

[60] HARVEY, A. (1998), 'Long Memory in Stochastic Volatility', in J. Knight and S. Satchell, eds., *Forecasting Volatility in the Financial Markets*, Oxford: Butterworth-Heinemann.

[61] HENRY, M., and P.M. ROBINSON (1996), 'Bandwidth Choice in Gaussian Semiparametric Estimation of Long-range Dependence', in P.M. Robinson and M. Rosenblatt, eds., *Athens Conference on Applied Probability and Time Series Analysis*, vol.II: Time Series Analysis. In Memory of E.J. Hannan, New York: Springer-Verlag, pp.220-32.

[62] HEYDE, C., and G. GAY (1993), 'Smoothed Periodogram Asymptotics and Estimation for Processes and Fields with Long-range Dependence', *Stochastic Processes and their Applications*, 45: 169-87.

[63] HIDALGO, F.J. (2001), 'Semiparametric Estimation of the Location of the Pole', preprint.

[64] ___________, and P.M. ROBINSON (2001), 'Adapting to Unknown Disturbance Autocorrelation with Long Memory', forthcoming, *Econometrica*.

[65] HIPEL, K.W., and A.J. McLEOD (1978), 'Preservation of the Adjusted Rescaled Range Parts 1, 2 and 3', *Water Resources Research*, 14: 491-518.

[66] HOSKING, J.R.M. (1981), 'Fractional Differencing', *Biometrika*, 68: 165-76.

[67] HOSOYA, Y. (1996), 'The Quasi-Likelihood Approach to Statistical Inference on Multiple Time Series with Long-range Dependence', *Journal of Econometrics*, 73: 217-36.

[68] _______ (1997), 'Limit Theory with Long-range Dependence and Statistical Inference of Related Models', *Annals of Statistics*, 25: 105-37.

[69] HURST, H. (1951), 'Long Term Storage Capacity of Reservoirs', *Transactions of the American Society of Civil Engineers*, 116: 770-99.

[70] HURVICH, C.M., and K.I. BELTRAO (1993), 'Asymptotics for the Low-Frequency Estimates of the Periodogram of a Long Memory Time Series', *Journal of Time Series Analysis*, 14: 455-72.

[71] __________, and K.I. BELTRAO (1994), 'Automatic Semiparametric Estimation of the Memory Parameter of a Long-memory Time Series', *Journal of Time Series Analysis*, 15: 285-302.

[72] __________, and J. BRODSKY (2001), 'Broadband Semiparametric Estimation of the Memory Parameter of a Long-memory Time Series using Fractional Exponential Model', *Journal of Time Series Analysis*, 22: 221-49.

[73] __________, and W.W. CHEN (2000), 'An Efficient Taper for Potentially Overdifferenced Long-memory Time Series', *Journal of Time Series Analysis*, 21: 155-80.

[74] __________, and R.S. DEO (1999), 'Plug-in Selection of the Number of Frequencies in Regression Estimates of the Memory Parameter of a Long-memory Time Series', *Journal of Time Series Analysis*, 20: 331-41.

[75] __________, R. DEO., and J. BRODSKY (1998), 'The Mean Squared Error of Geweke and Porter-Hudak's Estimates of the Memory Parameter of a Long Memory Time Series', *Journal of Time Series Analysis*, 19: 19-46.

[76] __________, and B.K. RAY (1995), 'Estimation of the Memory Parameter for Nonstationary or Noninvertible Fractionally Integrated Processes', *Journal of Time Series Analysis*, 16: 17-41.

[77] IBRAGIMOV, I.A., and Yu. V. LINNIK (1971), *Independent and Stationary Sequences of Random Variables*, Groningen: Wolters-Noordhoff.

[78] JANACEK, G.J. (1982), 'Determining the Degree of Differencing for Time Series via Log Spectrum', *Journal of Time Series Analysis*, 3: 177-83.

[79] JEFFREYS, H. (1939), *Theory of Probability*, Oxford: Clarendon Press.

[80] JEGANATHAN, P. (1999), 'On Asymptotic Inference in Cointegrated Time Series with Fractionally Integrated Errors', *Econometric Theory*, 15: 583-621.

[81] ___________ (2001), 'Correction to "On Asymptotic Inference in Cointegrated Time Series with Fractionally Integrated Errors"', preprint.

[82] KASHYAP, R., and K. EOM (1988), 'Estimation in Long-memory Time Series Model', *Journal of Time Series Analysis*, 9: 35-41.

[83] KÜNSCH, H.R. (1986), 'Discrimination Between Monotonic Trends and Long-range Dependence', *Journal of Applied Probability*, 23: 1025-30.

[84] _________ (1987), 'Statistical Aspects of Self-similar Processes', *Proceedings of the First World Congress of the Bernoulli Society*, VNU Science Press, 1: 67-74.

[85] LIPPI, M., and P. ZAFFARONI (1998), 'Aggregation and Simple Dynamics: Exact Asymptotic Results', preprint.

[86] LO, A.W. (1991), 'Long-term Memory in Stock Market Prices', *Econometrica*, 59: 1279-313.

[87] LOBATO, I.G. (1997), 'Consistency of the Averaged Cross-periodogram in Long Memory Series', *Journal of Time Series Analysis*, 18: 137-56.

[88] _________ (1999), ‘A Semiparametric Two-step Estimator in a Multivariate Long Memory Model’, *Journal of Econometrics*, 90: 129-53.

[89] _________, and P.M. ROBINSON (1996), ‘Averaged Periodogram Estimation of Long Memory’, *Journal of Econometrics*, 73: 303-24.

[90] _________, and P.M. ROBINSON (1998), ‘A Nonparametric Test for $I(0)$’, *Review of Economic Studies*, 65: 475-95.

[91] MANDELBROT, B.B. (1971), ‘A Fast Fractional Gaussian Noise Generator’, *Water Resources Research*, 7: 543-53.

[92] _____________ (1972), ‘Statistical Methodology for Non-periodic Cycles: From the Covariance to R/S Analysis’, *Annals of Economic and Social Measurement*, 1: 259-90.

[93] _____________ (1975), ‘Limit Theorems on the Self-normalized Range for Weakly and Strongly Dependent Processes’, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebeite*, 31: 271-85.

[94] _____________, and M.S. TAQQU (1979), ‘Robust R/S Analysis of Long-run Serial Correlations’, *Bulletin of the International Statistical Institute*, 48,2: 69-104.

[95] _____________, and J.W. VAN NESS (1968), ‘Fractional Brownian Motions, Fractional Noises and Applications’, *SIAM Review*, 10: 422-37.

[96] _____________, and T.R. WALLIS (1969), ‘Robustness of the Rescaled Range R/S in the Measurement of Noncyclic Long Run Statistical Dependence’, *Water Resources Research*, 5: 967-88.

[97] MARINUCCI, D., and P.M. ROBINSON (1999), 'Alternative Forms of Fractional Brownian Motion', *Journal of Statistical Planning and Inference*, 80: 111-22.

[98] __________ (2000), 'Weak Convergence of Multivariate Fractional Processes', *Stochastic Processes and their Applications*, 86:103-20.

[99] MOULINES, E., and P. SOULIER (1999), 'Broadband Log Periodogram Regression of Time Series with Long Range Dependence', *Annals of Statistics*, 27: 1415-39.

[100] ________, and P. SOULIER (2000), 'Data Driven Order Selection for Projection Estimates of the Spectral Density of Time Series with Long Range Dependence', *Journal of Time Series Analysis*, 21: 193-218.

[101] NELSON, D. (1991), 'Conditional Heteroskedasticity in Asset Returns: A New Approach', *Econometrica*, 59: 347-70.

[102] NEWCOMB, S. (1886), 'A Generalized Theory of the Comination of Observations so as to Obtain the Best Result', *American Journal of Mathematics*, 8: 343-66.

[103] PARZEN, E. (1986), 'Quantile Spectral Analysis and Long-memory Time Series', *Journal of Applied Probability*, 23A: 41-54.

[104] PHILLIPS, P.C.B. (1987), 'Time Series Regression with a Unit Root', *Econometrica*, 55: 277-301.

[105] __________ (1991a), 'Optimal Inference in Cointegrated Systems', *Econometrica*, 59: 283-306.

[106] __________ (1991b), 'Spectral Regression for Cointegrated Time Series', in W.A. Barnett, J. Powell and G. Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge: Cambridge University Press, 413-435.

[107] __________ (1999), 'Unit Root Log Periodogram Regression', preprint.

[108] ROBINSON, P.M. (1978a), 'Statistical Inference for a Random Coefficient Autoregressive Model', *Scandinavian Journal of Statistics*, 5: 163-8.

[109] ___________ (1978b), 'Alternative Models for Stationary Stochastic Processes', *Stochastic Processes and their Applications*, 8: 151-52.

[110] ___________ (1991), 'Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression', *Journal of Econometrics*, 47: 67-84.

[111] ___________ (1994a), 'Time Series with Strong Dependence', in C.A. Sims, ed., *Advances in Econometrics*, vol.1, Cambridge: Cambridge University Press, pp.47-95.

[112] ___________ (1994b), 'Semiparametric Analysis of Long-memory Time Series', *Annals of Statistics*, 22: 515-39.

[113] ___________ (1994c), 'Efficient Tests of Nonstationary Hypotheses', *Journal of the American Statistical Association*, 89: 1420-37.

[114] ___________ (1995a), 'Log-periodogram Regression of Time Series with Long Range Dependence', *Annals of Statistics*, 23: 1048-72.

[115] ___________ (1995b), 'Gaussian Semiparametric Estimation of Long-range Dependence', *Annals of Statistics*, 23: 1630-61.

[116] ___________ (2001), 'The Memory of Stochastic Volatility Models', *Journal of Econometrics*, 101: 195-218.

[117] ___________, and M. HENRY (2000), 'Higher-order Kernel Semiparametric M-Estimation of Long Memory', forthcoming *Journal of Econometrics*.

[118] ___________, and J.F. HIDALGO (1997), 'Time Series Regression with Long Range Dependence', *Annals of Statistics*, 25: 77-104.

[119] ___________, and J. HUALDE (2000), 'Cointegration in Fractional Systems with Unknown Integration Orders', preprint.

[120] ___________, and D. MARINUCCI (1997), 'Semiparametric Frequency-domain Analysis of Fractional Cointegration', preprint.

[121] ___________, and D. MARINUCCI (2000), 'The Averaged Periodogram for Non-stationary Vector Time Series', *Statistical Inference for Stochastic Processes*, 3: 149-60.

[122] ___________, and D. MARINUCCI (2001), 'Narrow-band Analysis of Nonstationary Processes', *Annals of Statistics,29: 947-986.*

[123] ___________, and P. ZAFFARONI (1997), 'Modelling Nonlinearity and Long Memory in Time Series', *Fields Institute Communications*, 11: 161-70.

[124] ___________, and P. ZAFFARONI (1998), 'Nonlinear Time Series with Long Memory: A Model for Stochastic Volatility', *Journal of Statistical Planning and Inference*, 68: 359-71.

[125] ROSENBLATT, M. (1961), 'Independence and Dependence', in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California Press, pp.411-43.

[126] SAMAROV, A., and M.S. TAQQU (1988), 'On the Efficiency of the Sample Mean in Long Memory Noise', *Journal of Time Series Analysis*, 9: 191-200.

[127] SINAI, Y. G. (1976), 'Self-similar Probability Distributions', *Theory of Probability and its Applications*, 21: 64-80.

[128] SOWELL, F.B. (1990), 'The Fractional Unit Root Distribution', *Econometrica*, 58: 495-505.

[129] __________ (1992), 'Maximum Likelihood Estimation of Stationary Univatiate Fractionally Integrated Time Series Models', *Journal of Econometrics*, 53: 165-88.

[130] STUDENT (1927), 'Errors of Routine Analysis', *Biometrika*, 19: 151-69.

[131] TAQQU, M.S. (1975), 'Weak Convergence to Fractional Brownian Motion and to the Rosenblatt Process', *Zeitschrift für Wahrscheinlichkeitstheorie*, 31: 287-302.

[132] TAYLOR, S.J. (1986), *Modelling Financial Time Series*, Chichester, UK.

[133] TEVEROVSKY, V., and M.S. TAQQU (1997), 'Testing for Long-range Dependence in the Presence of Shifting Means or a Slowly Declining Trend, Using a Variance-type Estimate', *Journal of Time Series Analysis*, 18: 279-304.

[134] VELASCO, C. (1999a), 'Non-stationary Log-periodogram Regression', *Journal of Econometrics*, 91: 325-71.

[135] _________ (1999b), 'Gaussian Semiparametric Estimation of Nonstationary Time Series', *Journal of Time Series Analysis*. 20: 87-127.

[136] _________ (2000), 'Non-Gaussian Log-periodogram Regression', *Econometric Theory*, 16: 44-79.

[137] _________, and P.M. ROBINSON (2000), 'Whittle Pseudo-maximum Likelihood Estimation for Nonstationary Time Series', *Journal of the American Statistical Association*, 95: 1229-43.

[138] VITALE, R.A. (1973), 'An Asymptotically Efficient Estimate in Time Series Analysis', *Quarterly Journal of Applied Mathematics*, 421-40.

[139] WHISTLER, D. (1990), 'Semiparametric Models of Daily and Intra-daily Exchange Rate Volatility', Ph.D. thesis, University of London.

[140] WHITTLE, R. (1951), *Hypothesis Testing in Time Series Analysis*, Uppsala: Almqvist.

[141] YAJIMA, Y. (1985), 'On Estimation of Long-memory Time Series Models', *Australian Journal of Statistics*, 27: 303-20.

[142] _______ (1988), 'On Estimation of a Regression Model with Long-memory Stationary Errors', *Annals of Statistics*, 16: 791-807.

[143] _______ (1991), 'Asymptotic Properties of the LSE in a Regression Model with Long Memory Stationary Errors', *Annals of Statistics*, 19: 158-77.

[144] YONG, C.H. (1974), *Asymptotic Behaviour of Trigonometric Series*, Hong Kong: Chinese University.