PARAMETRIC ESTIMATION UNDER LONG-RANGE DEPENDENCE*

by

Liudas Giraitis and Peter M Robinson
London School of Economics and Political Science

Contents:

Discussion paper
No. EM/01/416
May 2001

The Suntory Centre
Suntory and Toyota International Centres for Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020-7955 6698

* Research supported by ESRC Grant R000238212. The second author’s research was also supported by a Leverhulme Trust Personal Research Professorship.

Abstract

Parametric estimation is discussed in a variety of models exhibiting long-range dependence.

Keywords: Parametric estimation; long-range dependence.
JEL No.: C22

Contact address: Professor Peter M Robinson, Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, e-mail: pm.robinson@lse.ac.uk

1 Introduction

We focus on parametric estimation (and associated inference) in the sense that the joint distribution of the (scalar or multiple) time series need not necessarily be a parametric function, but that an aspect of interest is a parametric function which is estimated with 'parametric' rates of convergence. The overall setting can therefore be either purely parametric, or semiparametric. However, we do not concern ourselves with parameter estimates that depend on smoothed estimation of nonparametric nuisance functions (such as probability densities or spectral densities) which might be introduced, in particular, for the purpose of efficiency gain; indeed while semiparametric estimation of this kind has been greatly developed in case of independent and short range dependent processes, there is little work so far in the long range dependent case.

Loosely, one can think of parameters as describing either dynamic or stochastic properties of time series. Examples of 'dynamic' parameters are memory/self similarity parameters, as well as ARMA coefficients in FARIMA models. Examples of 'static' parameters are location and scale parameters and regression coefficients (including fractional cointegrating vectors). We discuss estimation of 'dynamic' parameters in Sections 2 and 3, and 'static' parameters in Section 4. While we do discuss theoretical properties, and give some idea of the circumstances which assure them, our treatment is biased towards 'useful' theory and avoids detailed regularity conditions, while we also give attention to modelling, and the merits of alternative methods of estimation, including computational considerations. Regularity conditions are sometimes rather lengthy when trade-oggs possible can to some degree reflect taste, so that quating regularity conditions from the literature can provide a misleading picture of the significance of certain conditions. Regularity conditions are sometimes rather lengthy and when trade-offs are possible can to some degree reflect taste, so that quoting regularity conditions from the literature can provide a misleading picture of the significance of certain conditions.

Section 2 is mainly motivated by the possibility of Gaussianity, insofar as we discuss, for stationary series, estimation of parametric autocovariance functions and spectral densities, which suffice to describe Gaussian dynamics, with some extension to nonstationary series. The term 'long range dependence' is often taken to imply stationarity, but of course many nonstationary series, such as unit root ones, exhibit even longer range dependence. Rival definitions of stationary long range dependent series $X(n)$, $n = 0, \pm 1, \cdots$, entail the existence of $d \in (0, \frac{1}{2})$, $0 < c < \infty$, $0 < C < \infty$, such that

$$r(n) \stackrel{def}{=} Cov\,(X(0), X(n)) \sim cn^{2d-1}, \text{ as } n \to \infty \tag{1.1}$$

or else

$$g(\lambda) \sim C\lambda^{-2d}, \text{ as } \lambda \to 0+, \tag{1.2}$$

where the spectral density function $g(\lambda)$ satisfies

$$r(n) = \int_{-\pi}^{\pi} g(\lambda) \cos n\lambda d\lambda, \; n = 0, \pm 1, \cdots. \tag{1.3}$$

The most popular parametric models, such as FARIMAS, satisfy both (1.1) and (1.3). As is well known, however, these definitions are not identical. A flaw with (1.1), which is heavily used at the probabilistic end of the literature, is that it does not cover short range dependent series, which can however be described by taking $d = 0$ in (1.2). Invertible negative dependent or antipersistent models, with $-\frac{1}{2} < d < 0$, are also covered by (1.2), and by (1.1) on taking $c < 0$ and $\sum_{n=-\infty}^{\infty} r(n) = 0$. Nonstationary series can be defined such that either a suitable degree of integer differencing produces a stationary series with $-\frac{1}{2} < d < \frac{1}{2}$, or such that fractional differencing produces a short range dependent series.

Much of Section 2 is relevant to many non-Gaussian series also, but Section 3 concerns series that are explicitly non-Gaussian in that the raw time series exhibits no autocorrelation, yet certain instantaneous nonlinear functions (such as squares) are long range dependent. This kind of property has been observed in asset returns and exchange rate data. The extent of rigorous justification of large sample inferences in this setting is presently extremely limited but on grounds of empirical importance we devote a separate section to it.

Section 4, concerning 'static' parameters, begins with discussion of estimating the mean and variance of a stationary series by first and second sample moments. The treatment is again biased towards Gaussianity in that we do not consider other types of estimates of these moments, or estimation of quantities such as the mode. Next we consider regression models, first with nonstochastic regressors and then with stochastic ones, developing the topic here to cointegrated nonstationary series.

The assumption (1.2), with $d > 0$, indicates the characteristic feature of a spectral pole at zero frequency, while the eventual monotonic decay in (1.1) has a similar meaning, and nonstationary extensions intensify these types of long run effect. The bulk of research views long range dependence in this way, as we shall, but one can also study spectral poles at nonzero frequencies and related phenomena, a recent review of this topic being Arteche and Robinson [2].

2 Estimation of 'dynamic' parameters, motivated by Gaussianity

Consider a stationary scalar series $X(n)$, $n = 0, \pm 1, \ldots$. We suppose $X(n)$ is observed at $n = 1, \cdots, N$, though the estimates we give are also relevant when $X(n)$ is an unobservable sequence which can be proxied in terms of observables; for example, $X(n)$ might be the error in a regression model. Bearing in mind that location estimation will be discussed subsequently (Section 4), we make the practically unrealistic assumption that the mean of $X(n)$ is known, and take its value to be zero; we shall however indicate implications of relaxing this assumption.

We suppose there exists a known function $r(n;\theta)$ of n and a $p \times 1$ vector θ, and an unknown $\theta_0 \in \mathbf{R^p}$, such that $r(n) \equiv r(n;\theta_0)$. We assume the spectral density $g(\lambda)$ of $X(n)$ exists, and so we correspondingly have a known function, $g(\lambda;\theta)$, of λ and θ, such that $g(\lambda) \equiv g(\lambda;\theta_0)$. Since we are referring to long range dependence, one element of θ will typically be the parameter d arising in (1.1) and (1.2), which is generally unknown in practice.

Using θ_i to denote the i^{th} element of θ, a simple model is the 'fractional noise', given by

$$r(n;\theta) = \frac{1}{2}\theta_1 \left(|n+1|^{2\theta_2+1} - 2\,|n|^{2\theta_2+1} + |n-1|^{2\theta_2+1} \right),\ n = 0, \pm 1, \cdots, \tag{2.1}$$

so $p = 2$. A process with increments having autocovariance (2.1), for $\theta_1 > 0$, $0 \le \theta_2 < \frac{1}{2}$, is self-similar, with self-similarity parameter $\theta_2 + \frac{1}{2}$, see Mandelbrot and Van Ness [37].

Another class of model, the FARIMA, is given by,

$$g(\lambda;\theta) = \frac{\theta_1}{2\pi} \left|1 - e^{i\lambda}\right|^{-2\theta_2} \frac{\left|a(e^{i\lambda};\theta)\right|^2}{\left|b(e^{i\lambda};\theta)\right|^2},\ \pi < \lambda \le \pi, \tag{2.2}$$

where, again, $0 < \theta_2 < \frac{1}{2}$ for long range dependence and $\theta_1 > 0$, while $a(z;\theta)$, $b(z;\theta)$ are finite-degree polynomials having no zero in common and all zeroes outside the unit circle. The model (1.2) originated in work of Adenstedt [1], following earlier work on short-range dependent ARMAS (with $\theta_2 = 0$), see e.g. Box and Jenkins [8], and was explicitly discussed in this generality by Granger and Joyeux [25]. Typically, $a(0;\theta) = b(0;\theta) = 1$ and the coefficients in a and b are distinct, freely varying elements of θ. This is what we term a "standard parameterization", which more generally we characterize by the property

$$\int_{-\pi}^{\pi} \log g(\lambda;\theta) d\lambda = 2\pi \log \theta_1, \tag{2.3}$$

where (as in (2.2) but not in (2.1)) θ_1 denotes the variance of the best linear predictor of $X(n)$ (the "innovations variance"). A related class of models combines the short range dependent exponential model of Bloomfield [6] with Adenstedt's model,

$$g(\lambda;\theta) = \frac{\theta_1}{2\pi} \left|1 - e^{i\lambda}\right|^{-2\theta_2} \exp\left(1 + \sum_{j=3}^{p} \theta_j \cos j\lambda\right), \tag{2.4}$$

also satisfying (2.3), see Robinson [42].

Each of the models (2.1), (2.2) and (2.4) satisfies both (1.1) and (1.2), with $\theta_2 = d$. However the fact that an autocovariance representation has been stressed for (2.1), but a spectral one for (2.2) and (2.4), warrants comment. Of course there is a spectral representation for (2.1) and there are autocovariance ones for (2.3) and (2.4) but these are cumbersome (except for very low order a and b in (2.2)) and their form can vary in a complicated way across the range of parameter values providing stationarity, for example the multiplicity of zeros of a and b in (2.2) varies. Mostly we are led to consider models for which the spectral density is the simpler formula, as this falls out immediately from a lag-operator representation of the process itself, for example if $X(n)$ has spectral density (2.2), we can write

$$(1 - L)^{\theta_2} b(L;\theta) X(n) = a(L;\theta)\varepsilon(n), \;\; n = 0, \pm 1, \cdots, \tag{2.5}$$

where L is the lag operator and $\varepsilon(n)$ is a sequence of uncorrelated zero-mean variates with variance θ_2.

As a further point, the models (2.1), (2.2) and (2.4) all contain a scale parameter θ_1. Unlike with the mean, we have chosen to include it in the current discussion for convenience of exposition, even though it is not a 'dynamic' parameter. However, assuming it varies freely from the other parameters we can, in the methods of estimation we consider, eliminate it at the start, writing its estimate as an explicit function of the estimates of the other parameters. If only large sample inference rules are to be established for the latter, this might have implications for the degree of regularity conditions (such as in relation to moment conditions and compactness). Nearly always, θ_1 is simply a nuisance parameter, of little or no interest in itself and present only to lend reality.

Perhaps the earliest method of estimation in long range dependent models, proposed in relation to (2.1), uses the adjusted rescaled range (R/S) statistic

$$R/S = \frac{\max_{1\le j\le N}\sum_{n=1}^{j}(X(n)-\overline{X}) - \min_{1\le j\le N}\sum_{n=1}^{j}(X(n)-\overline{X})}{\left\{\frac{1}{N}\sum_{n=1}^{N}(X(n)-\overline{X})^2\right\}^{\frac{1}{2}}}, \quad (2.6)$$

where $\overline{X} = N^{-1}\sum_{n=1}^{N}X(n)$ (see Hurst, [33], Mandelbrot and Wallis, [38]). Then $\log(R/S)/\log N - \frac{1}{2}$ can be a consistent estimate of $\theta_2 = d$ in (2.1). This statistic has been much used over the years, and it and modifications are still popular, for example in empirical finance. However, despite its interesting structure, this estimate of d has a limit distribution that is difficult to use in statistical inference and is not based on the traditional statistical principle of whitening, unlike some of its rivals introduced below, while it is not obviously optimal in an asymptotic sense for any class of distributions (and clearly not for Gaussian $X(n)$). This continued popularity, then, may be due in part to insufficient appreciation of the relative merits of alternative approaches.

For Gaussian $X(n)$ it is natural to consider first estimates maximizing the criterion

$$-\frac{1}{2N}\log|\Sigma(\theta)| - \frac{1}{2N}X'\Sigma(\theta)^{-1}X, \quad (2.7)$$

over a subset of the "stationary" domain of $\mathbf{R^p}$, where $X = (X(1), \cdots, X(N))'$ and $\Sigma(\theta)$ has (i,j)th element $r(i-j;\theta)$, because (2.7) is proportional to the log likelihood after omitting a constant. Such an objective function also, of course, arose in earlier work on estimating short range dependent series, where a variety of approximations to (2.7), that typically lead to estimates with the same first order limiting distribution but may have some advantages, have arisen.

To derive the first of these, note first that under regularity conditions

$$\frac{1}{N}\log|\Sigma(\theta)| \to \frac{1}{2\pi}\int_{-\pi}^{\pi}\log g(\lambda;\theta)d\lambda, \qquad \text{as } N\to\infty. \tag{2.8}$$

Now if we can write

$$X(n)=\varepsilon(n)+\sum_{j=1}^{\infty}\phi_j(\theta^*)\varepsilon(n-j),\ \ \sum_{j=1}^{\infty}\phi_j^2(\theta^*)<\infty, \tag{2.9}$$

such that the $\varepsilon(n)$ are uncorrelated with variance θ_1, and we have $\theta = (\theta_1, \theta^{*\prime})'$, so that the autocorrelations of $X(n)$ are free of θ_1, then the "standard parameterization" condition (2.3) is met, and we can write $\Sigma(\theta) = \Sigma^*(\theta^*)/\theta_2$, for some matrix function Σ^*. Thus (2.7) can be approximated by

$$-\frac{1}{2}\log\theta_1 - \frac{1}{2N\theta_1}X'\Sigma^*(\theta^*)^{-1}X. \tag{2.10}$$

It is easily seen that θ^* is thence estimated by minimizing

$$\frac{1}{N}X'\Sigma^*(\theta^*)^{-1}X, \tag{2.11}$$

while θ_1 is estimated by (2.11) at its minimum, illustrating the elimination procedure mentioned earlier. Of course we have (2.9) in case of "standard parameterizations" of (2.2) and (2.4), for example.

An alternative proxy to (2.10), again under (2.3), first replaces the MA representation (2.9) by an AR representation

$$X(n)-\sum_{j=1}^{\infty}\psi_j(\theta^*)X(n-j)=\varepsilon(n), \tag{2.12}$$

and then approximates this, for $n = 1,\cdots,N$, by

$$X(n)-\sum_{j=1}^{n-1}\psi_j(\theta^*)X(n-j)=\tilde{\varepsilon}(n;\theta^*), \quad n=1,\cdots N, \tag{2.13}$$

it being understood that the summation on the left vanishes for $n = 1$. Then in place of (2.10) consider

$$-\frac{1}{2N}\log\theta_1 - \frac{1}{2N}\sum_{n=1}^{N}\frac{\tilde{\varepsilon}(n;\theta^*)^2}{\theta_1}. \tag{2.14}$$

This type of procedure was stressed by Box and Jenkins [8] for estimates of short range dependent models, where it is especially convenient for AR models, indeed there is a closed form solution for the estimates of θ^*. For long range dependent models, however, the heavy truncation entailed in (2.14), for small n, might seem a disadvantage; the $\psi_j(\theta^*)$, though summable, typically converge relatively slowly.

A further approximation to (2.7) is

$$-\frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log g(\lambda;\theta)d\lambda - \frac{1}{4\pi}\int\limits_{-\pi}^{\pi}\frac{I(\lambda)}{g(\lambda;\theta)}d\lambda, \tag{2.15}$$

where $I(\lambda)$ is the periodogram

$$I(\lambda) = \frac{1}{2\pi n}\left|\sum_{n=1}^{N} X(n)e^{in\lambda}\right|^2 \tag{2.16}$$

and $g(\lambda;\theta^*) > 0$ is assumed for all λ. For the purpose of computation, denote by $S(\theta)$ the $N \times N$ matrix with (m,n)th element $S(m-n;\theta)$, where

$$S(j;\theta) = \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} g(\lambda;\theta)^{-1}e^{ij\lambda}d\lambda, \tag{2.17}$$

whence we have

$$\frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\frac{I(\lambda)d\lambda}{g(\lambda;\theta)} = \frac{X'S(\theta)X}{N} \tag{2.18}$$

and $S(\theta)$ can be viewed as approximating $\sum(\theta)^{-1}$ in (2.7).

In case of (2.9), (2.12), we can write $g(\lambda;\theta) = \theta_1 g^*(\lambda;\theta^*)$, for

$$g^*(\lambda;\theta^*) = \frac{1}{2\pi}\left|1+\sum_{j=1}^{\infty}\phi_j(\theta^*)e^{ij\lambda}\right|^2 = \frac{1}{2\pi}\left|1-\sum_{j=1}^{\infty}\psi_j(\theta^*)e^{ij\lambda}\right|^{-2} \tag{2.19}$$

and then, by (2.7), (2.15) is identical to

$$-\frac{1}{2}\log\theta_1 - \frac{1}{4\pi\theta_1}\int\limits_{-\pi}^{\pi}\frac{I(\lambda)}{g^*(\lambda;\theta^*)}d\lambda, \tag{2.20}$$

whence θ_1 can be eliminated as before, and we have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{I(\lambda)}{g^*(\lambda;\theta^*)}d\lambda = \frac{1}{N}X'S^*(\theta^*)X \tag{2.21}$$

for an $N \times N$ matrix S^*.

The final approximations we consider start from (2.16) and replace the integral by a discrete sum. Thus consider

$$-\frac{1}{2\pi}\sum_{j=1}^{N-1}\left\{\log g(\lambda_j;\theta) + \frac{I(\lambda_j)}{g(\lambda_j;\theta)}\right\}, \tag{2.22}$$

where $\lambda_j = 2\pi j/N$. For "standard parameterizations" we can alternatively consider

$$-\frac{1}{2}\log\theta_1 - \frac{1}{2N\theta_1}\sum_{j=1}^{N-1}\frac{I(\lambda_j)}{g^*(\lambda_j;\theta^*)}. \tag{2.23}$$

These forms correspond to an orthonormal transform of the $N \times 1$ vector X except that the sums contain only $N-1$ terms. Under our stated setting of a known (zero) mean for $X(n)$ this is optional and a summand for $j=0$ (or equivalently, by periodicity, $j=N$) can be included in (2.22) and (2.23). However, the omission of $j=0$ allows us to immediately drop the known-mean assumption. To see this, first consider the previous objective functions (2.7), (2.10), (2.14), and (2.15)/(2.20). If $EX(n) = \mu \neq 0$ then as presented these are liable to produce inconsistent estimates of θ, indeed they will be dominated as $N \to \infty$ by a term depending on μ. Of course this can be avoided. One can replace $X(n)$ by $X(n)-\mu$ in the formulae, and then treat μ as a parameter to be estimated simultaneously with the remainder. Or, more simply, one can replace $X(n)$ by $X(n)-\overline{X}$, and then adopt precisely one of the procedures decribed above. With either approach, the asymptotic properties of estimates, as described subsequently, will be the same relative to the zero mean case. However, as seen in Section 4 below, $\overline{X}$ is only N^{1-2d}-consistent for μ under, for example, (1.1), which entails a slower rate of convergence than that of the estimates of θ we have been discussing when $d > 0$ (as is true also of various other estimates of μ). It might then be anticipated that the dependence of estimates of θ on $\overline{X}$, might impair their precision in finite samples. On the other hand, at the frequencies λ_j, $j = 1, ..., N-1$, the periodogram of $X(n) - \overline{X}$ is

$$\frac{1}{2\pi N}\left|\sum_{n=1}^{N}\left(X(n)-\overline{X}\right)e^{it\lambda_j}\right|^2 = I(\lambda_j), \tag{2.24}$$

so that mean-correction is automatically incorporated in (2.22) and (2.23) without explicit dependence on $\overline{X}$; note that the missing periodogram $I(0) = (N/2\pi)\overline{X}^2$. Monte Carlo evidence of Cheung and Diebold [9] has demonstrated that indeed estimates minimizing (2.23) can have superior finite sample properties to some other of the Whittle estimates we have discussed. In short range dependent models, $\overline{X}$ is $N^{\frac{1}{2}}$-consistent and so this apparent finite-sample advantage of (2.22) and (2.23) over the others disappears, and it is thus ironic that whereas they have been quite often considered in the short range dependent literature, the theoretical literature on Whittle estimation under long range dependence has tended to ignore (2.22), (2.23) and instead stressed some of the other forms, especially (2.20).

Moreover, (2.22) and (2.23) might also be preferred on grounds of simplicity, and perhaps computational speed. The computation of (2.7), (2.10), (2.14), (2.16) and (2.21) requires formulae for such quanties as $X' \sum(\theta)^{-1} X$, $\sum_{t=1}^{N} \widetilde{\varepsilon}(n;\theta^*)^2$, and $X'S(\theta)X$ (see (2.19)) which are available (see e.g. Sowell, [50]) but of rather complex form for the models (2.2) (except in very simple versions) and (2.4), especially when $d > 0$, and moreover require knowledge of the multiplicity of zeros of a and b in (2.2). On the other hand, (2.22) and (2.23) directly depend on $g(\lambda_j;\theta)$, which is available automatically in case of (2.2) and (2.4). So again, (2.22) and (2.23) might seem especially suitable for important classes of long range dependent models.

Finally, the computation of periodograms at Fourier frequencies λ_j can be rapidly carried out by means of the fast Fourier transform. Statistics arising in the other Whittle estimates, such as sample autocovariances, can also be computed via the first Fourier transform, but the direct dependence of (2.22) and (2.23) on the $I(\lambda_j)$ appears to give them further advantage, especially in the long series which seem the most natural context for investigating long range dependence.

A preference for (2.22) and (2.23) must, however, be tempered by the lack of any very comprehensive numerical comparison of the various Whittle estimates. Certainly, all of them are liable to be consistent and have the same first order limit distribution and properties, the usual basis for statistical inference, with very similar conditions.

To discuss the asymptotic theory, a convenient starting point is the paper of Hannan [27], who studied (2.7), (2.21) and (2.23), under (2.3) (see also Dzhaparidze, [13]). Though various authors had previously worked on asymptotics for Whittle estimation under short range dependence, Hannan's paper represented a somewhat definitive treatment, and provides some benchmark for what might be achieved under long range dependence. Indeed, Hannan's basic dependence condition for (strong) consistency is ergodicity, which covers long range dependence. For asymptotic normality of estimates of θ^*, his regularity conditions explicitly rule out long range dependence, though in other respects they are noticeably weak, in particular the $\varepsilon(n)$ in (2.9) can be stationary martingale differences with only second moments existing (so that Gaussianity was not required). The central limit theorem has form

$$N^{\frac{1}{2}}(\widehat{\theta}^* - \theta^*) \rightarrow_d N\left(0, \left(\frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial}{\partial\theta}\log g^*(\lambda;\theta^*)\right\}\left\{\frac{\partial}{\partial\theta}\log g^*(\lambda;\theta^*)\right\}' d\lambda\right)^{-1}\right) \tag{2.25}$$

where $\widehat{\theta}^*$ is any of the above estimates of θ^*.

Robinson [40] considered instead (2.22), being concerned with nonstandard parameterizations, and though he again referred only to short range dependent cases his central limit theorem for the estimates of θ^* hints at how a degree of long range dependence might be covered given finite fourth moments. The reason is that he reduces the problem to a central limit theorem for sample autocovariances, which, from Hannan [28], essentially rests on only square integrability of $g(\lambda)$, which is satisfied for the models (2.1), (2.2) and (2.4), for example, when $d < \frac{1}{4}$. The central limit theorem has the form

$$N^{\frac{1}{2}}(\widehat{\theta} - \theta) \rightarrow_d N\left(0, 2\Omega^{-1} + \Omega^{-1}\Xi\Omega^{-1}\right), \tag{2.26}$$

where

$$\Omega = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\partial}{\partial\theta}\log g(\lambda;\theta)\right)\left(\frac{\partial}{\partial\theta}\log g(\lambda;\theta)\right)' d\lambda, \tag{2.27}$$

$$\Xi = \frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{f(\lambda,\mu,-\mu)}{g(\lambda,\theta)g(\mu,\theta)}\left(\frac{\partial}{\partial\theta}\log g(\lambda;\theta)\right)\left(\frac{\partial}{\partial\theta}\log g(\mu;\theta)\right)' d\lambda du \tag{2.28}$$

where f is the fourth cumulant spectral density of $X(n)$ (vanishing under Gaussianity) and $\widehat{\theta}$ is the estimate of θ. In order to employ this result in statistical inference, a consistent estimate of Ω is

$$\frac{1}{N}\sum_{j=1}^{N-1}\left\{\frac{\partial}{\partial\theta}\log g(\lambda_j;\widehat{\theta})\right\}\left\{\frac{\partial}{\partial\theta}\log g(\lambda_j;\widehat{\theta})\right\}', \tag{2.29}$$

with an analogous expression in case of (2.25). When $X(n)$ is Gaussian, $\Xi = 0$, otherwise its consistent estimation was discussed by Taniguchi [52]. Yajima [56] explicitly considered the central limit theorem, in this case $d < \frac{1}{4}$, for the estimates minimizing (2.14) and (2.20) under (2.3), in case of model (2.2) with $a \equiv b \equiv 1$ *a priori* that is

$$g(\lambda;\theta) = \frac{\theta_1}{2\pi}\left|1 - e^{i\lambda}\right|^{-2\theta_2}. \tag{2.30}$$

A major breakthrough, again with respect to the objective function (2.20) under (2.3) and for Gaussian $X(n)$ with long range dependence of a rather general parametric form, was Fox and Taqqu [17]. Their basic insight was that the vanishing of $g^*(\lambda;\theta^*)^{-1}$ at $\lambda = 0$ in (2.21) compensates for the blowing up of $I(\lambda)$ there, so that square integrability is no longer necessary and any $\theta_2 < \frac{1}{2}$ is permitted. Again under Gaussianity, Dahlhaus [10] studied both (2.7) and (2.15), showing that the Cramer-Rao efficiency bound is still obtained under long range dependence. For the same estimate as Fox and Taqqu [17], Giraitis and Surgailis [22] relaxed Gaussianity of $X(n)$ to a linear process representation in independent identically distributed innovations with finite fourth moments. Subsequent references are Heyde and Gay [31], Hosoya [32], who consider multivariate models and allow martingale difference innovations and more general models.

For short range dependent models, Whittle estimates will be asymptotically normal under a very wide variety of weak dependence conditions even if these have not all been explicitly studied, for example, various mixing conditions. With long range dependence, however, many situations have arisen, with a variety of statistics, in which non-normal limit distributions arise, due to forms of nonlinearity and starting with the work of Rosenblatt [48] (see Section 4). This is certainly the case with Whittle estimates also.

Giraitis and Taqqu [23] have shown that if $X(n) = P(Y(n))$ is a polynomial function of a Gaussian long range dependent process $Y(n)$ and $X(n)$ has spectral density $\theta_1 g^*(\lambda;\theta^*)$, satisfying (2.3), then the estimate of θ^* minimising the objective function (2.20) is consistent but not necessarily $\sqrt{N}$-consistent. They found that the compensation effect in (2.20) when $X(n)$ is Gaussian or linear is the exception rather than the rule, and that in general the class of limit distributions is much richer: $\hat{\theta}^* - \theta^*$ can behave like the sample mean $N^{-1}\sum_{n=1}^{N} Y(n)$ of $Y(n)$ and therefore be asymptotically Gaussian but with normalisation different from $\sqrt{N}$; the limit distribution of $\sqrt{N}(\hat{\theta}^* - \theta^*)$ may be Gaussian with a different covariance matrix from

(2.25), for example, $N^\gamma(\hat{\theta}^* - \theta^*)$, $(1/2 < \gamma < 1)$ can have a degenerate limit of type $\Omega^{-1}\rho\xi$ where ρ is a vector and ξ is a $N(0,1)$ scalar; $\hat{\theta}^* - \theta^*$ can have a limiting Rosenblatt distribution with normalisation different from $\sqrt{N}$.

While Whittle estimation has dominated the literature on estimating models with long range dependence, some mention must be made of alternative methods. In the simple model (2.30) Kashyap and Eom [34] proposed

$$\widehat{\theta}_2 = -\frac{1}{2}\frac{\sum_{j=1}^{N-1} \log \left|1 - e^{i\lambda_j}\right| \log I(\lambda_j)}{\sum_{j=1}^{N-1} (\log |1 - e^{i\lambda_j}|)^2}, \tag{2.31}$$

which comes out of logging (2.31), replacing $g(\lambda;\theta)$ by $I(\lambda)$, and employing least squares. This estimate is less efficient than a Whittle estimate for (2.30), having relative efficiency $6/\pi^2$, but, being defined in closed form, has some computational advantage over the implicitly-defined Whittle estimate. Robinson [42] extended (2.31) to the model (2.4) with $p > 1$, pointing out a desirable orthogonality property of the estimates of $\theta_3, ..., \theta_{p-1}$. Recently, asymptotic theory for such estimates has been given by Moulines and Soulier [39], for more general models and Gaussian $X(n)$. Though Gaussianity can doubtless be relaxed, it is unlikely that as neat conditions can be achieved here as for Whittle estimates, whose basic statistics are quadratic forms rather than the, mathematically rather inconvenient, weighted averages of logged periodograms. The main appeal of these parametric log periodogram estimates is their computational simplicity in case of the model (2.4), but the greater efficiency of the Whittle estimates can then be achieved by just one Newton-type step, in terms of Whittle function gradients.

Other estimates have been considered. Define the sample autocovariances

$$\widehat{r}(j) = \frac{1}{N}\sum_{j=1}^{N-j} \left(X(n) - \overline{X}\right)\left(X(n+j) - \overline{X}\right), \quad 0 \le j < N, \tag{2.32}$$

and the vector $\widehat{r} = (\widehat{r}(0), ..., \widehat{r}(s))'$, $s < N$. Writing $r(\theta) = (r(0;\theta), ..., r(s;\theta))'$, we can estimate θ by minimizing

$$(\widehat{r} - r(\theta))' A (\widehat{r} - r(\theta)), \tag{2.33}$$

where A is a prescribed $(s+1)\times(s+1)$ matrix. When θ has $p = s+1$ elements (cf (2.2), (2.4)), (2.32) amounts just to the method of moments, solving the simultaneous equations $\widehat{r}(j) = r(j;\widehat{\theta})$, $j = 0, ..., p-1$. With $s > p-1$, it is a version of what econometricians call ‘generalized method of moments’. While any $s \in [p-1, N-1]$ can legitimately be considered, it seems appropriate that s be regarded as increasing with N in long range dependent models, because intuitively estimation of d should make appropriate use of as long run information as is available. Whittle estimates, however, automatically achieve this, and naturally compensate for long memory, and while it is possible to discuss choices of A in (2.33) which can lead to a matching of Whittle efficiency, in general the limiting variance matrix of the estimates is relatively cumbersome, involving fourth cumulants even for ‘standard parameterizations’, while the relatively complicated nature of the $r(n;\theta)$ and the dependence on $\overline{X}$ makes it less attractive than (2.22) and (2.23) on computational grounds, and possibly finite sample statistical ones. While it may be that it can sometimes exhibit finite sample superiority over versions of Whittle it seems hard to see, at least on the bases of currently available information, how it might be preferred by a worker who fully understands the characteristics of Whittle estimation we have described.

We now consider extension to nonstationary processes. Let us look first at two modified versions of (2.30), namely

$$X(n) = (1-\alpha L)^{-1}\varepsilon(n), \quad n \geq 1 \tag{2.34}$$

$$X(n) = (1-L)^{-d}\varepsilon(n), \quad n \geq 1, \tag{2.35}$$

assuming that

$$\varepsilon(n) = 0, \quad n \leq 0. \tag{2.36}$$

The initial condition (2.36) ensures that versions of (2.34) and (2.35) with α and d in the ‘nonstationary’ regions $|\alpha| \geq 1$ and $d \geq \frac{1}{2}$ are well-defined (though this would be true of (2.34) under a milder condition, such as $\varepsilon(0) = 0$). Even for “stationary” values of α and d (i.e. $|\alpha| < 1$ and $d < \frac{1}{2}$), $X(n)$ given by (2.34) and (2.35) are only “asymptotically stationary”, but our interest here is in the nonstationary regions of the parameter space. A special case is $\alpha = 1$, $d = 1$, when (2.34) and (2.35) are identical. It is well known that least squares and other popular estimates of α have nonstandard limit distributions when $\alpha = 1$, while associated test statistics, such as score statistics (as usually defined) for testing $\alpha = 1$ have nonstandard null limit distributions. However, Robinson [43] showed that score tests for $d = 1$ in (2.35) have standard (χ^2) null limit distributions, indeed this is the case for all other values of d, stationary and nonstationary. These different outcomes also appear in more general versions of (2.35), in particular extending (2.34) to (2.5) with $\theta_2 < \frac{1}{2}$ but with $b(z;\theta)$ having a zero on the unit circle, and extending (2.35) to (2.5) with $d \geq \frac{1}{2}$ but $b(z;\theta)$ having all zeros outside the unit circle. Likewise, it seems that estimates of d and the other parameters can have standard asymptotics when $d \geq \frac{1}{2}$. Beran [4] asserts this when using the objective function (2.14). Velasco and Robinson [54] have considered a version of (2.22) after multiplying $\{X(n)\}$ by a data taper. Providing the data taper has sufficient "smoothness" properties an outcome similar to (2.5) is achieved, for any $d \geq \frac{1}{2}$, with an extended definition of the spectrum and multiplying the variance matrix in the limit distribution by a factor (greater than 1) of the data taper.

3 Estimation of 'dynamic' parameters, motivated by non-Gaussianity

It is important to stress that while the central limit theory for Whittle estimates is affected at the very most by the addition of a fourth cumulant term in the limiting variance matrix (see (2.25), (2.26)), when $X(n)$ is linear but non-Gaussian, they are not the most efficient approach here. Given a parametric form $f(\varepsilon;\nu)$ for the probability density of the $\varepsilon(n)$ in (2.9), (2.12), assumed independent and identically distributed, where ν is a vector of parameters, likelihood considerations suggest maximizing, for example,

$$\prod_{n=1}^{N} f\left(\widetilde{\varepsilon}(n;\theta);\nu\right) \tag{3.1}$$

with respect to θ and ν, where the $\widetilde{\varepsilon}(n;\theta)$ are given by (2.13) (cf (2.14)). If f has been correctly specified one expects the estimates to achieve the asymptotic Cramer-Rao bound, which will not be attained by Whittle estimates if f is not the normal density. Such an approach, incidentally, also allows one to impose empirically observed phenomena such as asymmetry or long-tailedness, for example. However, use of (3.1) is not necessarily guaranteed to produce consistent estimates of θ if f is mis-specified, so that something of the robustness property of Whittle estimation may be lost.

For some forms of non-Gaussianity that arise in practice, the Whittle estimates of Section 2 are not at all informative. Some time series, such as asset returns or exchange rates, can exhibit little or no autocorrelation, but cannot be regarded as independent across time because certain instantaneous nonlinear functions, such as squares and absolute values, are clearly correlated. When the correlations are consistent with short range dependence, the $ARCH(p)$ and $GARCH(p,q)$ models for "autoregressive conditional heteroscedasticity", see e.g. Engle [15], or the stochastic volatility models of Taylor [53], may be appropriate, though of course any number of nonlinear models can be designed to give rise to this kind of behaviour. However empirical evidence can also be suggestive of long range dependence in the nonlinear functions, and we discuss the parametric modelling and estimation of this phenomenon.

We first consider extensions of Engle's [15] $ARCH(p)$ model. This emphasizes autocorrelation in squares $X(n)^2$, and starts from the conditional moment restrictions

$$E\left(X(n)\,|F_{n-1}\right) = 0 \tag{3.2}$$

$$V\left(X(n)\,|F_{n-1}\right) = \sigma_n^2 \tag{3.3}$$

almost surely, where F_n is the σ-field of events generated by $X(m)$, $m \leq n$. Clearly (3.2) entails $EX(n) = 0$ (which can be relaxed) but also $r(n) = 0$, all $n \neq 0$. On the other hand, the conditional variance σ_n^2 in (3.3) is a function of $X(n-1)$, $X(n-2)$, ..., in general.

As a prescription for (3.3), consider first

$$\sigma_n^2 = \sigma^2 + (1 - \tau(L))\left(X(n)^2 - \sigma^2\right) \tag{3.4}$$

where $\sigma^2 = E\sigma_n^2 = EX(n)^2$, under the presumption that $X(n)$ is stationary, with $\tau(L) = 1 - \sum_{j=1}^{\infty} \tau_j L^j$. When the weights τ_j satisfy $\tau_j = 0$, $j > p$, we have the $ARCH(p)$ model of Engle [15], whereas for suitably chosen exponentially decaying τ_j we have the $GARCH(p,q)$ model of Bollerslev [7], both of which extail short range dependence. However, rewriting (3.4) as

$$\tau(L)(X(n)^2 - \sigma^2) = X(n)^2 - \sigma_n^2, \tag{3.5}$$

the right side is a martingale difference, in view of (3.3) so that, with reference to (2.2), say, choice of

$$\tau(L) = (1 - L)^d a(L)/b(L) \tag{3.6}$$

with $0 < d < \frac{1}{2}$, we might thereby produce squares $X(n)^2$ that have long range dependent autocorrelation, even though $X(n)$ itself is uncorrelated.

The general set up (3.2) - (3.4), and the special case $\tau(L) = (1-L)^d$ of (3.6) (cf (2.30)), was discussed by Robinson [41] in a hypothesis testing context, the latter special case also being considered by Ding and Granger [12], while extensions of (3.4) under (3.6) have been considered, for example one implying infinite variance $X(n)$ for any $d > 0$, under the acronym 'FIGARCH'. Giraitis, Kokoszka and Leipus [20] have considered sufficient conditions for a stationary solution of (3.5) when (3.2) and (3.3) are satisfied by $X(n) = \sigma_n \varepsilon(n)$, where σ_n is the positive square root of σ_n^2 and $\varepsilon(n)$ is an independent identically distributed sequence. Imposing the requirements $\tau_j \geq 0$, for all j, $\sum_{j=1}^{\infty} \tau_j < 1$, which are sufficient for $\sigma_n^2 \geq 0$ for all n, the sufficient conditions of Giraitis, Kokoszka and Leipus [20] do not, however, permit long range dependent $X(n)^2$.

The short range dependent ARCH literature contains many functional forms, so that there are other models besides (3.4) that might have the potential to entail long range dependence in $X(n)^2$ and other functions. Another case considered by Robinson [41] was

$$\sigma_n^2 = \left(1 + \sum_{j=1}^{\infty} \tau_j^2\right)^{-1} \left(\sigma + \sum_{j=1}^{\infty} \tau_j X(n-j)\right)^2 . \tag{3.7}$$

This can also be viewed as an extended type of one version of a bilinear model (see Granger and Andersen, [24]), though not a version that has been investigated in the bilinear time series literature. Giraitis, Robinson and Surgailis [21] consider a reparameterized version of (3.2), (3.3) and (3.7), with $X(n) = \sigma_n \varepsilon(n)$, where σ_n is not necessarily the positive square root of σ_n^2, but rather the linear-in-X's square root. Now the constraint $\tau_j \geq 0$ is not necessary. These authors show that there exist weights r_j such that the processes $X(n)^\ell$, $\ell \geq 2$, have autocorrelation consistent with long range dependence.

For estimates of a parameterized $\sigma_n^2 = \sigma_n^2(\theta)$ in (3.4) it is convenient to suppose that $X(n)$ is Gaussian conditional on F_{n-1}, as often in the short range dependence literature. Now given $X(1), ..., X(N)$, the $\sigma_n^2(\theta)$ arising from the examples (3.4), and (3.7) are not computable, so we need to proxy $\sigma_n^2(\theta)$ by $\tilde{\sigma}_n^2(\theta)$ depending only on $X(n-1), ..., X(1)$, for example by taking $\tau_j = 0$, $j \geq n$. Likelihood considerations then lead to

$$\sum_{j=1}^{N} \left\{ \log \tilde{\sigma}_n^2(\theta) + \frac{X(n)^2}{\tilde{\sigma}_n^2(\theta)} \right\} . \tag{3.8}$$

However, we know of no rigorous asymptotic theory for estimates of θ minimizing (3.8) in the long range dependent circumstances envisaged above.

There is a model which provides uncorrelated $X(n)$ and $X(n)^2$ with long range dependent autocorrelations for which asymptotic theory of estimates has been given. Robinson and Zaffaroni [46, 47] consider models including

$$X(n) = \eta(n) \left\{ \alpha + \sum_{j=1}^{\infty} \tau_j \varepsilon(n-j) \right\} \tag{3.9}$$

where $\{\varepsilon(n)\}$ and $\{\eta(n)\}$ are each sequences of independent and identically distributed random variables, and either $\eta(n) \equiv \varepsilon(n)$, when (3.9) is a nonlinear moving average model, or $\{\varepsilon(n)\}$ and $\{\eta(n)\}$ are mutually independent, when (3.9) can be compared with (short-range dependent) "two-shock" stochastic volatility models of Taylor [53]. In either case Robinson and Zaffaroni [46, 47] showed that the τ_j can be chosen to provide $X(n)^2$ with long range dependent autocorrelation. Unfortunately, after parameterizing the τ_j, approximate maximum likelihood estimation, cf (3.8), seems relatively intractable even computationally. Instead Robinson and Zaffaroni [46, 47] proposed Whittle estimation as considered in Section 2 but based on the $X(n)^2$ sequence, having derived formula for their spectral density and autocovariances function. Of course these estimates can never be asymptotically efficient, as Gaussian $X(n)^2$ is a logical impossibility, but Zaffaroni [59] has derived a central limit theorem analogous to (2.26). Harvey [30] considered Whittle estimation for an alternative functional form to (3.9), for the case of $\{\eta(n)\}$ independent of $\{\varepsilon(n)\}$, and the factor in braces replaced by $e^{\alpha + \sum_{j=1}^{\infty} \tau_j \varepsilon(n-j)}$ (cf Taylor, [53]) with certain long range dependent weights τ_j, but gave no asymptotic theory. Whittle estimation based on squares can also be used in the ARCH-type models discussed above, but again there is currently no asymptotic theory covering long range dependence.

4 Estimation of 'static' parameters

We first pass briefly over the topic of scale estimation, bearing in mind that Section 2 essentially covered estimation of the variance of uncorrelated innovations in long range dependent models. Here we consider direct estimation of the variance $\sigma^2 = V(X(n))$ of a stationary long range dependent series without necessarily having a parameterization of the dependence structure, though the parameter θ_1 in (2.1) is an example of such a variance. Assuming first that $X(n)$ has known, zero, mean, we can estimate σ^2 by

$$\widehat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N} X(n)^2. \tag{4.1}$$

For a very wide range of short range dependent $X(n)$, $N^{\frac{1}{2}}(\widehat{\sigma}^2 - \sigma^2)$ is asymptotically normally distributed. This can still be the case under (1.1) with $0 < d < \frac{1}{4}$. Taking $X(n)$ to be Gaussian, we then have

$$n^{\frac{1}{2}}(\widehat{\sigma}^2 - \sigma^2) \to_d N\left(0, 2\sum_{n=-\infty}^{\infty} r(n)^2\right) = N\left(0, \frac{4C^2}{4d-1}\right) \tag{4.2}$$

with a fourth cumulant term appearing in case of non-Gaussian $X(n)$ (assuming finite fourth moments). For $d \geq \frac{1}{4}$, however, $r(n)$ is not square summable, and Rosenblatt [48] showed that $N^{2d}(\widehat{\sigma}^2 - \sigma^2)$ converges to a certain nonnormal, nonstandard distribution, which Taqqu [51] termed the 'Rosenblatt distribution'. A similar result holds for certain more general, non-Gaussian, $X(n)$, as shown by Taqqu [51]. If $EX(n)$ is unknown, and instead we estimate σ^2 by

$$\widehat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N}\left(X(n) - \overline{X}\right)^2, \tag{4.3}$$

then, for $0 < d < \frac{1}{4}$, (4.2) still holds, but the limit distribution of $N^{2d}(\widehat{\sigma}^2 - \sigma^2)$ for $d > \frac{1}{4}$ contains an additional term, besides the Rosenblatt one.

In location and regression estimation, a prime interest is efficiency, assuming $X(n)$ has finite variance. The estimates of $\mu = EX(n)$ of particular interest in the Gaussian case are the sample mean or ordinary least squares (OLS) estimate and the generalized least squares (GLS) estimate

$$\widetilde{X} = \frac{1' \sum^{-1} X}{1' \sum^{-1} 1}, \tag{4.4}$$

where 1 is the column vector of N 1's, and $\sum$ is the $N \times N$ matrix with (m,n)th element $r(m-n)$. In practice $\sum$ will rarely be assumed known, but if parameterized it can be estimated as described in Section 2.

For short range dependent $X(n)$, in particular when $g(\lambda)$ is continuous and positive at $\lambda = 0$, then

$$V(\overline{X}) \sim V(\widetilde{X}) \sim \frac{2\pi}{n} g(0), \quad \text{as } n \to \infty, \tag{4.5}$$

so that no efficiency loss is incurred by $\overline{X}$ (see Grenander, [26]). Under long range dependence however, this is no longer the case. Nevertheless, available numerical evidence, in case of the simple model (2.30), is that the asymptotic efficiency loss of $\overline{X}$ is very small (see Samarov and Taqqu, [49]). It can be much greater if $X(n)$ is negative dependent, so $d < 0$, see Vitale [55], Adenstedt [1].

Beran and Künsch [5] showed that a large class of M - estimates of location have the same asymptotic efficiency as OLS in case of Gaussian long range dependent observations; this contrasts with the situation for independent and short range dependent observations, where M - estimators have a different asymptotic variance from (4.5).

Yajima [57, 58] considered the more general, linear regression, model

$$Y(n) = \beta' Z(n) + X(n), \tag{4.6}$$

where $Z(n)$ and β are $p \times 1$ vectors. We observe $Y(n)$, $Z(n)$, $n = 1, ..., N$, and because $Z(n)$ can include an intercept we take $EX(n) = 0$. Yajima [57, 58] treated the $Z(n)$ as deterministic sequences, including the case

$$Z(n) = \left(1, n, n^2, ..., n^{p-1}\right)'. \tag{4.7}$$

Again we know, from Grenander [26], that the OLS and GLS estimates of β have the same asymptotic variance, in case of (4.7), when $g(\lambda)$ is continuous and positive at $\lambda = 0$. For long range dependent $X(n)$, Yajima [57, 58] described the asymptotic variance of OLS and GLS under (4.7) and more generally, with numerical evidence including how the efficiency of OLS decreases with p under (4.7). Both estimates are asymptotically normal when $X(n)$ is a linear long range dependent process (early references being Eicker [14], Hannan, [29]), but not necessarily otherwise (see Taqqu, [51]). Dahlhaus [11] established similar results to Yajima [58], investigating the efficiency of GLS for $Z(n)$ that span the subspace spanned by certain Jacobi polynomials, and, when $X(n)$ has spectrum (2.30), showing that the same efficiency can be achieved when θ_2 is estimated. Beran [3], Koul [35], Koul and Mukherjee [36] extended results of Beran and Künsch [5] to M - and R - estimators of the slope coefficient elements of β in (4.6) with deterministic $Z(n)$ when $X(n)$ are long range dependent Gaussian. The results were further extended to linear long memory processes by Giraitis, Koul and Surgailis [18]. Giraitis and Koul [19] investigated the exact maximum likelihood estimator of the memory parameter of $X(n)$ in (4.6) when $X(n)$ is a nonlinear transformation of a Gaussian long range dependent process.

When $Z(n)$ is stochastic a rather different theory prevails. Let both $X(n)$ and $Z(n)$ in (4.6) have long range dependence, and $Z(n)$ have lag-n autocorrelation decaying like n^{2c-1}, $0 < c < \frac{1}{2}$ (cf. (1.1)), where for simplicity we take $p = 1$. While OLS can still be asymptotically normal when $c + d < \frac{1}{2}$, it has a non-standard limit distribution for $c + d \geq \frac{1}{2}$ (see Robinson, [42]). Robinson and Hidalgo [44] considered a general class of estimates which entail, in the frequency domain, a weight function $\phi(\lambda)$ which is zero at frequency zero and alleviates the possibly strong spectral poles of $X(n)$, $Z(n)$ at frequency zero. They showed such estimates to be asymptotically normal. Because GLS corresponds to taking $\phi(\lambda) = g(\lambda)^{-1}$ (which included zero at $\lambda = 0$ for long range dependent $X(n)$), we find a new advantage in GLS beyond the traditional one of improved efficiency. Robinson and Hidalgo [44] gave extensions to parametric autocorrelation in $X(n)$, with estimated parameters and nonlinear regression models.

When both $X(n)$ and $Z(n)$ are stationary, they need to be at least uncorrelated in order to avoid asymptotic bias in both OLS and GLS. In econometrics, regression models also arise in which $Z(n)$ is stochastic but nonstationary, for example having a unit root, and $X(n)$ can be stationary or nonstationary. In cointegration analysis (see e.g. Engle and Granger [16]) we interpret the regression relation (4.6) such that $X(n)$ is short range dependent, or stationary long range dependent, or nonstationary but less so than $Z(n)$, perhaps using a definition of nonstationary series in the spirit of (2.35). There is typically no natural reason to assume uncorrelatedness between $X(n)$ and $Z(n)$ here, but fortunately that is unnecessary for consistency of, for example, OLS, due to the asymptotic dominance of $X(n)$ by $Z(n)$. Nevertheless there is a bias due to the correlation, which can even affect rates of convergence. After writing OLS as a decomposition of frequency-domain quantities across the frequencies $\lambda_j = 2\pi j/n$, Robinson and Marinucci [45] found that an estimate based on a possibly arbitrarily slowly increasing number of such frequencies, nearest to frequency zero, incurs less bias and thereby has asymptotic properties that are at least as good as, and sometimes better than OLS, depending on the degrees of dependence of $X(n)$ and $Z(n)$.

References

[1] R. Adenstedt. On large-sample estimation for the mean of a stationary random sequence. *Annals of Mathematical Statistics*, 2:1095-107, 1974.

[2] J. Arteche and P.M. Robinson. Seasonal and cyclic long memory. In S. Ghosh, editor, *Asymptotics, Nonparametrics and Time Series: A Tribute to Madan Lal Puri.* Marcel Dekker, New York, 125-145, 1999.

[3] J. Beran. M-estimators for location for Gaussian and related processes with slowly decaying serial correlations. *Journal of the American Statistical Association*, 86:704-708, 1991.

[4] Beran, J. Maximum likelihood estimation of the differencing parameter for invertible short- and long-memory ARIMA models. *Journal of the Royal Statistical Society, Series B*, 57:659-672, 1995

[5] J. Beran and H. Künsch. Location estimators for processes with long range dependence. *Research Report No* 40, Seminar für Statistik, ETH, Zurich, 1985.

[6] P. Bloomfield. An exponential model for the spectrum of a scalar time series. *Biometrika*, 60:217-26, 1973.

[7] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307-327, 1986.

[8] G.E.P. Box and G.M. Jenkins. *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco, 1971.

[9] Y. Cheung and F. Diebold. On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean. *Journal of Econometrics*, 62:301-316, 1994.

[10] R. Dahlhaus. Efficient parameter estimation for self-similar processes. *Annals of Statistics*, 17:1749-1766, 1989.

[11] R. Dahlhaus. Efficient location and regression estimation for long range dependent regression models. *Annals of Statistics*, 23:1029-1047, 1996.

[12] Z. Ding and C.W.J. Granger. Modelling volatility persistence of speculative returns: a new approach. *Journal of Econometrics*, 73:185-215, 1996.

[13] K.O. Dzhaparidze. A new method for estimating spectral parameters of a stationary time series. *Theory of Probability and its Applications*, 19:122-132, 1974.

[14] F. Eicker. Limit theorems for regressions with unequal and dependent errors. In *Proc. Fifth Berkeley Symp. Math. Statist. Probab.*, vol. 1, 59-82. Univ. California Press, Berkeley, 1967.

[15] R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987-1008, 1982.

[16] R.F. Engle and C.W.J. Granger. Co-integration and error-correction: representation, estimation and testing. *Econometrica*, 55:251-276, 1987.

[17] R. Fox and M.S. Taqqu. Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics*, 14:517-532 1986.

[18] L. Giraitis, H. Koul and D. Surgailis. Asymptotic normality of regression estimators with long memory errors. *Statistics and Probability Letters*, 29:317-335, 1996.

[19] L. Giraitis, H. Koul. Estimation of the dependence parameter in linear regression with long-range dependent errors. *Stochastic Processes and their Applications*, 71:207-224, 1997.

[20] L. Giraitis, P. Kokoszka and R. Leipus. Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory*, 16:3-22, 2000.

[21] L. Giraitis, P.M. Robinson and D. Surgailis. A model for long memory conditional heteroscedasticity. *Annals of Applied Probability*, 10:1002-1024, 2000.

[22] L. Giraitis and D. Surgailis. 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, 1990.

[23] L. Giraitis and M.S. Taqqu. Whittle estimator for finite - variance non-Gaussian long-memory time series with long memory. *Annals of Statistics*, 27:178 - 203, 1999.

[24] L.W.J. Granger and A.P. Andersen. *An Introduction to Bilinear Time Series Models*. Vandenhoek and Rupecht, Gottingen, 1978.

[25] L.W.J. Granger and R. Joyeux. An introduction to long-memory time series and fractional differencing. *Journal of Time Series Analysis*, 1:15-30, 1980.

[26] U. Grenander. On the estimation of regression coefficients in the case of autocorrelated disturbance. *Annals of Mathematical Statistics*, 25:252-272, 1954.

[27] E.J. Hannan. The asymptotic theory of linear time series models. *Journal of Applied Probability*, 10:130-145, 1973.

[28] E.J. Hannan. The asymptotic distribution of serial covariances. *Annals of Statistics*, 4:396-399, 1976.

[29] E.J. Hannan. The central limit theorem for time series regression. *Stochastic Processes and their Applications*, 9:281-289, 1979.

[30] A. Harvey. Long memory in stochastic volatility. In J. Knight and S. Satchell, editors, *Forecasting Volatility in the Financial Markets*, Butterworth - Heinemann, Oxford, 1998.

[31] C. Heyde and G. Gay. Smoothed periodogram asymptotics and estimation for processes and fields with long-range dependence. *Stochastic Processes and their Applications*, 45:169-187, 1993.

[32] Y. Hosoya. Limit theory with long-range dependence and statistical inference of related models. *Annals of Statistics*, 25:105-137, 1997.

[33] H. Hurst. Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770-299, 1951.

[34] R.L. Kashyap and K.-B. Eom. Estimation in long-memory time series model. *Journal of Time Series Analysis*, 9:35-41, 1988.

[35] H.L. Koul. M-estimators in linear models with long range dependent errors. *Statistics and Probability Letters*, 14:153-164, 1992.

[36] H.L. Koul and K. Mukherjee. Asymptotics of R-, MD- and LAD estimators in linear regression models with long range dependent errors. *Probability Theory and Related Fields*, 95:538-553, 1993.

[37] B.B. Mandelbrot and J.W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10:422-37, 1968.

[38] B.B. Mandelbrot and J.R. Walliss. Noah, Joseph and operational hydrology. *Water Resources Research*, 4:909-918, 1968.

[39] E. Moulines and P. Soulier. Broad band log-periodogram regression of time series with long range dependence. *Annals of Statistics*, 27:1415-1439, 1999.

[40] P.M. Robinson. Alternative models for stationary stochastic processes. *Stochastic Processes and their Applications*, 8:141-152, 1978.

[41] P.M. Robinson. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47:67-84, 1991.

[42] P.M. Robinson. Time series with strong dependence. In C.A. Sims, editor, *Advances in Econometrics*, vol. 1, 47-95, Cambridge University Press, Cambridge, 1994.

[43] P.M. Robinson. Efficient tests of nonstationary hypotheses. *Journal of American Statistical Association*, 89:1420-1437, 1994.

[44] P.M. Robinson and J.F. Hidalgo. Time series regression with long range dependence. *Annals of Statistics*, 25:77-104, 1997.

[45] P.M. Robinson and D. Marinucci. Narrowband analysis of nonstationary processes, preprint, 1999.

[46] P.M. Robinson and P. Zaffaroni. Modelling nonlinearity and long memory in time series. *Fields Institute Communications*, 11:161-170, 1997.

[47] P.M. Robinson and P. Zaffaroni. Nonlinear time series with long memory: a model for stochastic volatility. *Journal of Statistical Planning and Inference*, 68:359-371, 1998.

[48] M. Rosenblatt. Independence and dependence. In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California Press, 411-443, 1961.

[49] A. Samarov and M.S. Taqqu. On the efficiency of the sample mean in long memory noise. *Journal of Time Series Analysis*, 9:191-200, 1988.

[50] F.B. Sowell. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53:165-188, 1992.

[51] M.S. Taqqu. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie*, 31:287-302, 1975.

[52] Taniguchi, M. On estimation of the integrals of the fourth order cumulant spectral density. *Biometrica*, 69:117-122, 1982.

[53] S.J. Taylor. *Modelling Financial Time Series*, Chichester, UK, 1986.

[54] C. Velasco and P.M. Robinson. Whittle pseudo-maximum likelihood estimation of nonstationary time series, preprint, 1999.

[55] R.A. Vitale. An asymptotically efficient estimate in time series analysis. *Quarterly Journal of Applied Mathematics*, 421-440, 1973.

[56] Y. Yajima. On estimation of long-memory time series models. *Australian Journal of Statistics*, 27:303-320, 1985.

[57] Y. Yajima. On estimation of a regression model with long-memory stationary errors. *Annals of Statistics*, 16:791-807, 1988.

[58] Y. Yajima. Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Annals of Statistics*, 19:158-177, 1991.

[59] P. Zaffaroni. Gaussian estimation of long-range dependent volatility in asset prices, preprint, 1999.