Long Memory in Time Series: Semiparametric Estimation and Conditional Heteroscedasticity

by

MARC HENRY

London School of Economics and Political Science

Thesis submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy at the University of London

To my parents

in his brains,-
Which is as dry as the remainder biscuit
After a voyage,- he hath strange places cramm'd
With observation, the which he vents
In mangled forms.

As You Like It

Abstract

This dissertation considers semiparametric spectral estimates of temporal dependence in time series. Semiparametric frequency domain methods rely on a local parametric specification of the spectral density in a neighbourhood of the frequency of interest. Therefore, such methods can be applied to the analysis of singularities in the spectral density at frequency zero to identify long memory. They can also serve as the basis for the estimation of regular parts of the spectrum. One thereby avoids inconsistency that might arise from misspecification of dynamics at frequencies other than the frequency under focus. In case of long financial time series, the loss of efficiency with respect to fully parametric methods (or full band estimates) may be offset by the greater robustness properties. However, if semiparametric frequency domain methods are to be valid tools for inference on financial time series, they need to allow for conditional heteroscedasticity which is now recognized as a dominant feature of asset returns. This thesis provides a general specification which allows the time series under investigation to exhibit this type of behaviour. Two statistics are considered. The weighted periodogram statistic provides asymptotically normal point estimates of the spectral density at zero frequency for weakly dependent processes. The local Whittle (or local frequency domain maximum likelihood) estimate provides asymptotically normal estimates of long memory in possibly strongly dependent processes. The asymptotic results hold irrespective of the behaviour of the spectral density at non zero frequencies. The asymptotic variances are identical to those that obtain under conditional homogeneity in the distribution of the innovations to the observed process. In semiparametric frequency domain estimation, the choice of bandwidth is crucial. Indeed, it determines the asymptotic efficiency of the procedure. Optimal choices of bandwidth are derived, balancing asymptotic bias and asymptotic variance. Feasible versions of these optimal bandwidths are proposed, and their performance is assessed in an extensive Monte Carlo study where the innovations to the observed process are simulated under numerous parametric submodels of the general specification, covering a wide range of persistence properties both in the levels and in the squares of the observed process. The

techniques described above are applied to the analysis of temporal dependence and persistence in intra-day foreign exchange rate returns and their volatilities. While no strong indication of returns predictability is found in the former, a clear pattern arises in the latter, indicating that intra-day exchange rate returns are well described as martingale differences with weakly stationary and fractionally cointegrated long memory volatilities.

Acknowledgements

The story of my PhD is no Greek tragedy. The plot lacks purity, I never managed to stay in one place and it took me considerably more than one day to write it. However, I was always confident my supervisor, Peter Robinson, would save it from becoming a farce. My gratitude goes to Peter for accepting yet another Ph.D. student in October 1994, notwithstanding the crowds already flocking at his doorstep. My gratitude goes to Peter for spending a considerable amount of time in collaborative work even after he had become aware of my shortcomings. My gratitude goes to Peter for always treating me as an equal, for never putting pressure on a much valued freedom and for providing me with incentives and encouragements when I most needed them.

I have him to thank also for my encounter with the Financial Markets Group which was to prove such a warm and fruitful work environment. A phonecall from Peter, the lack of interest of a colleague leading as I was the stray life of the Library Third Floor, and there I was, masquerading in front of David Webb as a chap interested in Finance. And so I became. The group owes David much of the dynamism and cohesion which made it such a wonderful place to spend four years in. The efficient and friendly assistance of Alison Brower, Lucy Gardiner, Kate Huddie, Marie Kearns, Tara Lehane, Pippa Madicott and Sarah Vink greatly contributed to that end as well. And with the help of Bob Nobay and the "FMG boys" Alvaro Almeida, Ward Brown, Simon Emrich, Sylvain Friederich, Urs Haegler, Richard Payne, Alex Stremme and Father Ted, there was never a day when I didn't enjoy LSE food, let alone beer at the Three Tuns or the Beaver's Retreat. Gleb Sandmann and Geoff Shuetrim proved excellent office mates in the only LSE office with a view of the Old Curiosity Shop and also wonderful academic holiday goers below the glitter of Topkapi and the smoke of a narghile. But the shades of the Bosphorus take me too far in my tale, long after a time when I spent my spare moments chatting away in the artificial lighting of the Lionel Robbins Building. Its winding corridors used to take me along the offices of all the real policy economists, Maia Guell, Alex Heath, Alice Mesnard, Claudio Michelacci, Rohini Pande, Barbara Petrongolo, Hélène Rey, Cecilia Testa and Etienne Wasmer, who helped restore my courage and will to get back to work. When confidence really withered, Javier Hidalgo, smoking away many a nostalgic cheap Spanish cigarette on the steps of St. Clements, would usher me into the entrails of the building to feed on Jörn Rothe's and Rob Northcott's faith in pure theory (and chess).

However, courage, will, faith and fine words did little to help me afford life in London and LSE fees. None of this (not even life in Ian Tonks' little Brixton flat I shared with Alice, Rohini and all the others who took an active part in the Tales of Brixton Motel) would have been possible without the help of a grant from the Ecole polytechnique, administered faithfully by Martine Guibert and awarded under the auspices of the Ecole des Hautes Etudes en Sciences Sociales and the shadow supervision of Gabrielle Demange. When this ran out, I turned to Christian Gouriéroux and was awarded a grant from the Centre de Recherche en Economie et Statistique (CREST). Je désire à ce titre manifester, en français, et on s'apecevra que même l'unité de langue est violée, toute ma reconnaissance à Christian qui m'a accueilli au CREST en enfant de la maison (bien qu'un peu prodigue) et qui m'a tiré d'une mauvaise passe financière en me permettant d'obtenir un financement CREST. J'ai passé d'excellents moments en compagnie de Julien Bechtel, Monica Billio, Cécile Boyer, Eric Burgayran, Laurence Carassus, Serge Darolles, Gaëlle le Fol, Nadine Guedj, Jean-Paul Laurent, Dietmar Leisen, Clotilde Napp, Huyen Pham, Olivier Scaillet, André Tiomo, Nizar Touzi, Fanda Traoré et Jean-Michel Zakoïan au café-comique (souvent mué en café-concert sous l'impulsion de Clotilde et Gaëlle) du cinquième étage de MK2. Ces moments furent très épisodiques, mais sont appelés, je l'espère, à se renouveler dans l'avenir.

Jumping back to the origins, my final words are for Katharina Gassner who gave pride to my econometrics and spice to my life, for Gilles Chemla and Sorbas von Coester who accompanied my first steps in a new life, and finally David Spector for being such a faithful and unflinching conversation topic during coffee breaks.

Contents

List of Tables

List of Figures

Chapter 1

Long memory and conditional heteroscedasticity

1.1 Long memory

The theory of econometric time series is the branch of econometrics concerned with the modelling of dependence across different realisations of a economic process x. Because of the description of the scale of realisations of the process as a time scale, this dependence is usually called temporal dependence and the process is indexed by t. Allowing for temporal dependence in the process implies relaxing the independence part of the traditional assumption of independence and identity of distribution (i.i.d.) for the stochastic process under focus. The identity of distribution part of the i.i.d. assumption is either defined by strict stationarity, meaning that for all positive integers n, $t_1, \ldots, t_n$ and h, the distributions of $(x_{t_1}, \ldots, x_{t_n})$ and $(x_{t_1+h}, \ldots, x_{t_n+h})$ are identical, or partly defined by weak (or covariance) stationarity, meaning that the covariance $E((x_t - E(x_t))(x_s - E(x_s)))$ depends only on $|s-t|$. The latter form of stationarity is implied by the former in case the process has finite variance. The latter form of stationarity is assumed throughout this work for the process x_t. From the Wold Decomposition Theorem (Wold (1938)), each realisation of a purely non-deterministic weakly stationary stochastic process x_t can be described as a result of an infinite weighted sum of uncorrelated errors of common variance, with square summable filtering weights.

$$x_t = E(x_t) + \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \quad \alpha_0 = 1, \quad \sum_{j=0}^{\infty} \alpha_j^2 < \infty, \tag{1.1}$$

with

$$E(\varepsilon_j) = 0 \quad \text{a.s.} \quad \text{and} \quad E(\varepsilon_j \varepsilon_k) = \delta_{jk} \sigma_\varepsilon^2 \quad \text{for all} \quad j, k \geq 0, \tag{1.2}$$

where δ stands for the Kronecker symbol.

A way of relaxing the independence within the i.i.d. assumption while retaining weak dependence between distant x's, or asymptotic independence, was introduced by Rosenblatt (1956) and Ibragimov (1959),(1962) with the notions of strong and uniform mixing. Let $(\Omega, \mathcal{A}, P)$ be the probability space in which the process x_t is defined, where $\mathcal{A}$ is the smallest Borel field including all the Borel sets of the form $\{\omega | (x_{j(k)}(n_k, \omega), k = 1, \ldots, m) \in B\}$ with B a Borel set in $\mathbb{R}^m$. Let $\mathcal{F}_p$ be the σ-field of events determined by x_t, $t \leq p$ and $\mathcal{F}^q$ be the σ-field of events determined by x_t, $t \geq p$. Define the sequences

$$\rho_x(k) := \sup\{|P(A \cap B) - P(A)P(B)|, \quad A \in \mathcal{F}_p, \quad B \in \mathcal{F}^{p+k}, \quad k \geq 1\} \tag{1.3}$$

and

$$\zeta_x(k) := \sup\{|\frac{P(A \cap B)}{P(A)} - P(B)|, \quad A \in \mathcal{F}_p, \quad B \in \mathcal{F}^{p+k}, \quad k \geq 1\}. \tag{1.4}$$

$\rho_x(k)$ was introduced by Rosenblatt (1956) and called α-mixing or strong-mixing sequence and $\zeta_x(k)$ was introduced by Ibragimov (1962) and called ϕ-mixing or uniform-mixing sequence (The usual α and ϕ notations are replaced by ρ and ζ respectively to avoid a clash of notation with what follows). The strictly stationary process x_t is called strong (resp. uniform) mixing if $\rho_x(k) \to 0$ (resp. $\zeta_x(k) \to 0$) when $k \to \infty$. It is easily seen that $2\rho_x(k) \leq \zeta_x(k)$ and therefore that uniform-mixing implies strong-mixing although the converse is not always true (see for instance Doukhan (1995) §1.3.2). Žurbenko (1986) gives an extensive discussion of these and other definitions of mixing behaviour. Alternatively, Brillinger (1975) applies summability conditions on cumulant moments of all orders, which are assumed to exist. Defining the kth order cumulant of a strictly stationary process by

$$\mathrm{cum}(X_1,\ldots,X_k) = \sum(-1)^p(p-1)!(E\prod_{j\in\nu_1} X_j)\ldots(E\prod_{j\in\nu_p} X_j) \tag{1.5}$$

where the summation extends over all partitions $(\nu_1,\ldots,\nu_p)$, $p = 1,\ldots,k$, and defining

$$c_{t_1,\ldots,t_{k-1}} = \mathrm{cum}(x_t, x_{t+t_1},\ldots,x_{t+t_{k-1}}), \quad \text{for} \quad t_1,\ldots,t_{k-1} = 0 \pm 1, \ldots\ , \tag{1.6}$$

Brillinger (1975) introduces a mixing condition of the form

$$\sum_{t_1,\ldots,t_{k-1}=-\infty}^{+\infty} |c_{t_1,\ldots,t_{k-1}}| < \infty \quad \text{for all} \quad k \geq 2. \tag{1.7}$$

It is easily seen that this condition includes absolute summability of autocovariances of the process, a condition which restricts the choice of filtering weights on the innovations ε_t in 1.1, because it is equivalent to

$$\sum_{l=-\infty}^{+\infty} |\sum_{j=0}^{\infty} \alpha_j\alpha_{j+l}| < \infty, \tag{1.8}$$

with the convention $\alpha_j = 0$, $j < 0$.

However, it is clear that specification 1.1-1.2 derived from the Wold Decomposition Theorem allows for much more far reaching patterns of correlations between observations, including the possibility of greater dependence between distant x's, and indeed, does not impose absolute summability of autocorrelations, or any condition of the mixing type. Time series processes with non summable autocovariances and, correspondingly, slow decay of Wold filtering weights, have been the focus of considerable attention, particularly recently, with the development of parsimonious parametric representations of long range temporal dependence and of robust semiparametric estimation methods. The failure of long range uncorrelatedness as highlighted by the nonsummability of autocovariances was first investigated in the field of Hydrology. Joseph's famous prophesy (Genesis 41, 29-30) promising Pharaoh seven years of abundance followed by seven years of famine in Egypt gives a speculative account of the long periods of drought followed by long periods of recurrent flooding of alluvial plains by the Nile river bringing prosperity to the region. A more definite account is provided by the particularly reliable measurements of the annual low levels of the Nile at the Ghoda Range collected between A.D. 622 and A.D. 1284 and appearing in Toussoun (1925) (the first missing observation is for year A.D. 1285 outside the sample chosen). Two characteristics of this series are consistent with non mixing behaviour: slow decay of sample autocorrelations and a sample mean with variance which decays at a markedly slower rate than n^{-1} (for graphical assessments, see, e.g. Beran (1994) p. 22). Hurst (1951) gives a quantitative account of a phenomenon later named after him "Hurst effect" together with a heuristic approach to the measurement of the degree of temporal dependence associated with this effect. He defined the rescaled adjusted range or R/S statistic which is the standardised ideal capacity of a reservoir between a time origin and time T, and he observed a pattern consistent with the relation

$$E[R/S] \sim cT^H \quad \text{as} \quad T \to \infty \quad \text{with} \quad H > \frac{1}{2} \tag{1.9}$$

where "$\sim$" indicates that the ratio of the left hand side and the right hand side tends to one when T tends to infinity, whereas H (often called the self-similarity parameter) should be equal to $\frac{1}{2}$ if the river flow behaved like a process under any type of mixing assumption. In his pioneering work on stock prices, Mandelbrot (1973) identified the same type of phenomenon and related it to the self-similarity distributional property introduced by Kolmogorov (1940), by which the joint distribution of $x_{t_1}, \ldots, x_{t_n}$ is identical to a^{-H} times the joint distribution of $x_{at_1}, \ldots, x_{at_n}$ for any $a > 0$, with the introduction of fractional Gaussian noise (in Mandelbrot and Ness (1968)), a Gaussian process with zero mean and autocovariances following

$$\mathrm{cov}(x_1, x_{1+j}) = \frac{1}{2}\mathrm{var}(x_1)\{|j+1|^{2H} - 2|j|^{2H} + |j-1|^{2H}\}, \quad \text{for} \quad j = 0, \pm 1, .. \tag{1.10}$$

Mandelbrot (1972) proposed the use of the latter model and the R/S statistic in the investigation of economic data supporting thereby Granger's view on *the typical spectral shape of an economic variable*, expounded in Granger (1966) referring to the recurrent pattern of estimated spectral densities with a peak at zero frequency (corresponding to a long range of temporal dependence in the process) and decreasing at higher frequencies. As Robinson (1994d) points out, "*such behaviour could* *be consistent with the presence of one or more unit roots, but the same series also displayed a tendency for the spectrum of first differences to exhibit a trough at zero frequency...*". The large proportion of variance concentrated around zero frequency is indeed a translation in the frequency domain of non summability of autocovariances, but it need not indicate non stationarity, let alone the presence of a unit root.

The most popular model, besides fractional Gaussian noise in 1.10, which encompasses this distinction, is the autoregressive fractionally integrated moving average model, where

$$(1-L)^{d_x} b(L)x_t = a(L)\varepsilon_t, \quad \text{with} \quad -\frac{1}{2} < d_x < \frac{1}{2}, \tag{1.11}$$

where $a(z)$ and $b(z)$ are both finite order polynomials with zeros outside the unit circle in the complex plane. This model was proposed by Adenstedt (1974), Hosking (1981) and Granger and Joyeux (1980). $(1-L)^d$ has a binomial expansion which is conveniently expressed in terms of the hypergeometric function

$$(1-L)^d = F(-d,1,1;L) = \sum_{k=0}^{\infty} \Gamma(k-d)\Gamma(k+1)^{-1}\Gamma(-d)^{-1}L^k \tag{1.12}$$

where $\Gamma(.)$ denotes the Gamma function. Writing

$$\alpha(z) = \sum_{j=0}^{\infty} \alpha_j z^j, \tag{1.13}$$

1.11 corresponds to a parametric representation nested in specification 1.1-1.2 with

$$\alpha(z) = (1-z)^{-d_x} \frac{a(z)}{b(z)}, \tag{1.14}$$

and when $\frac{1}{2} > d_x > 0$, this implies a slow decay of filtering weights

$$\alpha_j = O(j^{d_x - 1}) \quad \text{as} \quad j \to \infty \tag{1.15}$$

and of autocorrelations

$$\frac{\sum_{i=0}^{\infty} \alpha_i \alpha_{i+j}}{\sum_{i=0}^{\infty} \alpha_i^2} \sim c j^{2d_x - 1} \quad \text{as} \quad j \to \infty \tag{1.16}$$

consistent with the Hurst effect. Note that here, and in all that follows, $H - 1/2$ will be denoted d_x and will be called *long memory in the levels* because it measures the impulse response of the mean of the process.

Hyperbolic decay of the filtering weights as in 1.15 implies that the latter decay so slowly as to be non summable even though they remain square summable for $d_x < \frac{1}{2}$. In terms of impulse response, this implies a very long lived but non permanent response to shocks at any time in the series, consistent with very slow mean reversion in the process (a notion widely used in the economic literature, and which can be formally defined within this framework by a d_x value strictly below one, distinguishing it thereby from a stationarity concept). Hyperbolic decay of autocorrelations translates into the existence of a singularity of hyperbolic nature in the spectral density of the process in the neighbourhood of frequency zero. Conditions for equivalence between time domain and frequency domain representations of long memory are discussed in Yong (1974). Therefore, throughout this work, long memory in a weakly stationary time series x_t, $t = 0, \pm 1, \ldots$, with autocovariances satisfying

$$\operatorname{cov}(x_t, x_{t+j}) = \int_{-\pi}^{\pi} f(\lambda) \cos(j\lambda) d\lambda \quad j = 0, \pm 1, \ldots \ , \tag{1.17}$$

will be modelled semiparametrically by

$$f(\lambda) \sim L(\lambda)\lambda^{-2d_x} \quad \text{as} \quad \lambda \to 0^+, \quad \text{with} \quad -\frac{1}{2} < d_x < \frac{1}{2}, \tag{1.18}$$

where $L(\lambda) > 0$ and is continuous at $\lambda = 0$ when $d_x = 0$, and is otherwise a slowly varying function at zero defined by

$$\frac{L(t\lambda)}{L(\lambda)} \to 1 \quad \text{as} \quad \lambda \to 0 \quad \text{for all} \quad t > 0. \tag{1.19}$$

Under 1.18, $f(\lambda)$ has a pole at $\lambda = 0$ for $0 < d_x < \frac{1}{2}$ (when there is long memory in x_t), $f(\lambda)$ is positive and finite for d_x (which is identified with short memory in x_t) and $f(0) = 0$ for $-\frac{1}{2} < d_x < 0$ (which can be described as negative dependence or antipersistence, and which is characteristic of the first differences of a process which was mistakenly believed to hold a unit root, but was in fact mean reverting in the sense defined above). Parametric submodels of 1.18 such as autoregressive fractionally integrated moving average or fractional Gaussian noise in the time domain, can also be written in the frequency domain. This is the case with the generalisation of Bloomfield's exponential model (Bloomfield (1973)) proposed by Robinson (1994d), Beran (1993) and Janacek (1993) with an appended quasi maximum likelihood estimation method for the degree of dependence. The spectral density in the most common version of this model is given by

$$f(\lambda) = |1 - e^{i\lambda}|^{-2d_x} \exp\{\sum_{j=0}^{p} \eta_j \cos j\lambda\} \tag{1.20}$$

where the short memory part of the representation provides an arbitrarily accurate approximation of any positive function with a Fourier decomposition.

1.2 Sample mean of long memory processes

Long range dependence may also be detected through the behaviour of the sample mean of the process. Consider the partial sums

$$S_n = \sum_{t=1}^{n} x_t, \tag{1.21}$$

with variance $\sigma_n^2 = E|S_n|^2$, and suppose $Ex_t = 0$ without loss of generality. As noted by Robinson (1994d), σ_n^2 always exists and is equal to $2\pi n$ times the Cesaro sum, to $n-1$ terms, of the Fourier series of $f(\lambda)$, spectral density of the process x_t. Therefore, if f is continuous at the origin,

$$\frac{\sigma_n^2}{n} \rightarrow 2\pi f(0) \quad \text{as } n \rightarrow \infty.$$

Thus, if $f(0) \neq 0$ is estimated consistently by $\hat{f}(0)$, the Central Limit Theorem and Slutzky's Theorem yield

$$S_n(2\pi n \hat{f}(0))^{-\frac{1}{2}} \rightarrow_d N(0,1) \quad \text{as} \quad n \rightarrow \infty \tag{1.22}$$

See Hannan (1979) for a proof of 1.22 for a stationary x_t following 1.1 with i.i.d. innovations ε_t (this can be extended to conditionally homoscedastic and uniformly integrable martingale differences) and

$$\sum_{j=0}^{\infty} |\alpha_j| < \infty. \tag{1.23}$$

Eicker (1956) gives consistent estimates of the limiting variance of the LSE and other simple estimates of parametric models in the presence of parametric and nonparametric disturbance autocorrelation. 1.22 implies a convergence of the sample mean at the rate $n^{-\frac{1}{2}}$, a feature which generally fails when 1.23 does not hold. When the filtering weights α_j are not absolutely summable, but follow 1.15, with long memory parameter $d_x > 0$, the σ_n^2 need to be scaled with a factor $n^{-(2d_x+1)}$ instead of n^{-1} to converge. Moreover, the sample mean is no longer BLUE (Adenstedt (1974)). Samarov and Taqqu (1988) found that efficiency can be poor for $d < 0$ but is at least 0.98 for $d > 0$. Providing σ_n^2 does not have a finite limit, a central limit theorem continues to hold when x_t follows 1.1 with i.i.d. innovations (this can be relaxed to martingale difference innovations) in the form:

$$\sigma_n^{-1} S_n \to_d N(0,1) \quad \text{with} \quad n^{\frac{1}{2}+d_x}\sigma_n^{-1} \to c > 0, \tag{1.24}$$

in Ibragimov and Linnik (1971). Giraitis and Surgailis (1986) use the Appell generalisation of Hermite polynomials to extend the result 1.24 to nonlinear functions of processes satisfying 1.1 with i.i.d. innovations ε_t.

1.3 Conditional heteroscedasticity

Long memory therefore provides a framework for a very parsimonious representation of temporal dependence, in that the long range dependence is embodied in the one parameter d_x. To derive asymptotic distributional results for processes with strong temporal dependence typically outside the scope of any mixing assumption, the approach chosen here relies on a martingale difference or non predictability assumption for the Wold innovations ε_t of the process. In other words, letting $\mathcal{F}_t$ denote the filtration associated with the σ-field of events generated by $(\varepsilon_s, s \leq t)$, one needs to assume the innovations are *martingale differences*:

$$E(\varepsilon_t|\mathcal{F}_{t-1}) = 0 \quad \text{almost surely (a.s.).} \tag{1.25}$$

However, limit results for generalized linear processes under 1.25, such as Hannan

(1979)'s, require the assumption of constant conditional variance

$$E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2 \quad \text{a.s.} \tag{1.26}$$

that many time series, in particular long financial time series where a large degree of temporal dependence is (and can be) observed, are generally believed to violate. Financial returns, constructed from first differenced logged asset prices or foreign exchange bank quote midpoints sampled at weekly, daily or intra-daily frequencies, typically exhibit thick-tailed distributions and volatility clustering, i.e. conditional variances changing over time in such a way that periods of high movement are followed by periods displaying the same characteristic, and periods of low movement also. One therefore needs to allow for time varying volatilities for the innovations, and 1.26 needs to be replaced by

$$E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2 \quad \text{a.s.} \tag{1.27}$$

where σ_t^2 is a stochastic process whose temporal dependence properties can in turn be considered. The conditional variance σ_t^2 can be allowed to depend on some latent structure, as in the model due to Taylor (1980):

$$\begin{aligned}
&\varepsilon_t = \eta_t \sigma_t, \\
&\log \sigma_t = \gamma_0 + \gamma_1 \log \sigma_{t-1} + u_t, \\
&\eta_t, u_t \quad \text{independent i.i.d.}.
\end{aligned}$$

In this model, the latent variable σ_t can be construed as embodying the flow of heterogeneous information arrivals on the market, as in the work of Clark (1973). The latent variable σ_t can also be allowed to depend on the lagged values of the innovations. This approach was chosen by Engle (1982) in a form he called autoregressive conditional heteroscedasticity (ARCH), and generalised by Bollerslev (1986) who introduced lagged values of σ_t^2 thereby introducing a latent ARMA structure for the squared innovations (the GARCH model). A model ensuring the positivity of σ_t^2 and producing skewed conditional distributions is the exponential generalised autoregressive conditional heteroscedasticity model (EGARCH) proposed by Nelson

(1991), and some nonlinearities were introduced by Sentana (1995) with an extensive study of quadratic ARCH models and by Zakoïan (1995) with the threshold ARCH class of models. An extensive review of the literature in this field of econometric research is given by Bollerslev, Engle, and Nelson (1994). All of the above are based on a parameterisation of the one step-ahead forecast density, a particularly appealing feature -as pointed out by Shephard (1996)- as much of finance theory is concerned with one step-ahead moments or distributions defined with respect to the economic agent's information. Asymptotic theory for parametric ARCH modelling was proposed by Weiss (1986), Lee and Hansen (1994) Lumsdaine (1996) and Newey and Steigerwald (1994). Bollerslev, Chou, and Kroner (1992) give reviews of the GARCH modelling approach. A nonparametric specification encompassing both ARCH and GARCH as special cases was proposed by Robinson (1991b) where σ_t^2 is an infinite sum of lagged values of ε_t^2:

$$\sigma_t^2 = \sigma^2 + \sum_{j=1}^{\infty} \psi_j (\varepsilon_{t-j}^2 - \sigma^2) \quad \text{a.s.} \quad \text{with} \quad \sum_{j=1}^{\infty} \psi_j^2 < \infty. \tag{1.28}$$

This can be reparameterised as

$$\sigma_t^2 = \beta + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j}^2 \tag{1.29}$$

and includes both standard ARCH (when $\psi_j = 0,\ j > p$, for finite p) and GARCH (for which the ψ_j decay exponentially) models. However, as Robinson (1991b) indicated, long memory behaviour is also covered. This, and the semi-strong ARMA representation for the squared innovations implied by the above specification, is made apparent in the following reparameterisation. If, for complex valued z,

$$\psi(z) = 1 - \sum_{j=1}^{\infty} \psi_j z^j \tag{1.30}$$

satisfies

$$|\psi(z)| \neq 0, \ |z| \leq 1, \tag{1.31}$$

define

$$\phi(z) = \sum_{j=0}^{\infty} \phi_j z^j = \psi(z)^{-1}, \ \phi_0 = 1. \tag{1.32}$$

Then, Robinson (1991b) rewrote 1.28 as

$$\varepsilon_t^2 - \sigma^2 = \sum_{j=0}^{\infty} \phi_j \nu_{t-j}, \tag{1.33}$$

where

$$\nu_t = \varepsilon_t^2 - \sigma_t^2 \tag{1.34}$$

satisfies

$$E(\nu_t | \mathcal{F}_{t-1}) = 0 \quad \text{a.s.}, \tag{1.35}$$

by construction. As a result, the chosen specification does not include all weak GARCH processes as defined by Drost and Nijman (1993) as processes with the same linear projections as ordinary GARCH. However, as for weak ARMA processes, limiting distribution theory for weak GARCH processes, provided, for instance, by Francq and Zakoïan (1997), relies on mixing assumptions which may preclude the high levels of temporal dependence in the squares which are allowed by 1.33 with a suitable choice of filtering weights. To allow for specific types of nonlinearities in the squares, Robinson (1991b) also proposed a quadratic version of 1.28:

$$\sigma_t^2 = \Big(\sigma + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j}\Big)^2 \quad \text{a.s.} \tag{1.36}$$

which endogenises the positivity constraint on σ_t^2. It is apparent in the empirical literature on ARCH modelling of financial series, surveyed by Bollerslev, Chou, and Kroner (1992), that the degree of dependence in second moments is too large to be modelled in terms of mixing behaviour or with a latent stationary ARMA structure (and therefore exponentially decaying weights). Parameter estimates from GARCH(1,1) models on asset returns and foreign exchange, the most popular modelling technique, lie close to the boundary of stationarity for the process, prompting the introduction of a unit root in the autoregressive moving average equation describing the behaviour of the squares. Lumsdaine (1996) shows asymptotic normality of the quasi-maximum likelihood estimator in the integrated GARCH(1,1) model. However, the IGARCH model implies full persistence of shocks on the variance in a sense defined by Bollerslev and Engle (1993). According to Nelson's distinction

(Nelson (1990b)), it corresponds to persistence of shocks on both forecast moments of σ_t^2 and on forecast distributions of σ_t^2. A long memory representation of volatility, replacing for instance the unit root by a fractional filter in the equation for the squares, reconciles a high degree of temporal dependence in volatilities with lack of persistence and, possibly, with covariance stationarity. Denoting $s_t = \sigma_t^2 - \sigma^2$ and $\chi_t = \varepsilon_t^2 - \sigma^2$, for $l > 0$, we have

$$s_{t+l} = \psi_l \chi_t + \sum_{j \neq l}^{\infty} \psi_j \chi_{t-j+l} \quad \text{a.s.}\,. \tag{1.37}$$

Now as under 1.28, $\psi_l \chi_t \to 0$ almost surely when $l \to \infty$, χ is persistent in the volatility according to none of the definitions adopted by Nelson (1990b), i.e. persistence in probability, in L_p-norm or almost surely.

Besides, the analogy is apparent between the clustering of volatilities of financial returns and what Mandelbrot (1973) described as "Joseph effect". And, effectively, Whistler (1990), Lo (1991), Ding, Granger, and Engle (1993) and Lee and Robinson (1996), are among the first to show how well the long memory representation performs empirically. A general fractionally integrated GARCH model is obtained as a special case of specification 1.28 with the $\phi(z)$ polynomial defined as

$$\phi(z) = (1-z)^{-d_\varepsilon} \frac{b(z)}{a(z)} \tag{1.38}$$

for $0 < d_\varepsilon < \frac{1}{2}$ and finite order polynomials $a(z)$ and $b(z)$ whose zeros lie outside the unit circle in the complex plane. Note that the degree of fractional integration is called d_ε in this case to distinguish long memory in the squared innovations from long memory in the levels. Baillie, Bollerslev, and Mikkelsen (1996) apply 1.38 to asset prices with the addition of a drift parameter

$$(1-L)^{d_\varepsilon} a(L) \varepsilon_t^2 = \mu + b(L)\nu_t. \tag{1.39}$$

Nelson (1990b) proves almost sure convergence of the conditional variance σ_t^2 in the short memory case d_ε in 1.38 with $a(z)$ and $b(z)$ of degree one. Apart from 1.38, the requirement

$$0 < \sum_{j=0}^{\infty} \phi_j^2 < \infty \tag{1.40}$$

includes the other traditional long memory specification of moving average coefficients, the fractional noise case with autocorrelations satisfying

$$\operatorname{corr}\left(\varepsilon_t^2, \varepsilon_{t+j}^2\right) = \frac{\sum_{i=0}^{\infty} \phi_i \phi_{i+j}}{\sum_{i=0}^{\infty} \phi_i^2} = \frac{1}{2}\left\{|j-1|^{2d+1} - |j|^{2d+1} + |j+1|^{2d+1}\right\}. \quad (1.41)$$

Robinson (1991b) developed Lagrange multiplier tests for no-ARCH against alternatives consisting of general finite parameterisation of 1.28, specialising to 1.38 and 1.41. In both these cases, the autoregressive weights ψ_j satisfy

$$\sum_{j=1}^{\infty} |\psi_j| < \infty \quad (1.42)$$

Under 1.42 and

$$\max_t E\left(\varepsilon_t^4\right) < \infty, \quad (1.43)$$

it follows that

$$\begin{aligned} E\left(\nu_t^2\right) &\leq E\Big\{\sum_{j=0}^{\infty} \psi_j(\varepsilon_{t-j}^2 - \sigma^2)\Big\}^2 \\ &\leq \Big(\sum_{j=0}^{\infty} |\psi_j|\Big)\Big(\sum_{j=0}^{\infty} |\psi_j| E(\varepsilon_{t-j}^4)\Big) \\ &\leq K \end{aligned} \quad (1.44)$$

where K is a generic constant, so the innovations ν_t in 1.33 are square integrable martingale differences, ε_t^2 is well defined as a covariance stationary process and its autocorrelations can exhibit the usual long memory structure implied by 1.38 or 1.41. Even if 1.43 does not hold, the "autocorrelations" $\sum_{i=0}^{\infty} \phi_i \phi_{i+j} / \sum_{i=0}^{\infty} \phi_i^2$ are well defined under 1.40. Both parametric representations 1.38 and 1.41 have the implication that autocorrelations follow

$$\frac{\sum_{i=0}^{\infty} \phi_i \phi_{i+j}}{\sum_{i=0}^{\infty} \phi_i^2} \sim c j^{2d_\varepsilon - 1} \quad \text{as } j \to \infty \quad (1.45)$$

which in turn implies a rate of decay for the innovations filtering weights of

$$\phi_j = O(j^{d_\varepsilon - 1}) \quad \text{as } j \to \infty. \quad (1.46)$$

This is taken as a characterisation of long memory in the process ε_t^2 when $d_\varepsilon > 0$ and it implies nonsummability of weights ϕ_j and autocovariances

$$\gamma_j = \operatorname{cov}(\varepsilon_t^2, \varepsilon_{t+j}^2). \quad (1.47)$$

The rate of convergence of the sample mean is also characteristic of long memory processes when ϕ_j satisfies 1.46. Indeed, 1.35 and 1.44 imply that the partial sums of the squared innovations have variance

$$\begin{aligned} \mathrm{Var}[\sum_{t=1}^{n}(\varepsilon_t^2-\sigma^2)] &= \sum_{s,t=1}^{n}\sum_{j=0}^{\infty}\phi_j\phi_{j+s-t}E(\nu_{t-j}^2) && (1.48)\\ &= O(n\sum_{t=1}^{n}\Theta_t) && (1.49)\end{aligned}$$

with

$$\Theta_t := \sum_{j=0}^{\infty}|\phi_j\phi_{j+t}| = O(t^{2d_\epsilon-1}) \tag{1.50}$$

under 1.46. Therefore, we have the same rate upper bound as in 1.24, i.e.

$$\sum_{t=1}^{n}(\varepsilon_t^2-\sigma^2) = O_p(n^{d_\epsilon+\frac{1}{2}}). \tag{1.51}$$

This result, and nonsummability of the ϕ_j's, is to be contrasted with standard latent ARMA representations for the squares, where weights decay exponentially and are, therefore, absolutely summable. In view of the empirical evidence and the focus on possible long memory in financial returns x_t, it seems appropriate to allow for possible long memory in the ε_t^2 also. This thesis is concerned with the estimation of temporal dependence structures in a covariance stationary time series x_t via the analysis of its spectral density in a neighbourhood of zero frequency. This concerns the case where x_t displays short memory as well as the cases where x_t displays long memory; and a large part of the results provide asymptotic theory in case the squared innovations possibly exhibit long memory themselves.

1.4 Estimating dependence

Estimating the degree of dependence and carrying out inference on the process x_t may require estimation of the spectrum at zero frequency, when it is continuous or of the slope of the logged spectrum at the origin when the process is strongly autocorrelated.

1.4.1 Parametric estimation of long memory

Take x_t to be a covariance stationary series with mean μ_0 and spectral density $f(\lambda;\theta_0)$, $-\pi < \lambda \leq \pi$, where f is a given function of λ and θ. For a realisation of size n, we consider the discrete Fourier transform

$$w_x(\lambda) = (2\pi n)^{-\frac{1}{2}} \sum_{t=1}^{n} x_t e^{it\lambda} \tag{1.52}$$

and the periodogram

$$I_x(\lambda) = |w_x(\lambda)|^2. \tag{1.53}$$

This statistic was first proposed by Schuster (1898) to investigate hidden periodicities in time series. A useful general result, under various regularity conditions, is the following[1]:

$$\frac{\sqrt{n}}{4\pi} \int_{-\pi}^{\pi} \zeta(\lambda) \left[I_x(\lambda) - f(\lambda;\theta_0)\right] d\lambda \rightarrow_d N(0, A(\zeta) + B(\zeta)) \quad \text{as } n \rightarrow \infty, \tag{1.54}$$

where

$$A(\zeta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \zeta^2(\lambda) f^2(\lambda;\theta_0) d\lambda, \tag{1.55}$$

$$B(\zeta) = \frac{1}{8\pi} \int_{-\pi}^{\pi} \zeta(\lambda)\zeta(\mu) f_4(\lambda, -\mu, \mu;\theta_0) d\lambda d\mu, \tag{1.56}$$

and where

$$f_4(\lambda, \mu, \eta) = (2\pi)^{-3} \sum_{u,v,w} c_{u,v,w} e^{-i(u\lambda + v\mu + w\eta)} \tag{1.57}$$

is the fourth order cumulant spectral density, and

$$c_{u,v,w} = \mathrm{cum}(x_t, x_{t+u}, x_{t+v}, x_{t+w}) \tag{1.58}$$

is the fourth order cumulant moment of the process x_t.

[1] In case x_t are residuals from a fitted parametric model, taking $\mu_0 = 0$ does not, under regularity conditions (including conditions on the function $\zeta(\lambda)$), affect the asymptotic properties of the estimates of θ_0 considered below.

When x_t is Gaussian, $B(\zeta)$ vanishes, and under suitable regularity conditions on $\zeta(\lambda)$ and $f(\lambda;\theta)$, Fox and Taqqu(1986) show that 1.54 holds even when x_t is strongly autocorrelated, providing a pole of $f(\lambda;\theta)$ is matched by a zero of $\zeta(\lambda)$ of suitable order. Fox and Taqqu (1986) showed as a result that Whittle's estimate of θ_0 (Whittle (1962), Hannan (1973)), i.e. an estimate resulting from the minimization of

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\log f(\lambda;\theta)+\frac{I_x(\lambda)}{f(\lambda;\theta)}\right\}d\lambda,$$

is asymptotically normal with rate $n^{-1/2}$. Beran (1986) and Dahlhaus (1989) extend this result to prove that the Gaussian maximum likelihood estimate of θ_0 remains efficient in the Cramer-Rao sense when x_t is strongly autocorrelated. Robinson (1994d) shows that root-n asymptotic normality results also apply to the estimate resulting from the maximization of a discretised version of the Whittle likelihood

$$-\frac{1}{n}\sum_{j=1}^{n}\left\{\log f(\lambda_j;\theta)+\frac{I_x(\lambda_j)}{f(\lambda_j;\theta)}\right\} \tag{1.59}$$

where $\lambda_j = 2\pi j/n$ are the harmonic frequencies.

When x_t is possibly non Gaussian, the three estimates above (pseudo maximum likelihood, Whittle and discretized Whittle) continue to be root-n consistent and asymptotically normal under conditions involving weak autocorrelation (see Mann and Wald (1943), Whittle (1962), Hannan (1973) and Robinson (1978a)). Solo (1989) shows 1.54 for x_t satisfying 1.1 with 1.25 and restrictions on the Wold coefficients which include many strongly autocorrelated non Gaussian processes. In case x_t is non Gaussian, $B(\zeta)$ does not vanish in general. An important case where this occurs is when x_t follows 1.1 and 1.25 with dynamic conditional heteroscedasticity in the innovations as discussed in the previous section. In that case, $f_4(\lambda,-\mu,\mu)$ contains contributions from fourth cumulant moments of the innovations ε_t other than $\kappa = \mathrm{cum}(\varepsilon_t,\varepsilon_t,\varepsilon_t,\varepsilon_t)$. Under 1.26, which imposes constant second and fourth conditional moments, we have

$$\mathrm{cum}(\varepsilon_r,\varepsilon_s,\varepsilon_t,\varepsilon_u) \;=\; \kappa \quad \text{if} \quad r=s=t=u, \tag{1.60}$$

and zero otherwise. Under 1.27, with σ_t^2 defined by 1.28, however,

$$\begin{aligned}
\mathrm{cum}(\varepsilon_r, \varepsilon_s, \varepsilon_t, \varepsilon_u) &= \kappa \quad \text{if} \quad r = s = t = u, && (1.61)\\
&= \gamma_{r-s} \quad \text{if} \quad r = t \neq s = u, && (1.62)\\
&= \gamma_{r-t} \quad \text{if} \quad r = s \neq t = u, && (1.63)\\
&= \gamma_{r-s} \quad \text{if} \quad r = u \neq t = s, && (1.64)
\end{aligned}$$

and zero otherwise. Therefore, the fourth cumulant 1.58 is equal to

$$\begin{aligned}
c_{u,v,w} &= \kappa \sum_{k=0}^{\infty} \alpha_k \alpha_{k+u} \alpha_{k+v} \alpha_{k+w} && (1.65)\\
&+ \sum_{k \neq j} \gamma_{k-j} [\alpha_j (\alpha_{j+u} \alpha_{k+v} \alpha_{k+w} + \alpha_{j+v} \alpha_{k+u} \alpha_{k+w} \\
&\qquad + \alpha_{j+w} \alpha_{k+u} \alpha_{k+v})] && (1.66)
\end{aligned}$$

and zero otherwise. The ARCH special case of 1.27 was considered by Weiss (1986), and the GARCH(1,1) by Lee and Hansen (1994). Both show asymptotic normality of the quasi maximum likelihood. Lumsdaine (1996) allows for nonstationarity of the integrated form in the conditional variance equation, but long term dependence is not covered for x_t.

1.4.2 Smoothed periodogram spectral estimation

Semiparametric alternatives in the estimation of the slope of the logged spectrum at the origin rely on specification 1.18. Local specification around the frequency of interest avoids the pitfall of parametric estimation of the long memory parameter d_x: a mispecified spectrum at non zero frequencies may cause inconsistency in estimation of the long memory parameter (characterizing the low frequency dynamics of the system). This type of estimation is based on low frequency harmonics of the periodogram 1.53 whose properties are briefly discussed in this paragraph.

The periodogram is an asymptotically unbiased estimate for the spectral density, at continuity points of the latter. However, as Grenander (1953) first showed, the variance of the periodogram does not vanish with sample size, so that the periodogram is not a consistent estimate for the spectral density. More precisely, consider a stationary process x_t following 1.1 with i.i.d. innovations ε_t and filtering weights α_j satisfying

$$\sum_{j=0}^{\infty} j^{\frac{1}{2}} |\alpha_j| < \infty, \tag{1.67}$$

and with spectral density defined as in 1.17. The periodogram of x_t has the following asymptotic sampling properties:

$$\mathrm{cov}(I_x(\lambda), I_x(\mu)) = (1 + \delta_{0\lambda} + \delta_{\pi\lambda})\delta_{\lambda\mu}(f_x(\lambda)^2 + O(n^{-\frac{1}{2}})) + O(n^{-1}), \tag{1.68}$$

where δ stands for the Kronecker symbol. A proof of this result is given in Brockwell and Davis (1991). The same result is shown to hold by Brillinger (1975) for a strictly stationary process satisfying the mixing condition 1.7. 1.68 shows not only that periodogram ordinates are not mean-square consistent, but also that at distinct frequencies they are asymptotically uncorrelated under these conditions, which permits the construction of consistent estimates of the spectral density such as

$$\tilde{f}(\lambda) = \sum_{|j| \le m} W_n(j) I_x(\lambda + \frac{2\pi j}{n})$$

where m is a bandwidth sequence satisfying at least

$$\frac{1}{m} + \frac{m}{n} \to 0 \quad \text{as} \quad n \to \infty, \tag{1.69}$$

and $W_n(j)$ is a sequence of symmetric weight functions satisfying

$$\sum_{|j| \le m} W_n(j) = 1 \quad \text{and} \quad \sum_{|j| \le m} W_n^2(j) \to 0 \quad \text{as} \quad n \to \infty. \tag{1.70}$$

Under 1.67 and 1.43, we have (see Brockwell and Davis (1991) for a proof)

$$\lim_{n \to \infty} \left(\sum_{|j| \le m} W_n^2(j) \right)^{-1} \mathrm{cov}(\tilde{f}(\lambda), \tilde{f}(\mu)) = (1 + \delta_{0\lambda} + \delta_{\pi\lambda})\delta_{\lambda\mu} f_x(\lambda)^2,$$

so that, in view of 1.70, $\mathrm{var}(\tilde{f}(\lambda))$ shrinks to zero with sample size and $\tilde{f}(\lambda)$ is a consistent estimate for $f(\lambda)$. Bartlett (1950) and Tukey (1950) propose estimates of the spectral density based on earlier versions of the asymptotic uncorrelatedness property as discussed in Grenander and Rosenblatt (1966).

1.4.3 Semiparametric estimation of long memory

The estimation strategy based on low frequency periodogram ordinates which is considered in this work is related to a strategy first proposed by Hill (1975) in tail estimation for distributions with a high degree of leptokurtosis. Research in that field was fuelled in the last couple of years by institutional regulations allowing banks to derive their own method of estimation for the probability of extreme losses. Hill's semiparametric approach to the estimation of the tail of distributions relies on a parametric specification of the tail of the distribution and a nonparametric treatment of the rest of the distribution. The probability distribution is said to feature a heavy tail if it behaves asymptotically like the Pareto distribution

$$P(Y > y) = y^{-\gamma} L(y), \quad \gamma > 0, \quad y > 1, \tag{1.71}$$

where $L(y)$ is a slowly varying function at infinity. The Hill estimate for the "tail index" γ (which, similarly to the long memory parameter in the time domain or the frequency domain representations, appears as an exponent) maximises a conditional Pareto likelihood

$$\frac{1}{\hat{\gamma}} = \frac{1}{m} \sum_{j=1}^{m} \log\Big(\frac{Y_{(j)}}{Y_{(m+1)}}\Big) \tag{1.72}$$

where $Y_{(1)} \geq \ldots \geq Y_{(n)}$ are the order statistics of a sample of observations $Y_1, \ldots, Y_n$ and m is the number of statistics used in the estimation, satisfying 1.69.

Now suppose x_t is weakly stationary and follows 1.18. One wishes to estimate the degree of long memory d_x in a way that is robust to possible mispecification of the short range dynamics. A semiparametric estimate of d_x was proposed by Künsch (1987) and will be dwelt upon in chapter 3 of this thesis. It is based on the Whittle likelihood discretisation 1.59, but the optimisation is realised over the first m frequencies only, with 1.69, in accordance with the local specification 1.18. The function to minimise is

$$Q_n(d_x) = \frac{1}{m} \sum_{j=1}^{m} \left\{ \log G\lambda_j^{-2d_x} + \frac{I_x(\lambda_j)}{G\lambda_j^{-2d_x}} \right\} \tag{1.73}$$

noting that $f(\lambda)$ is replaced by its local parametric form in the neighbourhood of zero frequency 1.18 with $0 < L(\lambda) = G < \infty$. The estimate is not defined in closed form, so that a prior consistency result (as in Robinson (1995a) under very weak local smoothness conditions for the spectral density, in addition to assumption 1.26 is necessary. Robinson (1995a) also proves asymptotic normality under

$$f(\lambda) = G\lambda^{-2d_x}(1 + O(\lambda^{\beta})) \quad \text{as} \quad n \to \infty, \tag{1.74}$$

for some $\beta \in [0, 2)$, under slightly stronger local smoothness conditions and finite fourth moments for x_t still satisfying 1.26. The proof of the result

$$\sqrt{m}(\hat{d}_x - d_x) \to_d N(0, \frac{1}{4}) \tag{1.75}$$

assumes the following restrictions on the choice of "bandwidth" m,

$$\frac{1}{m} + \frac{m^{2\beta+1}\log^2 m}{n^{2\beta}} \to 0 \quad \text{as} \quad n \to \infty \tag{1.76}$$

which perforce restricts the rate of convergence of the estimate. The latter will therefore be inefficient with respect to correctly specified parametric Whittle estimation, when $m = [(n-1)/2]$, which has $n^{-\frac{1}{2}}$ rate of convergence. Robinson also conjectures that the theorem still holds under the milder and more natural condition

$$\frac{1}{m} + \frac{m^{2\beta+1}}{n^{2\beta}} \to 0 \quad \text{as} \quad n \to \infty. \tag{1.77}$$

An asymptotic normality result for the estimate may still hold when m is of exact order of magnitude $n^{\frac{2\beta}{2\beta+1}}$, corresponding to optimal smoothing. In that case, asymptotic bias will not be zero as in the cases of "oversmoothing" (cases where m is small in order to avoid asymptotic bias): 1.77 and 1.76.

This estimate has the advantage of an asymptotic variance which is free of unknown parameters and which is smaller than for any other known asymptotically normal semiparametric estimate of d_x under 1.18. The latter property may correspond to the minimiser of 1.73 retaining some of the optimality properties of its parametric (full-band) counterpart through optimal weighting of the low-periodogram ordinates. However, to date, no optimality theorem in the Cramer-Rao sense seems to be available for stationary processes with spectral densities satisfying 1.74. Giraitis, Robinson, and Samarov (1997) give a rate optimality theory but no lower bound for the asymptotic variance of estimates achieving that rate. Chapter 4 will be concerned with optimal smoothing and optimal bandwidth selection.

Other estimates of d_x following the same estimation principle are the log periodogram estimate proposed by Geweke and Porter-Hudak (1983), the averaged periodogram estimate proposed by Robinson (1994c) and the exponential estimate proposed by Janacek (1993). The log periodogram is based on a regression of the first m harmonics of the log periodogram on a simple function of frequency. An efficiency improving version of this estimate was proved in Robinson (1995b) to produce a consistent and asymptotically normal estimate of d_x, applying least squares to the regression

$$\log I_x(\lambda_j) = C + d_x(2\log\lambda_j) + U_j, \quad j = l+1, \ldots, m \tag{1.78}$$

where l is a trimming parameter which diverges at a slower rate than the bandwidth m. Hurvich, Deo, and Brodsky (1998) further show that for a slightly more specific local parameterisation, the original Geweke-Porter-Hudak estimate is also asymptotically normal and that no trimming of very low frequency harmonics is necessary. Janacek's estimate (Janacek (1993)) is a counterpart of the log periodogram estimate based on the fractional exponential model 1.20, but, to date, there seems to be no asymptotic theory for it. The averaged periodogram estimate proposed by Robinson (1994c) will be considered in Chapter 2 of this thesis in more depth. It is based on an analogy with the weak dependence case where averaging over approximately independent periodogram harmonics in a neighbourhood of zero frequency produces a consistent estimate of the spectral density at zero frequency. However, the asymptotic properties of low periodogram ordinates are considerably affected by long range dependence, and new results had to be derived. The idea of the log periodogram estimate is drawn from the identity 1.78 under 1.18, where $C = \log L(0) - E$, E is Euler's constant $E = 0.5772\ldots$, and $U_j = \log(L(\lambda_j)/L(0)) + \log(I_x(\lambda_j)/f(\lambda_j)) + E$. The U_j can be thought of as approximately zero mean and i.i.d. in view of the result, derived under a variety of assumptions -all of which include a form of weak dependence-, that for $\lambda \neq 0$ modulo π and J a finite positive integer, $I_x(\lambda_j), j = 1, \ldots, J$ are asymptotically independent $f(\lambda_j)\chi_2^2/2$ variates (see for instance Theorem 5.2.6 page 126 in Brillinger (1975) and Theorem 12 page 223 of Hannan (1970)). Asymptotic properties of the averaged periodogram estimate of the spectral density at zero frequency

$$\hat{f}(0) = \frac{1}{m}\sum_{j=1}^{m} I_x(\lambda_j), \tag{1.79}$$

m being the bandwidth following at least 1.69, and, more generally, of weighted periodogram spectral estimation (in Brillinger (1975)), follow from this asymptotic distributional result for ordinates of the periodogram under short memory.

For a process x_t where the conditional homogeneity condition 1.26 fails (and therefore the conditions applied by Hannan (1970) who assumed the ε_t to be i.i.d.), and is replaced by 1.27 with σ_t^2 defined by 1.28, the asymptotic distributional result for periodogram ordinates may not continue to hold, possibly because of non summable fourth cumulant contributions to asymptotic variances. In Chapter 2, it is proved that in spite of this, 1.79 remains an asymptotically normal estimate of $f(0)$ with a suitable choice of bandwidth m.

When x_t displays long memory (it follows 1.18), the asymptotic distributional result continues to hold for fixed positive frequencies (see Rosenblatt (1981) and Yajima (1989)) but not for periodogram ordinates in a neighbourhood of zero, as documented by Künsch (1996), Hurvich and Beltrao (1993), Comte and Hardouin (1995), and Robinson (1995b). The periodogram ordinates $I_x(\lambda_j)$ are no longer independent or identically distributed when the sample size n tends to infinity. In this setting, Theorem 2 of Robinson (1995b) gives a major result on asymptotic variance and correlations of low frequency periodogram ordinates which applies to the dependence structure considered in this thesis under 1.74: putting $v(\lambda) = w_x(\lambda)/G^{1/2}\lambda^{-d_x}$, where $w_x(\lambda)$ is the discrete Fourier transform defined in 1.52, and σ^2 is the unconditional variance of the innovations to the process, we have

$$E[v(\lambda_j)\bar{v}(\lambda_k)] = \delta_{jk}\sigma^2 + O(\frac{\log j}{k}) \tag{1.80}$$

$$E[v(\lambda_j)v(\lambda_k)] \;=\; O(\frac{\log j}{k}). \tag{1.81}$$

This result is instrumental to the proofs of the asymptotic properties of the log periodogram, the local Whittle and the averaged periodogram estimates of long memory, and it remains valid when the conditional homoscedasticity condition 1.26 is relaxed to 1.27 with σ_t^2 following 1.28. In this setting, Chapter 3 proves that the asymptotic normality result 1.75 continues to hold for the local Whittle estimate of long memory, and that it continues to hold with identical asymptotic variance so that no features of the ARCH structure defined by 1.28 or 1.36 enter. This result is due to additional smoothing of the periodogram via the slightly more stringent condition on the choice of bandwidth

$$m \log m = o(n^{\frac{1}{2}-d_\epsilon}) \quad \text{as} \quad n \to \infty \tag{1.82}$$

which ensures that the contribution to the variance of the periodogram of the errors ε_t from fourth cumulants 1.62-1.64 induced by long memory conditional heteroscedasticity is of small order of magnitude with regards to the suitable approximating martingale. This implicit effect of ARCH -restricting attainable rates of convergence for the estimates- is directly in contrast with parametric or adaptive estimation (see, e.g. Weiss (1986) and Kuersteiner (1997)) where ARCH-type behaviour directly affects limiting distributional properties.

This outcome (i.e. no explicit effect of ARCH) is especially desirable in the case of the local Whittle estimate. This is in the first place due to the simplicity of the limiting variance in 1.75, which is independent of G and d_x. Moreover, although maximum likelihood estimation of parametric versions of 1.33 such as 1.38 or 1.41 is implicit in the derivation of LM tests by Robinson (1991b), no rigorous asymptotic theory exists for such estimates, apart from the ARCH or GARCH special cases studies by Weiss (1986), Lee and Hansen (1994) and Lumsdaine (1996). Thirdly, there is no asymptotic theory available for semiparametric estimation of the memory parameter determining the asymptotic behaviour of the ψ_j's or ϕ_j's in 1.27 or 1.33. Chapter 2 and 3 develop for the first time asymptotic theory in a long memory context that allows for ARCH structure. The mixing conditions stressed above do not permit long memory, whereas long memory literature features either Gaussian processes (e.g. Fox and Taqqu (1986), Robinson (1995b)), non linear functions of Gaussian processes (e.g. Taqqu (1975)), linear functions of independently and identically distributed sequences (e.g. Giraitis and Surgailis (1990)), nonlinear functions of such linear filters ("Appell polynomials", see Giraitis and Surgailis (1986)), as well as the model defined by 1.1, 1.2, 1.25 and 1.26. None of these approaches represents conditional heteroscedasticity in a martingale difference sequence.

1.5 Choice of bandwidth

It is apparent from the discussion above, that the choice of bandwidth m, the number of periodogram ordinates used in the estimation procedure, is crucial in semiparametric estimation of long memory. It is crucial to both asymptotic distributional results and mean square optimality. Moreover, insofar as it determines from which point the practitioner starts to describe the behaviour of the series as asymptotic, bandwidth is central to the concept of long memory itself. In that regard, specifying the series only in the "asymptotic region" with a structure that does not impose itself on short run cycles, seems an intrinsically better approach, notwithstanding considerations of efficiency and robustness.

A discussion of semiparametric estimates of long memory would therefore not be complete without mean-square optimality theory. Bandwidth choice considerably affects kernel density estimates (see for instance Delgado and Robinson (1992), Silverman (1986), Härdle (1990)) and smoothed periodogram spectral estimates (see for instance Robinson (1983a), Robinson (1983b), Robinson (1991a)). Long memory semiparametric estimates are no exception. Henry and Robinson (1996) and Smith and Chen (1996) report extensive Monte Carlo experiments which show the huge variability of bias and variance with bandwidth for the local Whittle estimate of long memory. A similar picture is found in Monte Carlo studies of the log periodogram estimate in Robinson (1995b), Hurvich and Beltrao (1994), Hurvich, Deo, and Brodsky (1998), Taqqu and Teverovsky (1995a), and for the averaged periodogram estimate in Lobato and Robinson (1996), Delgado and Robinson (1996), Delgado and Robinson (1994). The need for an optimality theory for the determination of bandwidth is therefore evident. Giraitis, Robinson, and Samarov (1997) show that for long memory estimates, in a similar way as for smoothed periodogram estimates, one cannot improve on a rate of convergence which depends on the local smoothness properties of the spectral density following specification 1.74. They further show that the log periodogram estimate of long memory in the form proposed by Robinson (1995b) attains this optimal rate of convergence. Under the more restrictive specification

$$f(\lambda) = \left|2\sin\left(\frac{\lambda}{2}\right)\right|^{-2d_r} f^*(\lambda) \tag{1.83}$$

where $f^*(\lambda)$ is twice continuously differentiable and positive at $\lambda = 0$, Hurvich, Deo, and Brodsky (1998) give a precise expression for the mean squared error of the estimate and derive an optimal bandwidth formula. For spectral densities satisfying

$$f(\lambda) = L(\lambda)\lambda^{-2d_r}\left(1 + E_{\beta d_r}\lambda^\beta + o(\lambda^\beta)\right), \quad 0 < \left|E_{\beta d_r}\right| < \infty, \quad \beta \in (0,2], \tag{1.84}$$

as $\lambda \to 0^+$ (with $L(\lambda)$ defined as in 1.18), and defined nonparametrically on the rest of the spectral domain, Henry and Robinson (1996) propose a heuristic derivation of the mean squared error for the local Whittle estimate based on the assumption that the asymptotic variance in 1.75, i.e. $1/4m$, remains the same when the oversmoothing condition 1.77 fails. A full proof requires a treatment of the first two moments of the local Whittle scores and is the matter of further research. As for the averaged periodogram estimate of long memory, Robinson (1994b) provides theoretical values for the mean squared error, and corresponding optimal bandwidth formulae, and Delgado and Robinson (1996) provide an automatic bandwidth selection procedure, including estimation of the constant $E_{\beta d_r}$ in 1.84, and Henry and Robinson (1996) use a similar procedure applied to the local Whittle estimate. However, to date, there seems to be no fully automatic bandwidth selection procedure in the sense that every stage is supported by asymptotic theory. A fully automatic procedure was provided in Danielsson, de Haan, Peng, and de Vries (1997) for the determination of the bandwidth in the Hill estimation procedure of equation 1.72, but the sub-sample bootstrap technique employed relies on the i.i.d. assumption for the observations, and does not seem to be readily extendible to strong dependence. One therefore needs to rely on Monte Carlo experiments to assess the quality of optimal bandwidth selection formulae, and it remains advisable to report a wide range of bandwidth choices in empirical applications.

1.6 Long memory in speculative returns

One of the main areas of application of long memory estimates in economics is the investigation of the behaviour of speculative returns (i.e. first differences of log prices) on financial markets. The theoretical focus of this thesis is particularly well adapted to this analysis. Indeed, financial data on speculative returns sampled at high frequencies (weekly, daily and intra-daily) is now readily available in long and reliable data sets, in particular through data collection agencies such as Bloomberg and Reuters. Nonparametric and semiparametric procedures are well adapted to the analysis of temporal dependence in these long data series. First and foremost, time series methods are instrumental to the investigation of market efficiency. Market efficiency is defined broadly by Fama (1991) as the characteristic of a market with risk-neutral and rational agents where "*prices fully reflect all available information*". Fama (1991) draws distinctions between three types of tests for the Efficient Markets Hypothesis: "*tests for return predictability*", relating to the ability to forecast future returns from the knowledge of past returns, "*event studies*", or microstructure studies of issues relating to the speed with which prices incorporate public information announcements (such as central bank decisions or macroeconomic announcements), and "*tests for private information*", relating to the implications of microstructure models allowing for asymmetric information in the market (such as Glosten and Milgrom (1985)). The null hypothesis for the non predictability of returns is 1.25 with $\varepsilon_t = r_t$ the process of financial returns. In particular, the issue of predictability over long horizons can be addressed with long memory modelling for returns. Evidence of long range dependence in stock market returns is found in Greene and

Fielitz (1977). This finding raises a number of questions on the effects long memory in returns may have on portfolio decision and on derivative pricing using martingale methods. However, the finding of Greene and Fielitz (1977) is challenged by Lo (1991) with a slightly more powerful analysis based on a modified form of the R/S statistic. Lee and Robinson (1996) are the first to apply semiparametric methods to the measure of memory in stock price returns, and Lobato and Savin (1998) apply the Pitman-efficient test statistic developed in Lobato and Robinson (1998) to conclude with Lo (1991) that evidence of long memory in returns is spurious. They do, however, find strong evidence of long range dependence in the squared and absolute returns, as do Ding and Granger (1996). This refines the widely recognised stylised facts on conditionally heteroscedastic behaviour of financial returns (see Mandelbrot (1963) and Fama (1965) for a first description of the phenomenon) and reinforces the value of long memory estimation procedures robust to (possibly long memory) conditional heteroscedasticity when examining the long run predictability of returns.

Henry and Payne (1997) and Andersen and Bollerslev (1997a) find the same pattern in intra-day foreign exchange rate returns and give a rationale based on the aggregation of heterogeneous autoregressive information arrival processes on the market. In view of the more pervasive evidence of long range dependence in the volatilities than in the returns, the focus of interest is therefore naturally shifted to the investigation of temporal dependence in the volatility process. Modelling volatility is fundamental for several reasons. Volatility serves as a measure of risk, albeit very crude, and it is used in derivative pricing formulae such as the Black-Scholes formula. Secondly, the volatility process can be identified with an aggregate process of information arrivals on the market (see Clark (1973), Epps and Epps (1976) and Tauchen and Pitts (1983)). Thirdly, higher volatilities imply larger bid-ask spreads (difference between the buy and the sell quotations) and more generally, although knowledge of the volatility process alone is not sufficient to test whether such spreads are caused by asymmetric information or inventory control (or other microstructure questions relating to the "*tests for private information*"), modelling of volatility paves the way for the investigation of market efficiency in terms of information transmission ("*event studies*") and, in particular, the long run effect of transactions on the price process (see for instance Lyons (1985), and Hasbrouck (1991)).

1.7 Synopsis

The following two chapters are concerned with the effect of possibly long memory conditional heteroscedasticity on semiparametric estimation of long memory.

Chapter 2 considers the averaged periodogram statistic for a linear process with possibly long memory in the innovations conditional variance. An asymptotic normality result is given for averaged periodogram estimation of finite and positive spectral densities at zero frequencies. The proof is adapted from Robinson and Henry (1997). The robustness of the results in Robinson (1994c) regarding consistency of the averaged periodogram statistic in the presence of long memory is then shown and a Monte Carlo experiment assesses the effect of conditional heteroscedasticity in small sample averaged periodogram long memory estimation. The estimation of stationary cointegration is then discussed in this framework.

Chapter 3 presents the proofs of robustness to (possibly long memory) conditional heteroscedasticity of the consistency and asymptotic normality results for the local Whittle estimate of long memory in Robinson (1995a). A Monte Carlo study investigates the effect of conditional heteroscedasticity on local Whittle estimation of long memory in small samples. This chapter is based on a joint research with Peter Robinson, appearing in Robinson and Henry (1997).

Chapter 4 derives the asymptotic mean squared error of the local Whittle estimate of long memory and an automatic optimal bandwidth selection procedure. An extensive Monte Carlo study assesses its performance in comparison with automatic log periodogram estimation of long memory and automatic averaged periodogram estimation of long memory. The latter is proven to be valid under the current framework allowing for (possibly long memory) conditionally heteroscedastic innovations. An earlier version of this chapter is joint work with Peter Robinson and appears in

Henry and Robinson (1996).

Chapter 5 investigates patterns of long range dependence and co-dependence in several intra-day foreign exchange rate data series. Little evidence is found of predictable returns, but strong evidence is found of long range dependence in returns volatility, with a pattern of stationary cointegration in support of a mixture of distributions model for the inflow of information into the market. This chapter is the result of joint research with Richard Payne, part of which appears in Henry and Payne (1997). A very similar research was conducted simultaneously and independently by Andersen and Bollerslev (1997a) on a part of the same data set.

Chapter 2

Averaged periodogram statistic

2.1 Introduction

This second chapter is concerned with the use of an averaged periodogram statistic proposed by Grenander and Rosenblatt (1966) to investigate temporal dependence in weakly dependent time series. The process x_t considered is stationary and satisfies 1.1 and 1.2 with the martingale dependence assumption 1.25 on innovations ε_t. The approach is semiparametric in the sense that x_t is supposed to have spectral density $f(\lambda)$ satisfying the local specification 1.18 with $d_x \geq 0$; and the averaged periodogram statistic is used to investigate the behaviour of $f(\lambda)$ in a neighbourhood of zero frequency, estimating $f(0) = L(0)$ when $d_x = 0$ and estimating d_x when the latter is strictly positive. Section 2 of this chapter presents issues and past results.

In the use of a semiparametric approach, one may have in mind estimating dependence in long financial data series. To that end, asymptotic properties of the averaged periodogram statistic need to be justified when there is a possibly high degree of temporal dependence in conditional variances.

Asymptotic normality of the averaged periodogram estimate of $f(0)$ when $d_x = 0$ follows from Hannan (1970) under 1.26 and other regularity conditions. Section 3 of this chapter extends the validity of this result with unchanged asymptotic variance to the case 1.27 with σ_t^2 defined by 1.28 corresponding to (possibly long memory) conditional heteroscedasticity in the innovations of a generalised linear process. Consistency of the averaged periodogram based estimate of $d_x > 0$ is proved with a specific rate of convergence by Robinson (1994c). Section 4 of this chapter extends the validity of the latter result to processes satisfying 1.27 with σ_t^2 following 1.28.

A simple corollary is the extended validity of a consistent estimate of stationary cointegration proposed by Robinson (1994c). This is presented in Section 5 of this chapter while Section 6 proposes an investigation of the effect of conditional heteroscedasticity in small samples. Section 7 concludes this chapter.

2.2 Averaged periodogram statistic

Let the discrete Fourier transform of a covariance stationary process x_t be defined as in 1.52 and the periodogram $I_x(\lambda)$ as in 1.53. Define the averaged periodogram by

$$\hat{F}(\lambda) = \frac{2\pi}{n} \sum_{j=1}^{[\lambda n/2\pi]} I_x(\lambda_j) \tag{2.1}$$

where $\lambda_j = 2\pi j/n$, n is the sample size and $[x]$ denotes the largest integer smaller or equal to x. Because $I_x(\lambda_j)$ is invariant to location shift, no mean correction is necessary for 2.1. $\hat{F}(\lambda)$ is a discrete analogue of the more widely documented continuously averaged periodogram (see Ibragimov (1963)) where 1.53 is replaced by its demeaned version. The estimate $\hat{f}(0) = \hat{F}(\lambda_m)/\lambda_m$ given in 1.79 was proposed for $f(0)$ by Grenander and Rosenblatt (1966) and is readily generalisable to a wide class of weighted periodogram spectral estimates defined below. Let $K(\lambda)$ be a bounded function satisfying

$$\int_{-\infty}^{\infty} K(\lambda)d\lambda = 1, \quad K(-\lambda) = K(\lambda). \tag{2.2}$$

Defining

$$K_m(\lambda) = m \sum_{j=-\infty}^{\infty} K(m(\lambda + 2\pi j)) \tag{2.3}$$

where m is a positive integer called the bandwidth, weighted periodogram estimation of $f(0)$ is given by

$$\hat{f}_w(0) = \frac{2\pi}{n}\sum_{j=1}^{n-1} K_m(\lambda_j) I_x(\lambda_j). \tag{2.4}$$

The class of kernel functions such that

$$K_m(\lambda) = 0 \quad \text{for} \quad \lambda > \lambda_m \tag{2.5}$$

provides a basis for estimation of $f(0)$ under specification 1.18 with $d_x = 0$. Supposing 1.69 is satisfied, a set of sufficient conditions for

$$\hat{f}_w(0) \rightarrow_p f(0) \quad \text{as} \quad n \rightarrow \infty \tag{2.6}$$

includes absolute summability of fourth cumulants

$$\sum_{h,i,j=-\infty}^{+\infty} |\text{cum}(x_1, x_{1+h}, x_{1+i}, x_{1+j})| < \infty. \tag{2.7}$$

Suppose that a local Lipschitz condition is imposed on the spectral density in the form,

$$f(\lambda) = f(0)(1 + E_\beta \lambda^\beta) + o(\lambda^\beta) \quad \text{as} \quad \lambda \rightarrow 0^+, \tag{2.8}$$

with

$$\beta \in (0, 2], \quad 0 < f(0) < \infty, \quad 0 < E_\beta < \infty,$$

and suppose the bandwidth m satisfies 1.77. Under the conditions above, asymptotic normality of $\hat{f}(\lambda)$ given by 1.79

$$m^{\frac{1}{2}}(\hat{f}(0) - f(0)) \rightarrow_d N(0, f(0)^2) \quad \text{as} \quad n \rightarrow \infty \tag{2.9}$$

occurs under the two following sets of sufficient conditions: Brillinger (1975), Theorem 5.4.3, page 136 assumes 1.7 and existence of all moments of x_t; Hannan (1970), Theorem 13, page 224, assumes that x_t follows 1.1 with i.i.d. innovations. Hannan (1970), Theorem 13', page 227 also proves 2.9 under the uniform mixing condition 1.4 and fourth order stationarity with absolutely summable fourth cumulants as in 2.7. However, he needs the additional assumption that the spectral density of the process x_t, $f(\lambda)$, be absolutely continuous for all λ. Such a global condition is undesirable in this semiparametric framework where one wishes to allow for discontinuities in the spectrum, and indeed for any kind of behaviour for the spectrum at non zero frequencies, providing it remains integrable (a consequence of covariance stationarity).

Robinson (1983b) gives a survey of the possible applications of spectral estimation. One of the major applications of 1.79 is documented by Robinson and Velasco (1996). They show how a consistent estimate of the spectral density at zero of a weakly dependent process is instrumental in location inference, linear regression and more complex econometric models. As appears in 1.22 for instance, the sample mean of a process with nonparametric autocorrelation provides an asymptotically normal, if not efficient, estimation of the population mean where misspecified autocorrelation-corrected estimates might prove misleading.

2.3 Asymptotic normality of the averaged periodogram

As indicated in Section 2, the averaged periodogram statistic more commonly lends itself to the estimation of the spectral density at frequency zero for a weakly dependent process x_t (or indeed at any continuity point of the spectrum). The following theorem shows that the discretely averaged periodogram $(1/m)\sum_{j=1}^{m} I_x(\lambda_j)$ remains an asymptotically normal estimate for the spectral density at frequency zero of an observed generalised linear process with conditional heteroscedastic innovations.

We make the following assumptions:

Assumption A1 f satisfies 1.84. In addition, in a neighbourhood $(0,\delta)$ of the origin, $\alpha(\lambda)$ is differentiable and

$$\frac{d}{d\lambda}\log\alpha(\lambda) = O\left(\frac{|\alpha(\lambda)|}{\lambda}\right) \quad \text{as } \lambda \to 0^+, \tag{2.10}$$

where $\alpha(\lambda) = \sum_{j=0}^{\infty}\alpha_j e^{ij\lambda}$.

Assumption A2 m satisfies 1.77 and

$$\frac{m}{n^{1-2d_\varepsilon}} \to 0, \quad 0 < d_\varepsilon < \frac{1}{2}. \tag{2.11}$$

Assumption A3 x_t satisfies 1.1, 1.25, 1.27, 1.28, 1.33, 1.40 and 1.42. In addition, 1.46 holds with the same d_ε as in 2.11 and

$$\max_t E\varepsilon_t^8 < \infty, \tag{2.12}$$

$$E\left(\varepsilon_t^4 \varepsilon_u | \mathcal{F}_{t-1}\right) = E\left(\varepsilon_t^4 \varepsilon_u^2 \varepsilon_v | \mathcal{F}_{t-1}\right) = 0, \text{ a.s., } t \geq u \geq v, \tag{2.13}$$

and the α_j are quasi monotonically convergent, that is, $\alpha_j \to 0$ as $j \to \infty$ and there exists $J < \infty$ and $B < \infty$ such that for all $j \leq J$,

$$\alpha_{j+1} < \alpha_j \left(1 + \frac{B}{j}\right). \tag{2.14}$$

Consistency of $\hat{F}(\lambda_m)/\lambda_m$ holds when $f(\lambda)$ is only continuous at frequency zero, but a Lipschitz condition is necessary for asymptotic normality. 2.10 is needed for the treatment of fourth cumulant moments of the errors, to justify the martingale approximation. 1.77 is a minimum requirement for asymptotic normality in view of the fact that an optimal bandwidth rate is $n^{2\beta/2\beta+1}$ at which rate bias and asymptotic variance have the same order of magnitude. 2.11 strengthens 1.77 unless $d_\varepsilon \leq 1/(4\beta+2)$. 2.11 is required for the left-hand side of 2.9 to converge to a finite random variable. Quasi-monotonicity of the Fourier coefficients α_j of f, and boundedness of $f(0)$ imply absolute summability of the Fourier coefficients α_j:

$$\sum_{j=0}^{\infty} |\alpha_j| < \infty. \tag{2.15}$$

The requirement 2.13 that conditional odd moments be non stochastic up to seventh order is restrictive, but satisfied if ε_t has a conditionally symmetric density, or, more specially, if

$$\varepsilon_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2). \tag{2.16}$$

Note that 2.12 itself entails a restriction on the magnitude of the ϕ_j; see for instance the results of Engle (1982), Bollerslev (1986) for ARCH(1) and GARCH(1,1) processes under 2.16, and of Nelson (1990b) under more general distributional assumptions. However, 2.12 is only a sufficient condition in this setting. The quasi-monotonicity assumption on the α_j entails (see Yong (1974)), for all sufficiently large j,

$$|\alpha_j - \alpha_{j+1}| \leq K\frac{|\alpha_j|}{j}. \tag{2.17}$$

In fact, we believe that this requirement could be removed or relaxed by a more detailed proof, but the quasi-monotonicity requirement does not seem very onerous, while 2.11 is also needed when the ε_t^2 have long memory, and there always exists an m sequence satisfying both 1.77 and 2.11.

Theorem 1 Under Assumptions A1-A3, 2.9 holds.

Proof[1]

$$\begin{aligned} \frac{\sqrt{m}}{f(0)}\left(\frac{\hat{F}(\lambda_m)}{\lambda_m} - f(0)\right) &= \frac{m^{-\frac{1}{2}}}{f(0)}\sum_{j=1}^{m}\left(I_x(\lambda_j) - \frac{2\pi}{\sigma^2}f(\lambda_j)I_\varepsilon(\lambda_j)\right) && (2.18)\\ &+ \frac{m^{-\frac{1}{2}}}{f(0)}\sum_{j=1}^{m}f(\lambda_j)\left(\frac{2\pi I_\varepsilon(\lambda_j)}{\sigma^2} - 1\right) && (2.19)\\ &+ \frac{m^{-\frac{1}{2}}}{f(0)}\sum_{j=1}^{m}\left(f(\lambda_j) - f(0)\right). && (2.20)\end{aligned}$$

From 2.8,

$$\frac{\sqrt{m}}{f(0)}\left\{\frac{F(\lambda_m)}{\lambda_m} - f(0)\right\} = O\left(m^{-\frac{1}{2}}\sum_{j=1}^{m}\lambda_j^{\beta}\right) = O\left(\frac{m^{\beta+\frac{1}{2}}}{n^{\beta}}\right).$$

2.18 can be further decomposed into

$$m^{-\frac{1}{2}}\sum_{j=1}^{m}\left(\frac{I_x(\lambda_j)}{f(\lambda_j)} - \frac{2\pi}{\sigma^2}I_\varepsilon(\lambda_j)\right) \tag{2.21}$$

[1] The proof is adapted from the proof of Theorem 2 in Robinson and Henry (1997).

$$+ \quad m^{-\frac{1}{2}} \sum_{j=1}^{m} \left(\frac{1}{f(0)} - \frac{1}{f(\lambda_j)} \right) I_x(\lambda_j)$$

$$+ \quad \frac{2\pi}{\sigma^2} m^{-\frac{1}{2}} \sum_{j=1}^{m} \left(1 - \frac{f(\lambda_j)}{f(0)} \right) I_\varepsilon(\lambda_j).$$

Because $E(I_\varepsilon(\lambda_j)) = \sigma^2/2\pi$, the third term has first absolute moment bounded by

$$m^{-\frac{1}{2}} \sum_{j=1}^{m} \left| 1 - \frac{f(\lambda_j)}{f(0)} \right| = O\left(m^{\beta+\frac{1}{2}} n^{-\beta} \right) = o(1),$$

from 2.8 and 1.77. The second term has first absolute moment bounded by

$$m^{-\frac{1}{2}} \sum_{j=1}^{m} \left| 1 - \frac{f(\lambda_j)}{f(0)} \right| E\left(I_x(\lambda_j) \right). \tag{2.22}$$

Now,

$$\begin{aligned} E\left(I_x(\lambda) \right) &= \frac{1}{2\pi n} \sum_{t,s=1}^{n} E(x_t x_s) e^{i(t-s)\lambda} \\ &= \frac{1}{2\pi n} \sum_{t,s=1}^{n} \sum_{j,k=0}^{\infty} \alpha_j \alpha_k E(\varepsilon_{t-j} \varepsilon_{s-k}) e^{i(t-s)\lambda} \\ &= \frac{\sigma^2}{2\pi} \sum_{j=0}^{\infty} \sum_{t=-j}^{n-1} \alpha_t \alpha_{t+j} e^{it\lambda} \leq K \left(\sum_{j=0}^{\infty} |\alpha_j| \right)^2 < \infty, \end{aligned}$$

where the second equality is derived from application of 1.1, the third equality is derived from application of 1.25 and the last inequality follows from 2.15. Therefore, 2.22 is bounded by

$$K m^{-\frac{1}{2}} \sum_{j=1}^{m} \left| 1 - \frac{f(\lambda_j)}{f(0)} \right| = O\left(m^{\beta+\frac{1}{2}} n^{-\beta} \right) = o(1),$$

from 2.8 and 1.77. 2.21 has second moment equal to

$$m^{-1} \sum_{j,k=1}^{m} E\left(\frac{I_x(\lambda_j)}{f(\lambda_j)} - \frac{2\pi}{\sigma^2} I_\varepsilon(\lambda_j) \right) \left(\frac{I_x(\lambda_k)}{f(\lambda_k)} - \frac{2\pi}{\sigma^2} I_\varepsilon(\lambda_k) \right). \tag{2.23}$$

Robinson (1995a) proves that this is $o(1)$ under assumptions such that $\mathrm{cum}(\varepsilon_r, \varepsilon_s, \varepsilon_t, \varepsilon_u) = \kappa$ when $r = s = t = u$ and zero otherwise. Under the present assumptions, 1.62-1.64 also contribute. The complete fourth cumulant contribution to 2.23 is the following:

$$\frac{1}{m(2\pi n)^2}\sum_{j,k=1}^{m}\frac{1}{f(\lambda_j)f(\lambda_k)}\sum_{r,s,t,u=1}^{n}\mathrm{cum}(x_r,x_s,x_t,x_u)e^{i(r-s)\lambda_j-i(t-u)\lambda_k} \tag{2.24}$$

$$+\frac{1}{m(2\pi n)^2}\sum_{j,k=1}^{m}\left(\frac{2\pi}{\sigma^2}\right)^2\sum_{r,s,t,u=1}^{n}\mathrm{cum}(\varepsilon_r,\varepsilon_s,\varepsilon_t,\varepsilon_u)e^{i(r-s)\lambda_j-i(t-u)\lambda_k} \tag{2.25}$$

$$-\frac{1}{m(2\pi n)^2}\sum_{j,k=1}^{m}\frac{1}{f(\lambda_k)}\frac{2\pi}{\sigma^2}\sum_{r,s,t,u=1}^{n}\mathrm{cum}(\varepsilon_r,\varepsilon_s,x_t,x_u)e^{i(r-s)\lambda_j-i(t-u)\lambda_k} \tag{2.26}$$

$$-\frac{1}{m(2\pi n)^2}\sum_{j,k=1}^{m}\frac{1}{f(\lambda_j)}\frac{2\pi}{\sigma^2}\sum_{r,s,t,u=1}^{n}\mathrm{cum}(x_r,x_s,\varepsilon_t,\varepsilon_u)e^{i(r-s)\lambda_j-i(t-u)\lambda_k} \tag{2.27}$$

Now, applying 1.1,

$$\begin{aligned}\mathrm{cum}(x_r,x_s,x_t,x_u) &= \sum_{p=-\infty}^{r}\sum_{q=-\infty}^{s}\sum_{l=-\infty}^{t}\sum_{v=-\infty}^{u}\alpha_{r-p}\alpha_{s-q}\alpha_{t-l}\alpha_{u-v}\mathrm{cum}(\varepsilon_p,\varepsilon_q,\varepsilon_l,\varepsilon_v)\\ &= \kappa\sum_{p=-\infty}^{n}\alpha_{r-p}\alpha_{s-p}\alpha_{t-p}\alpha_{u-p}\\ &+ \sum_{\substack{p\neq q\\ -\infty}}^{n}\gamma_{p-q}\left(\alpha_{r-p}\alpha_{s-p}\alpha_{t-q}\alpha_{u-q}+\alpha_{r-p}\alpha_{s-q}\alpha_{t-p}\alpha_{u-q}\right.\\ &\qquad\left.+\alpha_{r-p}\alpha_{s-q}\alpha_{t-q}\alpha_{u-p}\right)\end{aligned}$$

in view of 1.61-1.64 and with the convention that $\alpha_j = 0,\ j < 0$. In the same way,

$$\begin{aligned}\mathrm{cum}(\varepsilon_r,\varepsilon_s,x_t,x_u) &= \sum_{p=-\infty}^{t}\sum_{q=-\infty}^{u}\alpha_{t-p}\alpha_{u-q}\mathrm{cum}(\varepsilon_r,\varepsilon_s,\varepsilon_p,\varepsilon_q)\\ &= \kappa\delta_{rs}\alpha_{t-r}\alpha_{u-r}+\delta_{rs}\sum_{p=-\infty}^{\min(t,u)}\gamma_{r-p}\alpha_{t-p}\alpha_{u-p}\\ &+ \gamma_{r-s}(\alpha_{t-r}\alpha_{u-s}+\alpha_{t-s}\alpha_{u-r}),\end{aligned}$$

and a symmetric expression can be written for $\mathrm{cum}(x_r,x_s,\varepsilon_t,\varepsilon_u)$. The contributions from κ is also proved to be $o(1)$ in Robinson (1995a). The contribution of 1.62-1.64 is the following:

$$\frac{m^{-1}n^{-2}}{(2\pi)^2}\sum_{j,k=1}^{m}\frac{1}{f(\lambda_j)f(\lambda_k)}\sum_{r,s,t,u=1}^{n}\sum_{\substack{p\neq q\\ -\infty}}^{n}\gamma_{p-q}\left(\alpha_{r-p}\alpha_{s-p}\alpha_{t-q}\alpha_{u-q}\right.$$

$$\left.+\alpha_{r-p}\alpha_{s-q}\alpha_{t-p}\alpha_{u-q}+\alpha_{r-p}\alpha_{s-q}\alpha_{t-q}\alpha_{u-p}\right)e^{i(r-s)\lambda_j-i(t-u)\lambda_k} \tag{2.28}$$

$$+\frac{m^{-1}n^{-2}}{\sigma^4}\sum_{j,k=1}^{m}\sum_{\substack{r\neq s\\ 1}}^{n}\gamma_{r-s}\left(1+e^{i(r-s)(\lambda_j+\lambda_k)}+e^{i(r-s)(\lambda_j-\lambda_k)}\right) \tag{2.29}$$

$$- \frac{2m^{-1}n^{-2}}{2\pi\sigma^2} \sum_{j,k=1}^{m} \frac{1}{f(\lambda_k)} \sum_{r,s,t,u=1}^{n} \gamma_{t-u} \left(\alpha_{r-t}\alpha_{s-u} + \alpha_{r-u}\alpha_{s-t}\right) e^{i(r-s)\lambda_j - i(t-u)\lambda_k} \tag{2.30}$$

$$- \frac{2m^{-1}n^{-2}}{2\pi\sigma^2} \sum_{j,k=1}^{m} \frac{1}{f(\lambda_k)} \sum_{r,s,t=1}^{n} \sum_{p=-\infty}^{\min(t,u)} \gamma_{r-p}\alpha_{t-p}\alpha_{u-p} e^{-i(t-u)\lambda_k}. \tag{2.31}$$

From 2.15, the four contributions above are $O\left(mn^{-1}\sum_{t=1}^{\infty}|\gamma_t|\right)$, which is $O(n^{2d_\varepsilon-1}m)$ by 1.46, and therefore, from 2.11, it is $o(1)$ as required.

There remains to prove that

$$m^{-\frac{1}{2}} \sum_{j=1}^{m} f(\lambda_j)\left(\frac{2\pi I_{\varepsilon j}}{\sigma^2} - 1\right) \rightarrow_p N(0,1). \tag{2.32}$$

The left-hand side is a martingale equal to $\sum_{t=1}^{n} z_t$ with $z_t = \varepsilon_t \sum_{s=1}^{t-1} \varepsilon_s c_{t-s}$, where $c_s = 2m^{-1/2}\mu_s/f(0)$ and $\mu_s = \frac{1}{n}\sum_{j=1}^{m} f(\lambda_j)\cos s\lambda_j$. We wish to show that as $n \to \infty$

$$\sum_{t=1}^{n} E(z_t^4) \to 0, \tag{2.33}$$

$$\sum_{t=1}^{n} E\left(z_t^2 | \mathcal{F}_{t-1}\right) \rightarrow_p \sigma^4, \tag{2.34}$$

which, following Brown's Martingale Central Limit Theorem (in Brown (1971)), implies 2.32. By the Schwarz inequality, $E(z_t^4) \le (E\varepsilon_t^8)^{\frac{1}{2}}(E\xi_t^8)^{\frac{1}{2}}$. Because $\xi_t = \sum_{s=1}^{t-1} \varepsilon_s c_{t-s}$ is a martingale, by Burkholder's inequality (Burkholder (1973)),

$$E\left(\xi_t^8\right) \le KE\Big(\sum_{s=1}^{t-1} c_{t-s}^2 \varepsilon_s^2\Big)^4 \le \max_s E\varepsilon_s^8 \Big(\sum_{s=1}^{n} c_s^2\Big)^4.$$

Now, putting $f_j = f(\lambda_j)/f(0)$,

$$\begin{aligned} \sum_{s=1}^{n} c_s^2 &= \sum_{s=1}^{n} 4m^{-1}n^{-2}\left(\sum_{j=1}^{m} f_j \cos s\lambda_j\right)^2 \\ &= 4m^{-1}n^{-2}\sum_{j=1}^{m} f_j^2 \sum_{s=1}^{n}\cos^2 s\lambda_j \\ &+ 2m^{-1}n^{-2}\sum_{\substack{j\neq k \\ 1}}^{m} f_j f_k \sum_{s=1}^{n}\left[\cos s(\lambda_j+\lambda_k) + \cos s(\lambda_j - \lambda_k)\right]. \end{aligned} \tag{2.35}$$

2.35 is $O(1/n)$ whereas, using trigonometric identities in Zygmund (1977) p. 49,

$$\sum_{s=1}^{n}[\cos s(\lambda_j+\lambda_k)+\cos s(\lambda_j-\lambda_k)] =$$
$$\frac{\sin(n+\frac{1}{2})(\lambda_j+\lambda_k)}{2\sin\frac{1}{2}(\lambda_j+\lambda_k)}+\frac{\sin(n+\frac{1}{2})(\lambda_j-\lambda_k)}{2\sin\frac{1}{2}(\lambda_j-\lambda_k)}-1=0$$

for $j \neq k$, so that

$$\sum_{s=1}^{n} c_s^2 = O\left(\frac{1}{n}\right), \tag{2.36}$$

which in turn implies

$$\sum_{t=1}^{n} E\left(z_t^4\right) \leq \frac{K}{n} \to 0$$

to verify 2.33. To check 2.34, write

$$E\left(z_t^2|\mathcal{F}_{t-1}\right) = \sigma_t^2\xi_t^2 = \sigma^2\xi_t^2 + (\sigma_t^2-\sigma^2)\xi_t^2.$$

From (4.14) and (4.15) of Robinson (1995a),

$$\sum_{t=1}^{n}\xi_t^2 - \sigma^2 = \sum_{t=1}^{n-1}\chi_t r_{n-t} + \sigma^2\left\{\sum_{t=1}^{n-1} r_{n-t} - 1\right\} + \sum_{t=2}^{n}\sum_{r\neq s}\varepsilon_r\varepsilon_s c_{t-r}c_{t-s}, \tag{2.37}$$

writing $\chi_t = \varepsilon_t^2 - \sigma^2$ and $r_t = c_1^2 + \ldots + c_t^2$. The first term on the right has mean zero and variance

$$\sum_{t=1}^{n-1}\sum_{u=1}^{n-1}\gamma_{t-u}r_{n-t}r_{n-u}, \tag{2.38}$$

where $\gamma_j = \mathrm{cov}(\varepsilon_t^2, \varepsilon_{t+j}^2)$. $\sum_{t=1}^{n-1} r_{n-t}$ can be written

$$\frac{4}{mn^2}\sum_{j=1}^{m} f_j^2 \sum_{t=1}^{n-1}\sum_{s=1}^{n-t}\cos^2(s\lambda_j) + \frac{2}{mn^2}\sum_{j\neq k} f_j f_k \sum_{t=1}^{n-1}\sum_{s=1}^{n-t}[\cos(s(\lambda_j+\lambda_k))+\cos(s(\lambda_j-\lambda_k))].$$

Robinson (1995a) shows that

$$\sum_{t=1}^{n-1}\sum_{s=1}^{n-t}\cos^2(s\lambda_j) = \frac{(n-1)^2}{4},$$

$$\sum_{t=1}^{n-1}\sum_{s=1}^{n-t}[\cos(s(\lambda_j+\lambda_k))+\cos(s(\lambda_j-\lambda_k))] = -n.$$

Because

$$\frac{1}{m}\sum_{j=1}^{m} f_j^2 = 1 + O\left(\lambda_m^{2\beta}\right) = 1 + o(m^{-1}),$$
$$\frac{1}{mn}\sum_{j\neq k} f_j f_k = O\left(\frac{1}{n}\sum_{j=1}^{m} f_j^2\right) = O\left(\frac{m}{n}\right),$$

it follows that

$$\sum_{t=1}^{n-1} r_{n-t} \to 1, \quad \text{as } n \to \infty. \tag{2.39}$$

As, moreover,

$$|\gamma_j| \le K\Phi_0\Phi_j = O\left(j^{2d_\epsilon - 1}\right) \to 0, \quad \text{as } j \to \infty \tag{2.40}$$

by 2.12 and 1.46, it follows from the Toeplitz lemma that 2.38 tends to zero. Clearly, the second term in 2.37 thus tends to zero, whereas the last term has mean zero and variance bounded by

$$2\left(\max_t E\varepsilon_t^4\right) \sum_{t,u=2}^{n} \sum_{\substack{r \ne s \\ 1}}^{\min(t-1,u-1)} |c_{t-r}c_{t-s}c_{u-r}c_{u-s}|. \tag{2.41}$$

This follows from the corresponding derivation in Robinson (1995a), but upper bounding $E(\varepsilon_t^2 \varepsilon_s^2)$ by the Schwarz inequality. Following Robinson (1995a) step by step, we bound 2.41 by

$$K\sum_{t=2}^{n}\sum_{\substack{r \ne s \\ 1}}^{t-1} c_{t-r}^2 c_{t-s}^2 + K\sum_{t=3}^{n}\sum_{u=2}^{t-1}\sum_{\substack{r \ne s \\ 1}}^{u-1} |c_{t-r}c_{t-s}c_{u-r}c_{u-s}|. \tag{2.42}$$

From 2.36, the left-hand side is $O(1/n)$. By summation by parts, we have

$$|c_s| \;=\; 2m^{-1/2}n^{-1}\left|\sum_{j=1}^{m-1}(f_j - f_{j+1})\sum_{l=1}^{j}\cos s\lambda_l + f_m \sum_{j=1}^{m}\cos s\lambda_l\right|. \tag{2.43}$$

Now $|\sum_{l=1}^{j}\cos s\lambda_l| = O(n/s)$ for $1 \le j \le m$ and $1 \le s \le n/2$ by Zygmund (1977) page 2. Moreover, by 1.84,

$$\sum_{j=1}^{m}|f_j - f_{j+1}| = O\left(\sum_{j=1}^{m}|\lambda_j^\beta - \lambda_{j+1}^\beta|\right) = O\big((m/n)^\beta\big). \tag{2.44}$$

Therefore, $|c_s| = O(m^{\beta-1/2}n^{-\beta}s^{-1} + m^{-1/2}s^{-1}) = O(m^{-1/2}s^{-1})$. It is immediate to see that $|c_s|$ is also $O(m^{1/2}n^{-1})$. Therefore, from Robinson's derivation (in Robinson (1995a)), the right-hand side of 2.41 is bounded by

$$Knr_n\left(\sum_{j=2}^{[n/n^{2/3}]} jc_j^2 + \sum_{[n/n^{2/3}]+1}^{[n/2]} jc_j^2\right)$$

which is

$$O\left(\frac{n^3}{m^{4/3}}\left(\frac{m^{1/2}}{n}\right)^2+n^2\left(\frac{1}{m^{1/2}}\right)^2\sum_{s=[n/m^{2/3}]}^{\infty}s^{-2}\right)=O\left(\frac{n}{m^{1/3}}\right)$$

as $n\to\infty$, so that 2.41 is $O\left(m^{-1/3}\right)$ in view of 2.36. It remains to show that

$$\sum_{t=1}^{n}(\sigma_t^2-\sigma^2)\xi_t^2\to_p 0.$$

The left side is

$$\sigma^2\sum_{t=1}^{n}\left(\sigma_t^2-\sigma^2\right)r_{t-1}+\sum_{t=1}^{n}\left(\sigma_t^2-\sigma^2\right)\sum_{s=1}^{t-1}c_{t-s}^2\chi_s+\sum_{t=1}^{n}\left(\sigma_t^2-\sigma^2\right)\sum_{\substack{v\neq s\\ 1}}^{t-1}\varepsilon_v\varepsilon_s c_{t-v}c_{t-s} \quad (2.45)$$

The first term is

$$\sigma^2\sum_{t=2}^{n}\sum_{j=1}^{\infty}\psi_j\chi_{t-j}r_{t-1}=\sigma^2\left(S_1+S_2\right),$$

where

$$S_1=\sum_{j=1-n}^{n-1}\chi_j\sum_{t=1}^{n-1}r_t\psi_{t-j+1},\quad S_2=\sum_{j=-\infty}^{-n}\chi_j\sum_{t=1}^{n-1}r_t\psi_{t-j+1},$$

and $\psi_j=0,\ j\leq 0$. Now S_1 has mean zero and variance

$$\begin{aligned}\sum_{j,k=1-n}^{n-1}\gamma_{j-k}\sum_{s,t=1}^{n-1}r_s r_t\psi_{s-j+1}\psi_{t-k+1} &\leq Knr_{n-1}^2\Big(\sum_{j=1}^{\infty}|\psi_j|\Big)^2\sum_{j=0}^{2n-2}|\gamma_j|\\ &= O\left(\frac{1}{n^{1-2d_\varepsilon}}\right)\to 0,\quad \text{as } n\to\infty,\end{aligned}$$

using 1.42, 2.40 and 2.36. On the other hand

$$E\left|S_2\right|\leq K\sum_{t=1}^{n-1}r_t\sum_{j=n}^{\infty}|\psi_j|\to 0,\quad \text{as } n\to\infty,$$

from 1.42 and 2.39, so that the first term in 2.45 is $o_p(1)$. The second term in 2.45 is

$$\sum_{t=2}^{n}\sum_{v=-\infty}^{0}\psi_{t-v}\sum_{s=1}^{t-1}c_{t-s}^2\chi_v\chi_s \quad (2.46)$$

$$+\sum_{t=2}^{n}\sum_{v=1}^{t-1}\psi_{t-v}\sum_{s=1}^{t-1}c_{t-s}^2\chi_v\chi_s. \quad (2.47)$$

The expectation of the absolute value of 2.46 is bounded by

$$K\left(\max_t E\varepsilon_t^4\right)\sum_{t=2}^{n}\sum_{j=t}^{\infty}|\psi_j|r_{t-1}\to 0$$

using 1.42, 2.39 and the Toeplitz lemma. 2.47 includes the component

$$\sum_{t=2}^{n}\sum_{s=1}^{t-1}\psi_{t-s}c_{t-s}^2\chi_s^2,$$

whose absolute value has expectation which likewise tends to zero. The remainder of 2.47 can be written

$$\sum_{t=2}^{n}\sum_{v=1}^{t-1}\psi_{t-v}\sum_{s=1}^{v-1}c_{t-s}^2\chi_v\chi_s+\sum_{t=2}^{n}\sum_{v=1}^{t-1}\psi_{t-v}\sum_{s=v+1}^{t-1}c_{t-s}^2\chi_v\chi_s. \tag{2.48}$$

The first term in 2.48 has mean square

$$\sum_{t,u=2}^{n}\sum_{v=1}^{t-1}\psi_{t-v}\sum_{s=1}^{v-1}c_{t-s}^2\sum_{q=1}^{u-1}\psi_{u-q}\sum_{p=1}^{q-1}c_{u-p}^2E\left(\chi_v\chi_s\chi_q\chi_p\right). \tag{2.49}$$

Now each (v,s,q,p) such that $s<v$, $p<q$ satisfies one of the relations $v=q$, $s\le q<v$, $q<s<v$, $p\le v<q$ or $v<p<q$. The contribution from summands in 2.49 such that $v=q$ is bounded by

$$\begin{aligned}&K\left(\max_t E\chi_t^4\right)\sum_{t,u=2}^{n}\sum_{v=1}^{\min(t-1,u-1)}\left|\psi_{t-v}\psi_{u-v}\right|\sum_{s=1}^{v-1}c_{t-s}^2\sum_{p=1}^{v-1}c_{u-p}^2\\ &\le K\left(\max_t E\varepsilon_t^8\right)r_{n-1}^2 n\Big(\sum_{j=1}^{\infty}|\psi_j|\Big)^2=O\left(1/n\right)\to 0.\end{aligned}$$

Next, for $v>q\ge s$, $p<q$,

$$E\left(\chi_v\chi_s\chi_q\chi_p\right)=E\left\{\sum_{j=-\infty}^{q}\phi_{v-j}\nu_j\chi_s\chi_q\chi_p\right\}, \tag{2.50}$$

because

$$E\left(\chi_v|\mathcal{F}_q\right)=\sum_{j=-\infty}^{q}\phi_{v-j}\nu_j,\quad \text{a.s.},\quad v>q, \tag{2.51}$$

as follows from 1.33 and

$$E\left(\nu_j|\mathcal{F}_q\right)=E\left(\varepsilon_j^2|\mathcal{F}_q\right)-E\left(E\left(\varepsilon_j^2|\mathcal{F}_{j-1}\right)|\mathcal{F}_q\right)=0,\quad \text{a.s.},\quad q<j.$$

Now 2.50 is bounded in absolute value by

$$E\left|\Big(\sum_{j=-\infty}^{q}\phi_{v-j}\nu_j\Big)\chi_s\chi_q\chi_p\right|\;\le\;\left\{E\left(\sum_{j=-\infty}^{q}\phi_{v-j}\nu_j\right)^4\left(\max_t E\chi_t^4\right)^3\right\}^{\frac{1}{4}}$$

$$\le K\left\{E\left(\sum_{j=-\infty}^{q}\phi_{v-j}^2\nu_j^2\right)^2\right\}^{\frac14}$$
$$\le K\Phi_{v-q}^{\frac12}\left(\sum_{j=-\infty}^{q}\phi_{v-j}^2E\left(\nu_j^4\right)\right)^{\frac14}$$
$$\le K\Phi_{v-q},$$

where the second inequality employs Burkholder's (1973) inequality and the final one $E\left(\nu_j^4\right) \le 8\left[E\left(\varepsilon_j^8\right)+E\left\{E\left(\varepsilon_j^2|\mathcal{F}_{j-1}\right)\right\}^4\right] \le K$, by 2.12. Considering similarly the three cases $\{p<q<s<v\}$, $\{p\le v<q$ and $s<v\}$ and $\{s<v<p<q\}$, we have

$$|E(\chi_v\chi_s\chi_q\chi_p)| \le K(\Phi_{v-q}+\Phi_{v-s}+\Phi_{q-v}+\Phi_{q-p})$$

whenever $s<v$, $p<q$ and $v\neq q$, where $\Phi_j=0$ for $j<0$. Thus the contribution to 2.49 for $v\neq q$ is bounded in absolute value by

$$K\sum_{t,u=2}^{n}\sum_{v=1}^{t-1}|\psi_{t-v}|\sum_{s=1}^{v-1}c_{t-s}^2\sum_{q=1}^{u-1}|\psi_{u-q}|\sum_{p=1}^{q-1}c_{u-p}^2(\Phi_{v-q}+\Phi_{v-s}+\Phi_{q-v}+\Phi_{q-p})$$

$$\begin{aligned}&\le K\sum_{t,u=2}^{n}\left\{\sum_{v=1}^{t-1}\sum_{q=1}^{u-1}|\psi_{t-v}\psi_{u-q}|\Phi_{v-q}\right\}r_{t-1}r_{u-1}\\&+K\sum_{j=1}^{\infty}|\psi_j|\sum_{u=2}^{n}r_{u-1}\sum_{t=2}^{n}\sum_{s=1}^{t-1}c_{t-s}^2\left\{\sum_{v=1}^{t-1}|\psi_{t-v}|\Phi_{v-s}\right\}.\end{aligned} \tag{2.52}$$

The terms in braces are bounded respectively by

$$\sum_{i,j=0}^{\infty}|\psi_i\psi_{i+j+u-t}|\Phi_j,\quad \sum_{i=1}^{\infty}|\psi_i|\Phi_{t-s-i},$$

which tend to zero as $|u-t|\to\infty$ and $|t-s|\to\infty$ respectively, in view of 1.40 and 1.42 and the Toeplitz lemma. Thus, 1.42, 2.39 and the Toeplitz lemma further imply that 2.52$\to$0 as $n\to\infty$, completing the proof that the first term of 2.48 is $o_p(1)$. The second term of 2.48 can be treated in the same way to conclude that 2.47 is $o_p(1)$. The last term of 2.45 is

$$2\sum_{t=2}^{n}\sum_{j=-\infty}^{t-1}\psi_{t-j}\chi_j\sum_{\substack{v<s\\1}}^{t-1}\varepsilon_v\varepsilon_s c_{t-v}c_{t-s}. \tag{2.53}$$

Now, note that

$$E\left(\chi_j \varepsilon_s \varepsilon_v \chi_k \varepsilon_r \varepsilon_u\right) = 0, \quad v < s,\ u < r,\ v \neq u \text{ or } s \neq r.$$

This follows by proceeding recursively using 1.28 and nested conditional expectations, and the fact that $E\left(\varepsilon_t|\mathcal{F}_{t-1}\right)$, $E\left(\varepsilon_t^3|\mathcal{F}_{t-1}\right)$, $E\left(\varepsilon_t^4\varepsilon_u|\mathcal{F}_{u-1}\right)$, $t \geq u$ and $E\left(\varepsilon_t^4\varepsilon_u^2\varepsilon_v|\mathcal{F}_{v-1}\right)$, $t \geq u \geq v$, are all a.s. zero under A3. On the other hand, for all indices,

$$\left|E\left(\chi_j \varepsilon_s \varepsilon_v \chi_k \varepsilon_r \varepsilon_u\right)\right| \leq \max_t E\left(\varepsilon_t^8\right) < \infty$$

by Hölder's inequality. It follows that 2.53 has second moment

$$4 \sum_{t,u=2}^{n} \sum_{j=-\infty}^{t-1} \psi_{t-j} \sum_{k=-\infty}^{u-1} \psi_{u-k} \sum_{\substack{v<s \\ 1}}^{min(t,u)-1} c_{t-v}c_{t-s}c_{u-v}c_{u-s} E\left(\chi_j \chi_k \varepsilon_v^2 \varepsilon_s^2\right)$$
$$\leq K \sum_{t,u=2}^{n} \sum_{\substack{v<s \\ 1}}^{min(t,u)-1} \left|c_{t-v}c_{t-s}c_{u-v}c_{u-s}\right| = O\left(\frac{1}{m^{\frac{1}{3}}}\right)$$

as in 2.41, to complete the proof that 2.37 $\rightarrow_p 0$ and thus of 2.34.

2.4 Consistency under long memory

In case $f(\lambda)$ satisfies 1.18, the singularity at zero naturally precludes estimation of $f(0)$, but the strength of the dependence in the process is embodied in the slope of the spectrum in a neighbourhood of frequency zero. Properties of the discrete Fourier transforms in this framework are discussed in Rosenblatt (1981) and in Chapter 1 of this thesis. Yajima (1989) proves a central limit theorem for discrete Fourier transforms of strictly stationary processes with finite moments of all orders and absolutely summable higher-order cumulants at fixed non zero frequencies, whereas Theorem 2 of Robinson (1995b) gives the orders of magnitude 1.80 and 1.81 for moments of discrete Fourier transforms in a neighbourhood of frequency zero under 1.18, and very weak additional regularity conditions. Their asymptotic distribution under the current framework is still an open question. However, as shown in Robinson (1994c), the slope of the spectrum at frequency zero can be consistently estimated using the averaged periodogram statistic providing

$$\frac{\hat{F}(\lambda_m)}{F(\lambda_m)} \rightarrow_p 1 \quad \text{as } n \rightarrow \infty. \tag{2.54}$$

This chapter proves that 2.54 continues to hold when 1.26 is replaced by 1.27 with 1.28. Robinson (1994c) proved 2.54 under minimal moment conditions capable of delivering convergence in probability only, whereas this chapter requires 1.43. This is particularly unfortunate, as 1.43 reduces the scope of GARCH specifications covered by 1.28 and is generally not supported by empirical evidence on financial returns (see He (1997) for an investigation of the fourth moment structure of the GARCH model).

We now consider the case where the observed process x_t displays long memory, with the degree of temporal dependence embodied in the long memory parameter d_x. The following theorem shows that the weak consistency result 2.54 continues to hold when the error process ε_t displays (possibly long memory) conditional heteroscedasticity.

The following assumptions are introduced:

Assumption B1 f satisfies 1.18, for $0 < d_x < \frac{1}{2}$. In addition, 2.10 holds for $\alpha(\lambda) = \sum_{j=0}^{\infty} \alpha_j e^{ij\lambda}$.

Assumption B2 m satisfies 1.69.

Assumption B3 x_t satisfies 1.1, with 1.25, 1.28-1.33, 1.40, 1.42 and 1.43. In addition either

$$E\left(\varepsilon_t^3 | \mathcal{F}_{t-1}\right) = E(\varepsilon_t^3) \text{ a.s.}, \quad t = 0, \pm 1 \ldots, \tag{2.55}$$

or

$$\sum_{j=0}^{\infty} |\phi_j| < \infty. \tag{2.56}$$

Assumptions B1 and B2 correspond to assumptions A and B in Robinson (1994c) with the addition of 2.10 which is used to control the martingale approximation term 2.59 below. 2.10 is added here for clarity of the proof. It is not strictly necessary for the consistency results of Theorems 2 and 3. Only the magnitude of the singularity is specified for the spectrum at frequency zero while Assumption B2 is a minimal assumption for semiparametric estimation based on a degenerating band of harmonic frequencies. Assumption B3 relaxes the restriction on fourth cumulants (condition C(ii) in Robinson (1994c)) through the introduction of conditional heteroscedasticity. Robinson only assumed the innovations are uncorrelated, whereas in Assumption B3, they follow a martingale difference sequence. Condition C(iii) in Robinson (1994c) is replaced by the fourth moment condition 1.43. Assumption A3 implies Condition C(iv), as

$$\sum_{t=1}^{n}\left(\varepsilon_t^2-\sigma^2\right)=o_p(n) \tag{2.57}$$

Indeed, left side of 2.57 has mean zero and variance

$$\sum_{t,s=1}^{n}\sum_{j,k=0}^{\infty}\phi_j\phi_k E(\nu_{t-j}\nu_{s-k})=\sum_{t,s=1}^{n}\sum_{j=0}^{\infty}\phi_j\phi_{j+s-t}E(\nu_{t-j}^2) \tag{2.58}$$

in view of 1.25, with $\phi_j=0,\ j<0$. In view of 1.44 and the Cauchy inequality, 2.58 is, with $\Phi_j=\left(\sum_{i=j}^{\infty}\phi_i^2\right)^{\frac{1}{2}}$, equal to

$$O\left(n\sum_{j=0}^{\infty}\phi_j^2+n\Phi_0\sum_{j=1}^{n-1}\Phi_j\right)=o\left(n^2\right)$$

by the Toeplitz lemma and 2.56, thus verifying 2.57.

Again, the requirement 2.55 that conditional third moments be nonstochastic is restrictive, but again is satisfied if ε_t has a conditionally symmetric density, or, more specially, if it follows 2.16. The alternative requirement 2.56 rules out long memory in ε_t^2 but covers standard ARCH and GARCH specifications as well as many processes for which autocorrelation in squares decays more slowly than exponentially. As noted above, 1.43 entails a restriction on the magnitude of the ϕ_j. However, 1.43 is not a necessary condition, and indeed, under 2.56 it can be shown to be unnecessary by means of a longer argument, involving truncations, than that in the proof of the following theorem.

<u>Theorem 2</u> Under Assumptions B1-B3, 2.54 holds.

Proof Calling $I_\varepsilon(\lambda)$ the periodogram of the innovations, Robinson (1994c) wrote

$$\hat{F}(\lambda_m) - F(\lambda_m) = \frac{2\pi}{n}\sum_{j=1}^{m}\left(I_x(\lambda_j) - \frac{2\pi}{\sigma^2}f(\lambda_j)I_\varepsilon(\lambda_j)\right) \tag{2.59}$$

$$+ \frac{2\pi}{n}\sum_{j=1}^{m} f(\lambda_j)\left(\frac{2\pi I_\varepsilon(\lambda_j)}{\sigma^2} - 1\right) \tag{2.60}$$

$$+ \frac{2\pi}{n}\sum_{j=1}^{m} f(\lambda_j) - F(\lambda_m). \tag{2.61}$$

The first parts of Propositions 1 and 2 of Robinson (1994c) carry through to Assumptions A1 to A3, so that 2.61 is $o(F(\lambda_m))$.

From (3.17) of Robinson (1995a) and Theorem 2 of Robinson (1995b) whose proofs are not affected by conditional heteroscedaticity,

$$E\left|I_x(\lambda_j) - \frac{2\pi}{\sigma^2}f(\lambda_j)I_\varepsilon(\lambda_j)\right| = O\left(f(\lambda_j)\left(\frac{\log j}{j}\right)^{\frac{1}{2}}\right),$$

so that 2.59 is $O_p\left(\frac{1}{n}\sum_{j=1}^{m} f(\lambda_j)\left(\frac{\log j}{j}\right)^{\frac{1}{2}}\right)$. From 1.18 and lemma 3(ii) in Robinson (1994c), 2.59 is therefore $O_p\left(\left(\frac{\log m}{m}\right)^{\frac{1}{2}} F(\lambda_m)\right)$ which is $o_p(m^{\eta-\frac{1}{2}}F(\lambda_m))$, for any $\eta > 0$.

There remains to prove that 2.60 is $o_p(F(\lambda_m))$. The left-hand side of 2.60 is proportional to

$$\left\{\frac{1}{n}\sum_{j=1}^{m} f(\lambda_j)\right\}\frac{1}{n\sigma^2}\sum_{j=1}^{n}(\varepsilon_t^2 - \sigma^2) + \frac{2}{n\sigma^2}\sum_{s<t}\varepsilon_t\varepsilon_s\mu_{t-s}, \tag{2.62}$$

where $\mu_t = \frac{1}{n}\sum_{j=1}^{m} f(\lambda_j)\cos t\lambda_j$. In view of 2.57, it is sufficient to show that

$$\sum_{s<t}\varepsilon_t\varepsilon_s\mu_{t-s} = o_p(nF(\lambda_m)). \tag{2.63}$$

The left hand side of 2.63 has variance

$$\sum_{s<t} E(\varepsilon_t^2\varepsilon_s^2)\mu_{t-s}^2 + 2\sum_{t>s>u} E(\varepsilon_t^2\varepsilon_s\varepsilon_u)\mu_{t-s}\mu_{t-u} \tag{2.64}$$

from 1.25. Substituting 1.33 in the second term of 2.64 yields

$$2E\left(\sum_{\substack{u<s<t \\ 1}}^{n}\left(\sigma^2 + \sum_{j=0}^{\infty}\phi_j\nu_{t-j}\right)\varepsilon_u\varepsilon_s\mu_{t-s}\mu_{t-u}\right)$$

$$= 2 \sum_{\substack{u<s<t \\ 1}}^{n} \phi_{t-s} E\left(\nu_s \varepsilon_u \varepsilon_s\right) \mu_{t-s}\mu_{t-u}$$

$$= 2 \sum_{\substack{u<s<t \\ 1}}^{n} \phi_{t-s} E\left(\varepsilon_s^3 \varepsilon_u\right) \mu_{t-s}\mu_{t-u} \ ,$$

where the first equality applies nested conditional expectations and 1.25 for $j > t-s$, and 1.35 for $j < t-s$, whereas the second equality employs 1.34 and nested conditional expectations with 1.25 to verify $E(\sigma_s^2 \varepsilon_s \varepsilon_u) = 0$ for $u < s$. Under 2.55, this is identically zero. Under 2.56, it is bounded in absolute value by

$$\begin{aligned} 2n \max_t E(\varepsilon_t^4) \sum_{t>s=1-n}^{n-1} |\phi_s \mu_s \mu_t| &\le KnF(\lambda_m) \sum_{j=0}^{\infty} |\phi_j| \sum_{s=1}^{n} |\mu_s| \\ &\le KnF(\lambda_m) \sum_{s=1}^{n} |\mu_s|, \end{aligned} \tag{2.65}$$

where the first inequality follows from $\mu_t = O(F(\lambda_m))$ from Proposition 1 in Robinson (1994c).

Now, following Robinson (1994c),

$$\sum_{s=1}^{n} |\mu_s| = O\left(rF(\lambda_m) + n \max_{r<t\le n/2} |\mu_t|\right) \tag{2.66}$$

for $n/m < r < n/2$, where the first term on the right follows again from $\mu_t = O(F(\lambda_m))$, and, be Lemma 1(ii) of Robinson (1994c),

$$\max_{r<t\le n/2} |\mu_t| = O(r^{2(d_x - \frac{1}{2})}) \quad \text{as} \quad r \to \infty.$$

Choosing $r \sim nm^{(1-2d_x)/(2d_x-2)}$, which is indeed larger than n/m, yields the tightest bound for 2.66, i.e. $O(nF(\lambda_m)m^{(2d_x-1)/(2-2d_x)})$. It follows that 2.65 is $O(n^2F(\lambda_m)^2$ $m^{(2d_x-1)/(2-2d_x)})$ which is $o(n^2F(\lambda_m)^2)$ as required.

The first term in 2.58 is $O_p\left(n\sum_{t=1}^{n}\mu_t^2\right)$, which is proven in the same way to be $O(n^2F(\lambda_m)^2m^{2(2d_x-1)/(3-4d_x)})$, which is also $o(n^2F(\lambda_m)^2)$ as required.

The consistency result of Theorem 2 is sufficient for a number of applications, including consistent estimation of d_x, the long memory parameter which determines the extent of temporal dependence in the observable process x_t. However, in order to determine the scale of the hyperbolic pole of its spectral density in the neighbourhood of frequency zero, it is necessary to derive an upper bound for the rate of convergence of $\hat{F}(\lambda_m)/F(\lambda_m)$ to 1. To derive this upper bound, a smoothness assumption is added to Assumption B1 and a rate of decay is given for the ϕ_j, coefficients of the infinite moving average decomposition for the squares of the errors ε_t^2. This rate of decay controls the degree of temporal dependence in the squared error process.

Assumption C1 $f(\lambda)$ satisfies 1.74 with $0 < d_x < \frac{1}{2}$ and 2.10 holds.

Assumption C2 1.69 holds.

Assumption C3 Assumption B3 holds. In addition, when 2.56 does not hold, 1.46 does with $d_\varepsilon < \frac{1}{2}$.

Assumption C1 is weaker than Condition A' in Robinson (1994c) while 1.46 is needed to derive the upper bound 1.51 for the partial sums of squared innovations.

Theorem 3 Under Assumptions C1-C3,

$$\frac{\hat{F}(\lambda_m)}{F(\lambda_m)} - 1 = O_p\left(n^{d_\varepsilon - \frac{1}{2}} + \left(\frac{m}{n}\right)^\beta + m^{-\delta}\right) \quad \text{as } n \to \infty$$

for $\delta < (\frac{1}{2} - d_x)/(3 - 2d_x)$.

Proof The proof of Theorem 3 in Robinson (1994c) still applies to show that 2.61 is $O_p[((\frac{m}{n})^\beta + m^{-\delta})F(\lambda_m)]$. The proof of Theorem 1 established that 2.59 was $o_p(m^{\eta-\frac{1}{2}})$ for any $\eta > 0$, so that 2.59 is $O_p(m^{-\delta})$ for any $\delta < (\frac{1}{2} - d_x)/(3 - 2d_x)$ as $d_x \in (0, \frac{1}{2})$. As for 2.60, the first term in 2.62 is $O(n^{d_\varepsilon - \frac{1}{2}}F(\lambda_m))$ from 1.51, whereas the second is shown to be $O([m^{(2d_x-1)/(3-4d_x)} + m^{(d_x-\frac{1}{2})/(2-2d_x)}]F(\lambda_m))$ in the proof of Theorem 2. The result follows from the inequalities $(\frac{1}{2} - d_x)/(3 - 2d_x) < (\frac{1}{2} - d_x)/(1 - 2d_x) < (\frac{1}{2} - d_x)/(3 - 4d_x)$ for any $d_x \in (0, \frac{1}{2})$.

2.5 Estimation of long memory

An estimate based on the average periodogram statistic was proposed by Robinson (1994c)

$$\hat{d}_{xq} = 1 - \frac{\log\left(\hat{F}(q\lambda_m)/\hat{F}(\lambda_m)\right)}{2\log q}$$

motivated by the fact that $F(q\lambda)/F(\lambda) \sim q^{2(d_x-\frac{1}{2})}$ for any $q > 0$. From Theorem 1, (4.3) of Robinson (1994c) and Slutzky's Theorem, it is immediate that

<u>**Corollary 1**</u> Under Assumptions B1-B3,

$$\hat{d}_{xq} \to_p d_x \quad \text{as } n \to \infty, \quad \text{for any } q \in (0,1).$$

Under the additional requirement

$$m = O(n^{\gamma}), \quad 0 < \gamma < 1 \tag{2.67}$$

on the bandwidth, and

$$L(\lambda) = GM(\lambda), \quad G > 0, \tag{2.68}$$

and $M(\lambda)$ is a known function, a rate can be specified for the convergence of $\hat{d}_{xq}$ and G can be consistently estimated by

$$\hat{G}_q = \frac{2(\frac{1}{2} - \hat{d}_{xq})\hat{F}(\lambda_m)\lambda_m^{2(\hat{d}_{xq}-\frac{1}{2})}}{M(\lambda_m)}.$$

<u>**Corollary 2**</u> Under Assumptions C1-C3, 2.67 and 2.68,

$$\hat{d}_{xq} - d_x = O_p(n^{-\delta}) \quad \text{as } n \to \infty, \quad \text{for some } \delta > 0,$$

$$\hat{G}_q \to_p G \quad \text{as } n \to \infty, \quad \text{for any } q \in (0,1).$$

This estimate is particularly convenient for its great computational simplicity compared to the local Whittle estimate which is not defined in closed form. Its asymptotic distribution under Gaussianity (see Lobato and Robinson (1996)), however, is only normal when $0 < d_x < 1/4$ and, unlike that of the local Whittle estimate, it is not free of d_x. When $1/4 < d_x < 1/2$, its asymptotic distribution is influenced by the Rosenblatt process.

2.6 Finite sample investigation of the averaged periodogram long memory estimate

While the weak consistency result proposed for $\hat{d}_{xq}$ in Corollary 1 is valuable for the investigation of long financial time series, it is of interest to examine its relevance to series of more moderate length. Moreover, as Corollary 1 does not provide any limiting distributional result[2], it is important to provide simulation values for standard errors, and thereby to investigate the robustness of distributional results given by Lobato and Robinson (1996).

The finite sample results presented here do not consider the sensitivity of the estimate to the choice of the constant q ($q = 1/2$ is chosen arbitrarily), but concentrate on the sensitivity to conditional heteroscedasticity in the errors. Robustness to departures from finite fourth moment condition is also considered.

Finite sample performance of $\hat{d}_{xq}$ was examined under the presumption of no conditional heteroscedasticity (and indeed unconditional Gaussianity of the errors) in Lobato and Robinson (1996). We present here results of a Monte Carlo study of the averaged periodogram estimate applied to simulated series x_t following an ARFIMA(0,d_x,0) parametric version of 1.1 with conditionally Gaussian innovations ε_t (see 2.16) satisfying the following five models for the conditional variance σ_t^2:

(i) IID: $\sigma_t^2 = \sigma^2$. The ε_t are independent and identically distributed, so that there is no conditional heteroscedasticity. We can take $\sigma^2 = 1$ with no loss of generality.

[2]It is only conjectured that $\hat{d}_{xq}$ remains asymptotically normal for $0 < d_x < 1/4$.

(ii) ARCH: $\sigma_t^2 = .5 + .5\varepsilon_{t-1}^2$. The ε_t are ARCH(1) with modest autocorrelation in the ε_t^2; they satisfy 1.43 (Engle (1982)).

(iii) GARCH: $\sigma_t^2 = .05 + .5\varepsilon_{t-1}^2 + .45\sigma_{t-1}^2$. The ε_t are GARCH(1,1), with strong autocorrelation in the ε_t^2 at "short" lags (nearly IGARCH); they do not satisfy 1.43 (Bollerslev (1986)).

(iv) LMARCH: $\sigma_t^2 = \left\{1 - (1-L)^{.25}\right\}\varepsilon_t^2$. The ε_t have (moderate) long memory ARCH structure satisfying 1.28-1.33 and 1.38 with $a(z) = b(z) = 1$, so that the ε_t^2 follow the ARFIMA(0,d_ε,0) structure discussed in Section 4 of Robinson (1991b), with $d_\varepsilon = .25$.

(v) VLMARCH: $\sigma_t^2 = \left\{1 - (1-L)^{.45}\right\}\varepsilon_t^2$. The ε_t have "very long memory" ARCH structure, such that the ε_t^2 follow the same type of model as in (iv) but with $d_\varepsilon = .45$, close to the stationarity boundary.

So far as the ARFIMA(0,d_x,0) model for x_t is concerned, so that in relation to 1.1, $\sum_{j=0}^{\infty}\alpha_j L^j = (1-L)^{-d_x}$, we consider:

(a) "Moderate long memory": d_x=.2,

(b) "Very long memory": d_x=.45.

We study each of (i)-(v) with (a)-(b), covering a range of long/very long memory in x_t and a range of short/long memory in ε_t^2.

Tables 2.1-2.3 and 2.4-2.6 deal respectively with each of the two d_x values (a)-(b). In each case the results are based on n=64, 128 and 256 observations, with bandwidths m= $n/16$, $n/8$, $n/4$, and 10000 replications. In tables 2.1-2.2 and 2.4-2.5, we report, for the conditional variance specifications (i)-(v), Monte Carlo bias of the averaged periodogram estimate and Monte Carlo root mean squared error. In tables 2.3 and 2.6, we report the efficiency of the averaged periodogram estimate relative to the local Whittle estimate, that is the ratio of the Monte Carlo mean squared errors.

We make the comparison with the local Whittle estimate because it is extensively investigated in Chapter 3 and because asymptotic distributional theory is provided.

Table 2.1: Moderate long memory averaged periodogram biases

Monte Carlo BIASES for the averaged periodogram estimate of long memory applied to an ARFIMA(0, .2, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	-0.17	-0.10	-0.07	-0.10	-0.06	-0.04	-0.05	-0.03	-0.02
ARCH	-0.17	-0.11	-0.07	-0.09	-0.06	-0.05	-0.05	-0.04	-0.03
GARCH	-0.17	-0.11	-0.08	-0.10	-0.07	-0.05	-0.07	-0.05	-0.04
LMARCH	-0.18	-0.11	-0.07	-0.10	-0.06	-0.04	-0.05	-0.04	-0.03
VLMARCH	-0.17	-0.11	-0.08	-0.09	-0.06	-0.05	-0.06	-0.04	-0.03

Table 2.2: Moderate long memory averaged periodogram RMSEs

Monte Carlo ROOT MEAN SQUARED ERRORS for the averaged periodogram estimate of long memory applied to an ARFIMA(0, .2, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.41	0.25	0.17	0.26	0.16	0.11	0.16	0.11	0.07
ARCH	0.41	0.25	0.18	0.24	0.16	0.12	0.16	0.11	0.09
GARCH	0.40	0.28	0.20	0.26	0.20	0.15	0.19	0.15	0.12
LMARCH	0.42	0.26	0.17	0.25	0.17	0.12	0.16	0.12	0.08
VLMARCH	0.41	0.27	0.19	0.25	0.17	0.14	0.18	0.13	0.10

Table 2.3: Moderate long memory relative efficiencies

RELATIVE EFFICIENCY of the averaged periodogram compared to the local Whittle estimate of long memory applied to an ARFIMA(0, .2, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.86	0.92	1.00	1.00	0.88	1.00	0.88	1.00	1.00
ARCH	0.81	0.92	1.00	0.92	0.88	1.00	1.00	1.00	0.79
GARCH	0.85	1.00	1.00	0.93	1.00	1.00	1.00	0.87	0.84
LMARCH	0.82	0.93	0.89	0.92	1.00	1.00	0.88	1.00	1.00
VLMARCH	0.81	1.00	1.00	0.86	0.89	0.86	1.00	1.00	1.00

Table 2.4: Very long memory averaged periodogram biases

Monte Carlo BIASES for the averaged periodogram estimate of long memory applied to an ARFIMA(0, .45, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	-0.25	-0.18	-0.14	-0.16	-0.12	-0.09	-0.12	-0.09	-0.07
ARCH	-0.26	-0.19	-0.14	-0.17	-0.12	-0.10	-0.12	-0.09	-0.08
GARCH	-0.27	-0.19	-0.15	-0.17	-0.13	-0.11	-0.13	-0.10	-0.08
LMARCH	-0.26	-0.18	-0.14	-0.17	-0.12	-0.10	-0.12	-0.09	-0.07
VLMARCH	-0.26	-0.18	-0.14	-0.17	-0.13	-0.10	-0.12	-0.09	-0.08

Table 2.5: Very long memory averaged periodogram RMSEs

Monte Carlo ROOT MEAN SQUARED ERRORS for the averaged periodogram estimate of long memory applied to an ARFIMA(0, .45, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.40	0.26	0.18	0.25	0.16	0.12	0.16	0.11	0.09
ARCH	0.40	0.26	0.19	0.24	0.16	0.13	0.16	0.12	0.09
GARCH	0.41	0.27	0.21	0.26	0.19	0.15	0.18	0.14	0.12
LMARCH	0.41	0.26	0.19	0.25	0.17	0.13	0.16	0.12	0.09
VLMARCH	0.41	0.27	0.19	0.26	0.17	0.14	0.17	0.13	0.10

Table 2.6: Very long memory relative efficiencies

RELATIVE EFFICIENCY of the averaged periodogram compared to the local Whittle estimate of long memory applied to an ARFIMA(0, .45, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.90	0.78	0.60	0.77	0.66	0.56	0.56	0.52	0.44
ARCH	0.85	0.78	0.70	0.76	0.76	0.59	0.56	0.44	0.60
GARCH	0.86	0.86	0.65	0.78	0.62	0.53	0.60	0.51	0.44
LMARCH	0.87	0.78	0.62	0.70	0.58	0.47	0.66	0.44	0.44
VLMARCH	0.87	0.79	0.70	0.71	0.67	0.51	0.58	0.47	0.49

Indeed, Theorem 5 below proves asymptotic normality of the local Whittle estimate under moment conditions which include models (i), (iv) and (v), and the Monte Carlo results of Chapter 3 indicate that Theorem 5 continues to hold for models (ii) and (iii). In case of i.i.d. errors ε_t, the relative efficiency results reported are to be compared with theoretical ratios of asymptotic variances based on Theorem 5 for the local Whittle estimate on the one hand (i.e. $1/4m$ for all values of d_x), and Theorem 1 of Lobato and Robinson (1996) for the averaged periodogram estimate on the other hand. When $d_x = .2$, the ratio of asymptotic variances is $(1/4)/((3 - 2^{1.4})(.3)^2/(.2\log^2 2)) \simeq .74$. With the choice $q = .4$ (instead of $q = .5$ which is chosen here) which is the asymptotic variance minimizing value for q when $d_x = .2$ (see Lobato and Robinson (1996)), the theoretical relative efficiency is .76 instead.

The errors ε_t were sampled from a conditionally normal distribution (see 2.16) with conditional variance σ_t^2 in a recursive procedure with iid normal startup values subsequently discarded. Namely, for $t = -1000$ to 0, ε_t were generated as iid normal and σ_t^2 were identically set to one; and for $t = 1$ to $2n$, $\sigma_t^2 = \sigma^2 + P(L)\varepsilon_t^2$ and $\varepsilon_t = \sqrt{\sigma_t^2}\eta_t$, where η_t are iid normal and σ^2 and $P(L)$ are the relevant intercept and operator in cases (i) to (v), truncated to 1000 lags in the two long memory cases (iv) and (v). The Gauss random number generator RNDN was used with random seed starting at the value 12145389. A method based on the Cholevsky decomposition $(m_{i,j})_{i,j=1}^{2n}$ of the Toeplitz matrix $\left(\rho_{|i-j|}\right)_{i,j=1}^{2n}$, where ρ_j are the autocovariances of an ARFIMA(0,d_x,0), was then used to simulate x_t from the errors ε_t as $x_t = \sum_{i=1}^{t} m_{ti}\varepsilon_i$, $t = 1, \ldots, 2n$, the first n values being subsequently discarded. For each of the series simulated, the periodogram was computed by the Gauss Fast Fourier Transform algorithm.

Monte Carlo biases seem relatively unaffected by the model specification for the errors. Monte Carlo biases when the errors follow the near unit root GARCH process are largest (in absolute value) 9 times and tie largest 8 times out of the 18 d_x, n, m combinations, but the difference with other error specifications is very slight. In any case, the difference is more likely to be due to the effect of a near unit root rather than a failure of moment conditions 1.43 and 2.12. Other error specifications lead to almost identical performances to i.i.d. errors. The typical decay of biases with n is in line with the consistency theorem of the previous section while the decay of biases with m is rather more surprising. Biases are all negative and on average twice as large in case of very long memory ($d_x = .45$) than in case of moderately long memory (d_x=.2), with an indication that for very small values of m and n ($n = 64$ and $m = 4$), estimates of d_x are centered on 0 whatever the true value.

Monte Carlo MSEs seem hardly more affected by conditional heteroscedasticity than biases, apart from GARCH which again results in the worst performance. GARCH MSEs are largest 13 times and tie largest 4 times among the 18 d_x, n, m combinations. The discrepancy between GARCH and other cases increases with n and m, indicating that asymptotic behaviour may be significantly different as well. To investigate this point further, empirical distributions of estimates of the memory parameter are plotted for three different values of the sample size, $n = 500$, $n = 1000$ and $n = 2000$, and for two models for the errors, GARCH and i.i.d., to investigate the possible divergence of the two distributions. The empirical distributions of $1.7\sqrt{m}(\hat{d}_x - d_x)$ are plotted in figures 2.1 for $n = 500$, 2.2 for $n = 1000$ and 2.3 for $n = 2000$. In each graph, the empirical distributions of the estimate with GARCH errors and with i.i.d. errors are compared with the normal distribution function. There is clear indication of convergence of the empirical distribution with i.i.d. errors to the normal distribution function in accordance with asymptotic theory, and there is clear indication that the tails of the empirical distribution with GARCH errors become fatter with sample size, so that the latter is very unlikely to converge to the normal distribution function.

MSEs are very similar for the two reported values of d_x. They are identical in 22 cases and different by more than .01 only in 2 cases out of the 45 combinations of n, m and error models.

Finally, the relative efficiency of the averaged periodogram estimate of long memory compared to the local Whittle estimate for $d_x = .2$ is significantly larger in small samples than would be expected from the theoretical value in the i.i.d. case. The average periodogram even performs equally well as the local Whittle in 23 cases.

Figure 2.1: Averaged periodogram empirical distribution for $n = 500$

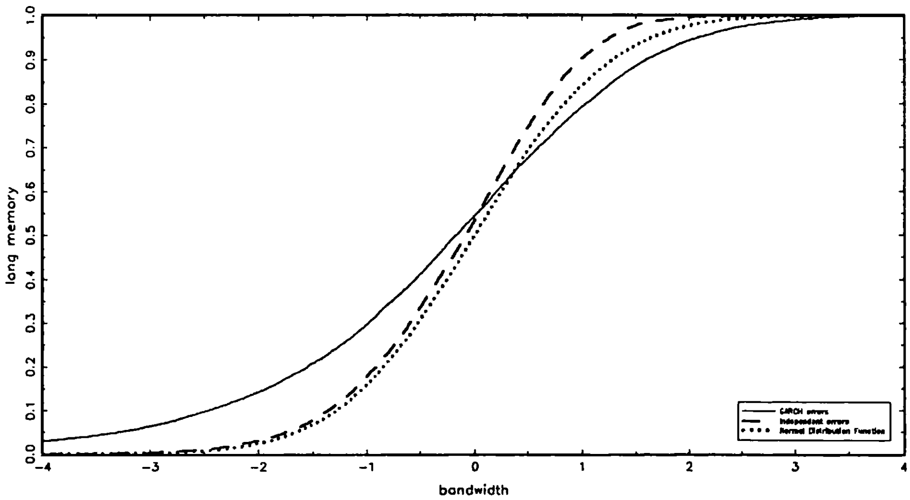


Figure 2.2: Averaged periodogram empirical distribution for $n = 1000$

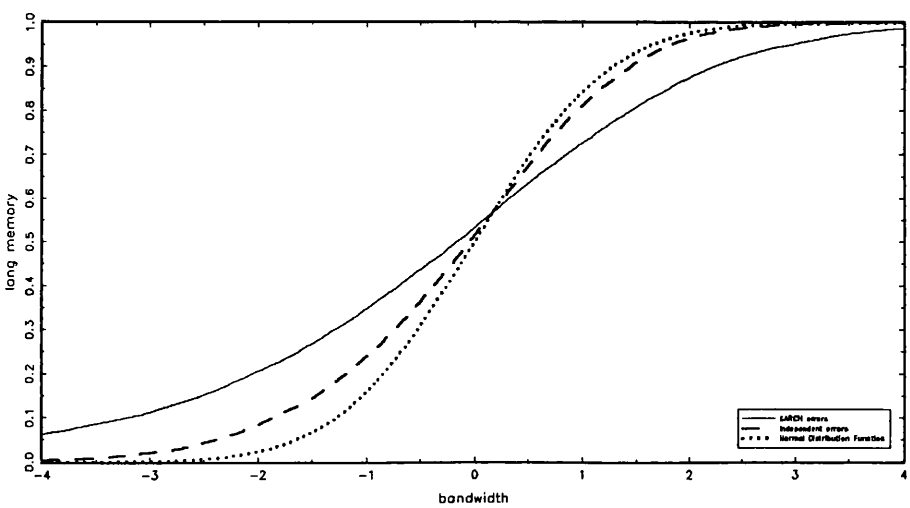


Figure 2.3: Averaged periodogram empirical distribution for $n = 2000$

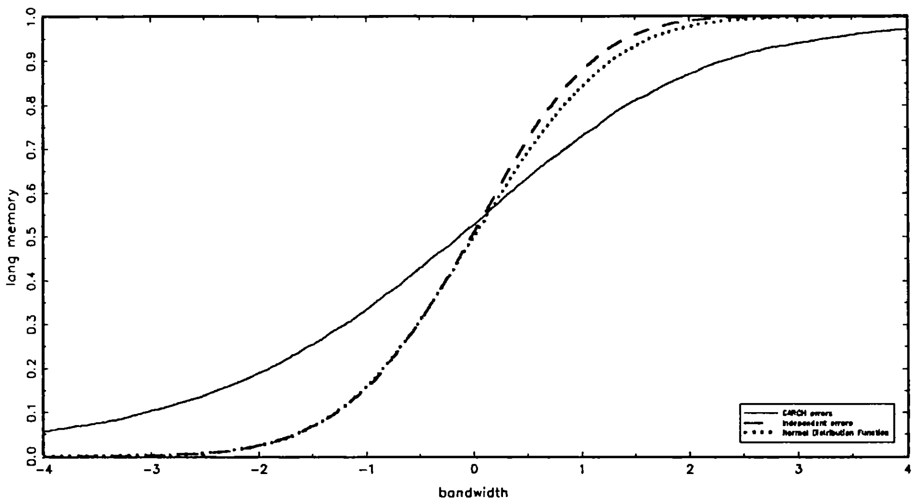


Table 2.7: Averaged periodogram relative efficiencies for larger sample sizes

RELATIVE EFFICIENCY of the averaged periodogram compared to the local Whittle estimate of long memory applied to and ARFIMA(0,d_x,0) series with five specified error structures, sample size 1000 and bandwidth 250.

MODEL	IID	ARCH	GARCH	LMARCH	VLMARCH
$d_x = .2$	0.62	0.74	0.82	0.71	0.78
$d_x = .45$	0.16	0.24	0.43	0.21	0.32

The relative performance is always at least 5% higher than the known theoretical value for the i.i.d. case. There is no evidence of a worsening of relative efficiency with sample size for the sample sizes reported. To see whether this pattern persists, relative efficiencies are reported also for a sample size of $n = 1000$ in table 2.7 and one observes that in the case $d_x = .2$, relative efficiencies become close to the theoretical value given for i.i.d. errors, conditional heteroscedasticity appearing to have again no significant effect. In no cases does the model chosen for the errors seem to influence relative efficiency, supporting the conjecture that the asymptotic normality result given in Lobato and Robinson (1996) for $0 < d_x < \frac{1}{4}$ continues to hold when the errors are conditionally heteroscedastic. For $d_x = .45$, where the asymptotic distribution is non standard even for i.i.d. errors (see Lobato and Robinson (1996)), relative efficiency of the average periodogram decreases steadily with sample size and bandwidth and reaches values as low as 16% for a sample size of $n = 1000$.

2.7 Estimation of stationary cointegration

Another main application of the averaged periodogram statistic in a long memory environment is the estimation of stationary cointegrating relationships. A large strand of literature has recently investigated the possibility of long run relationships between nonstationary variables. In the equation

$$y_t = \beta z_t + x_t,$$

y_t and z_t were typically assumed to have a unit root, and x_t to be weakly dependent. This notion of cointegration was extended by Granger (1987) to cases where x_t can be a long memory process, and prompted a new type of investigation of the Long Run Purchasing Power Parity Hypothesis (see e.g. Cheung and Lai (1993)). For stationary y_t and z_t series, a notion of cointegration can be defined as in Robinson (1994c) when all three variables are possibly long memory covariance stationary time series, providing x_t has a lesser degree of dependence. A similar idea was proposed by Gouriéroux and Peaucelle (1991) to identify codependence between moving average series in the sense that a linear combination of the original variables has a lower moving average order.

However, ordinary least squares estimation of the fractional cointegration coefficient β is inconsistent in the stationary case unless z_t and x_t are orthogonal. Robinson (1994c) pointed out that one should envisage the equilibrium relation between y_t and z_t at high lags only, and carry out the regression in the frequency domain on a degenerating band of frequencies. We assume that y_t and z_t follow the same assumptions as x_t with long memory parameters d_{xy} and d_{xz} such that $0 \leq d_x < d_{xy} \leq d_{xz} < 1/2$, and we denote by

$$I_{yz}(\lambda) = w_y(\lambda)\bar{w}_z(\lambda) \tag{2.69}$$

the cross periodogram of y_t and z_t, and by

$$\hat{F}_{yz}(\lambda) = \frac{2\pi}{n} \sum_{j=1}^{[n\lambda/2\pi]} I_{yz}(\lambda) \tag{2.70}$$

the dicretely averaged cross periodogram of y_t and z_t. If, moreover, $\hat{F}_z(\lambda)$ denotes the discretely averaged periodogram of z_t,

$$\hat{\beta} = \frac{\mathcal{R}e(\hat{F}_{yz}(\lambda_m))}{\hat{F}_z(\lambda_m)}$$

is the low-pass frequency domain least squares estimate of β. More precisely, we make the following assumptions on the process z_t:

<u>Assumption B1'</u> z_t has spectral density $f_z(\lambda)$ satisfying

$$f_z(\lambda) \sim L_z(\lambda)\lambda^{-2d_{x_z}} \text{ as } \lambda \to 0^+ \quad \text{with } d_x < d_{x_z} < \frac{1}{2}$$

and $L_z(\lambda)$ is a slowly varying function at 0.

<u>Assumption B3'</u>

$$z_t = \sigma_z^2 + \sum_{j=0}^{\infty} \alpha_{z_j} \varepsilon_{z_{t-j}}$$

and the conditions in Assumption B3 are satisfied when σ^2, α_j, ε_t, ψ_j and ϕ_j are replaced by σ_z^2, α_{z_j}, ε_{z_t}, ψ_{z_j} and ϕ_{z_j} respectively.

We can now state the following Corollary to Theorem 2:

<u>**Corollary 3**</u> Under Assumptions B1, B1', B2, B3 and B3',

$$\hat{\beta} \to_p \beta \quad \text{as} \quad n \to \infty.$$

<u>Proof</u> As pointed out by Robinson (1994c),

$$|\hat{\beta} - \beta| \leq \left| \frac{\mathcal{R}e(\hat{F}_{xz}(\lambda_m))}{\hat{F}_z(\lambda_m)} \right| \leq \left| \frac{\hat{F}_{xz}(\lambda_m)}{\hat{F}_z(\lambda_m)} \right| \leq \left\{ \frac{\hat{F}(\lambda_m)}{\hat{F}_z(\lambda_m)} \right\}^{\frac{1}{2}}$$

by the Cauchy inequality. The result is therefore an immediate consequence of Theorem 2 applied to both x_t and z_t.

This result will be shown in Chapter 5 to be valuable for the investigation of common persistence patterns in the conditional variances of financial returns.

2.8 Conclusion

This chapter has focused on the simple averaged periodogram statistic insofar as it provides insights into the dependence structure of a time series. It was shown in this chapter that the averaged periodogram is an asymptotically normal estimate of the spectral density at zero frequency for a weakly dependent process. The conditions of this theorem do not presume anything on the short memory structure, and, more importantly allow for a very high degree of temporal dependence (including long memory) in the conditional variance and higher moments. It is conjectured that this property continues to hold when one focuses on non zero frequencies where the spectral density is continuous. However, to gain insights into the short memory structure of the process, functional estimation of the spectral density would then be required, and global conditions on the smoothness of the spectral density would have to be imposed. This chapter has also extended the applicability of Robinson (1994c)'s results on the averaged periodogram statistic in the presence of long range dependence. It has been shown that all the results in Robinson (1994c) still hold for long memory processes with (possibly long memory) conditionally heteroscedastic errors. This permits consistent estimation of long memory and stationary cointegration which proves especially useful in the investigation of dependence and codependence in foreign exchange rate returns analysed in chapter 5.

Chapter 3

Local Whittle estimation of long memory with conditional heteroscedasticity

Joint work with Peter Robinson

3.1 Introduction

This chapter is concerned with the Gaussian, or local Whittle estimate of long memory proposed by Künsch (1987). Only recently has asymptotic distributional theory of this estimate been laid down by Robinson (1995a). Despite the efficiency improvements Robinson (1995b) made to the log periodogram estimate, the local Whittle remains the more efficient and its asymptotic variance, given by 1.75 is free of unknown parameters. Unlike the log periodogram estimate, it is not defined in closed form, but nonlinear optimisation is only needed with respect to a single parameter, d_x, and can be accomplished rapidly.

The asymptotic theory of Robinson(1995a,b) rules out the possibility of conditional heteroscedasticity, and this seems a drawback in case of financial series for which semiparametric estimates otherwise seem appropriate. Indeed, Robinson (1995b) analysed the log periodogram under the assumption that x_t is Gaussian, whereas for the Gaussian estimate he did not assume Gaussianity but 1.1 where the ε_t satisfy at least 1.25 and 1.26. This chapter is concerned with relaxing 1.26 in the derivation of the asymptotic distribution of the local Whittle estimate of long memory, to allow for the possibility of autocorrelation in the ε_t^2, for example in some financial applications, the levels x_t can be approximated by a martingale sequence (so $\alpha_j = 0,\ j > 0$) but the squares $x_t^2 = \varepsilon_t^2$ cannot, so that the sequence x_t is not a sequence of independent random variables. In fact, empirical evidence (discussed in chapter 1) has also suggested that dependence in the squares can fall off very slowly, in a way that is possibly more consistent with long memory than with standard short memory ARCH and GARCH specifications.

Here, again, we adopt specification 1.28 for the conditional variance σ_t^2 and consider the local Whittle estimate of d_x in these circumstances, partly because it is well motivated by superior efficiency properties under the previous conditions, and because the log periodogram estimate (and some others) are technically more complex and cumbersome to handle when Gaussianity is relaxed, due to their highly nonlinear structure.

The following section describes the Gaussian estimate of d_x. Because the estimate is of the implicitly defined extremum type, one has to establish consistency prior to deriving limiting distributional behaviour, and these tasks are carried out in Sections 3 and 4. Section 5 reports a Monte Carlo study of finite sample behaviour. Section 6 contains some concluding comments.

3.2 Local Whittle estimate

On the basis of observations x_t, $t = 1, \ldots, n$, define the periodogram $I_x(\lambda)$ as in 1.53. Again, let $\lambda_j = 2\pi j/n$ and consider the discrete local likelihood defined in 1.73. for $0 < m < [n/2]$. As indicated by Robinson (1995a), we can concentrate out

G, and then consider estimating d_x by

$$\hat{d}_x = \text{argmin}_{\Delta_1 \le d \le \Delta_2} R(d), \tag{3.1}$$

where $-\frac{1}{2} < \Delta_1 < \Delta_2 < \frac{1}{2}$ and

$$R(d) = \log\left\{\frac{1}{m}\sum_{j=1}^{m}\frac{I_x(\lambda_j)}{\lambda_j^{2d}}\right\} - (2d)\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j. \tag{3.2}$$

For $m = [(n-1)/2]$, $\hat{d}_x$ is a form of Gaussian estimate under the parametric model $f(\lambda) = G|\lambda|^{-2d_x}$, all $\lambda \in (-\pi, \pi]$, and its asymptotic properties would be approximately covered by Fox and Taqqu (1986), Giraitis and Surgailis (1990) and others, under Gaussianity, or more generally the assumption that x_t is linear with independent and identically distributed innovations. (These authors considered continuous, rather than discrete, averaging over frequencies.) When $m < [n/2]$ such that 1.69 is satisfied, $\hat{d}_x$ can be viewed as a semiparametric estimate based on 1.18 with $0 < L(\lambda) = G < \infty$. Under 1.1, 1.25 and 1.26, and other regularity conditions, Robinson (1995a) showed that $\hat{d}_x$ is consistent for d_x, and under further conditions that 1.75 holds (see also Taqqu and Teverovsky (1995b) for a discussion of this result). The bandwidth parameter m is analogous to that employed in weighted periodogram estimates of the spectral density of short memory processes. Clearly 1.69 is a minimal requirement for consistency under 1.18. Chapter 4 discusses optimal choices of m in the determination of $\hat{d}_x$.

The compact set $[\Delta_1, \Delta_2]$ of admissible d values in Robinson (1995a) can include ones between 0 and $\frac{1}{2}$, where there is long memory, ones between $-\frac{1}{2}$ and 0, where there is negative dependence or antipersistence, and $d = 0$, where there is short memory. It seems desirable to avoid assuming, say, $0 < d_x < \frac{1}{2}$, a priori, but rather to allow also for the possibility that $d_x \le 0$, especially in view of the very mixed evidence of the existence of long memory in levels of financial series (see, e.g. Lo (1991), Lee and Robinson (1996)), in view of the efficient markets hypothesis, and in view of the possibility that log price levels may be nonstationary with less than a unit root, in which case returns can exhibit negative dependence (as will be seen in chapter 5). By contrast, the bulk of asymptotic theory relevant to long memory assumes a priori that long memory exists.

It turns out that not only is $\hat{d}_x$ still consistent for d_x in the presence of the (possibly long memory) ARCH behaviour described in the previous section (although with stronger moment conditions), but 1.75 holds in detail with the same asymptotic variance, so that no features of the ARCH structure defined by 1.28 or 1.33 enter (see discussion of this point in chapter 1). Our derivation of the asymptotic properties of $\hat{d}_x$ follows the main steps of the proof in Robinson (1995a), and uses a number of properties established there, but it also differs significantly, posing new challenges, primarily the control of fourth cumulants and the convergence of partial sums of conditional variances of the approximating martingale.

3.3 Consistency of the local Whittle estimate

We introduce the following assumptions.

Assumption D1 For $d_x \in [\Delta_1, \Delta_2]$, $-\frac{1}{2} < \Delta_1 < \Delta_2 < \frac{1}{2}$, and $0 < L(\lambda) = G < \infty$, $f(\lambda)$ satisfies 1.18. In addition, in a neighbourhood $(0, \delta)$ of the origin, $f(\lambda)$ is differentiable and

$$\frac{d}{d\lambda} \log f(\lambda) = O(\lambda^{-1}) \quad \text{as} \quad \lambda \to 0^+. \tag{3.3}$$

Assumption D2 Assumption B2 holds.

Assumption D3 Assumption B3 holds.

Assumptions D1 and D2 are identical to Assumptions A1, A2 and A4 of Robinson (1995a). We stress that only local (to zero) assumptions are made on $f(\lambda)$, so that it need not be smooth, or even bounded (or nonzero) outside a neighbourhood of the origin. In place of the current Assumption A3, Robinson (1995a) assumed 1.1, 1.25 and 1.26 with a homogeneity condition, so that we require more moments while allowing for ARCH behaviour, possibly with long memory. The discussion of Assumptions B1 to B3 applies.

Theorem 4 Under Assumptions D1-D3,

$$\hat{d}_x \to_p d_x, \quad \text{as } n \to \infty.$$

Proof The main part of the proof of the corresponding Theorem 1 of Robinson (1995a) applies except for the proof that

$$\sum_{r=1}^{m-1} \left(\frac{r}{m}\right)^{2(\Delta-d_x)+1} \frac{1}{r^2} \left| \sum_{j=1}^{r} \left(2\pi I_\varepsilon(\lambda_j) - \sigma^2\right) \right| \to_p 0, \tag{3.4}$$

where

$$I_\varepsilon(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} \varepsilon_t e^{it\lambda} \right|^2$$

and $\Delta = \Delta_1$ when $d_x < \frac{1}{2} + \Delta_1$ and $\Delta \in (d_x - \frac{1}{2}, d_x]$ otherwise. (Note that unlike in Robinson (1995a), we take the unconditional variance of ε_t to be σ^2, not unity.)

The justification for the above claim rests on the fact that the remainder of the aforementioned proof depends only on unconditional second moment properties. 2.57 holds under the present conditions, so that in view of (3.18) of Robinson (1995a), 3.4 is implied if

$$\sum_{\substack{s \neq t \\ 1}}^{n} \varepsilon_s \varepsilon_t A_{st}^{(r)} = o_p(r^{1-\eta} n), \text{ some } \eta > 0, \tag{3.5}$$

uniformly in $r \in [1, m-1]$, where $A_{st}^{(r)} = \sum_{j=1}^{r} \cos\left[(s-t)\lambda_j\right]$. To prove 3.5, the left hand side has variance

$$4E\left(\sum_{\substack{u<v \\ 1}}^{n} \sum_{\substack{s<t \\ 1}}^{n} \varepsilon_s \varepsilon_t \varepsilon_u \varepsilon_v A_{st}^{(r)} A_{uv}^{(r)}\right). \tag{3.6}$$

In view of 1.25 of Assumption D3, it is clear that no summands for which $t \neq v$ can contribute. Thus, 3.6 is

$$4E\left(\sum_{\substack{s<t \\ 1}}^{n} \varepsilon_t^2 \varepsilon_s^2 A_{st}^{(r)2}\right) + 8E\left(\sum_{\substack{u<s<t \\ 1}}^{n} \varepsilon_t^2 \varepsilon_s \varepsilon_u A_{st}^{(r)} A_{ut}^{(r)}\right). \tag{3.7}$$

The first term in 3.7 is bounded by

$$4 \max_t E(\varepsilon_t^4) \sum_{\substack{s<t \\ 1}}^{n} A_{st}^{(r)2} = O(rn^2),$$

from (3.20) of Robinson (1995a). Substituting 1.33 in the second term of 3.7 gives

$$\begin{aligned} 8E\left(\sum_{\substack{u<s<t \\ 1}}^{n}\left(\sigma^2 + \sum_{j=0}^{\infty}\phi_j \nu_{t-j}\right)\varepsilon_u \varepsilon_s A_{st}^{(r)} A_{ut}^{(r)}\right) \\ = 8\sum_{\substack{u<s<t \\ 1}}^{n} \phi_{t-s} E\left(\nu_s \varepsilon_u \varepsilon_s\right) A_{st}^{(r)} A_{ut}^{(r)} \\ = 8\sum_{\substack{u<s<t \\ 1}}^{n} \phi_{t-s} E\left(\varepsilon_s^3 \varepsilon_u\right) A_{st}^{(r)} A_{ut}^{(r)}, \end{aligned}$$

where the first equality applies nested conditional expectations and 1.25 for $j > t-s$, and 1.35 for $j < t-s$, whereas the second equality employs 1.34 and nested conditional expectations with 1.25 to verify $E(\sigma_s^2 \varepsilon_s \varepsilon_u) = 0$ for $u < s$. Under 2.55, this is identically zero. Under 2.56, it is bounded in absolute value by

$$8\left|\sum_{t=3}^{n}\sum_{s=2}^{t-1}\sum_{u=1}^{s-1} \phi_{t-s} E(\varepsilon_t^3 \varepsilon_u) A_{st}^{(r)} A_{ut}^{(r)}\right|.$$

$|A_{st}^{(r)}| \leq r$ from Robinson (1995a), so the quantity above is bounded by

$$8r \max_t E(\varepsilon_t^4) \sum_{t=3}^{n}\sum_{s=2}^{t-1}\sum_{u=1}^{s-1} |\phi_{t-s} A_{ut}^{(r)}|.$$

By 1.43, this is further bounded by

$$\begin{aligned} Kr\sum_{t=3}^{n}\sum_{s=2}^{t-1}\sum_{u=1}^{t-s-1} |\phi_s A_{ut}^{(r)}| &\leq Kr\sum_{j=0}^{\infty}|\phi_j| \sum_{\substack{s<t \\ 1}}^{n} \left|A_{st}^{(r)}\right| \\ &\leq Krn\left(\sum_{\substack{s<t \\ 1}}^{n} A_{st}^{(r)2}\right)^{\frac{1}{2}} = O\left(r^{\frac{3}{2}} n^2\right) \end{aligned}$$

because $\sum_{s<t}^{n} A_{st}^{(r)2} = O(rn^2)$ is proven in Robinson (1995a). Thus, 3.5 is verified.

As explained by Robinson (1995a), there is a lack of uniformity in the convergence of $R(d)$ around $d = d_x - \frac{1}{2}$ which is of concern when $d_x \geq \frac{1}{2} + \Delta$, and then one has to show also that

$$\frac{1}{m}\sum_{j=1}^{m}(a_j - 1)\left(2\pi I_\varepsilon(\lambda_j) - \sigma^2\right) \rightarrow_p 0 \tag{3.8}$$

where $a_j = (\frac{j}{p})^{2(\Delta - d_x)}$ for $1 \leq j \leq p$, and $a_j = (\frac{j}{p})^{2(\Delta_1 - d_x)}$ for $p < j \leq m$, where $p = \exp\left(\frac{1}{m}\sum_{j=1}^m \log j\right)$. We have

$$\begin{aligned}
2\pi I_\varepsilon(\lambda_j) - \sigma^2 &= \frac{1}{n}\sum_{t,s=1}^{n} \varepsilon_t\varepsilon_s \cos(s-t)\lambda_j - \sigma^2 \\
&= \frac{1}{n}\sum_{t=1}^{n}(\varepsilon_t^2 - \sigma^2) + \frac{1}{n}\sum_{\substack{s\neq t \\ 1}}^{n} \varepsilon_t\varepsilon_s \cos(s-t)\lambda_j. \qquad (3.9)
\end{aligned}$$

The contribution of the first term to the left-hand side of 3.8 is equal to

$$\left\{\frac{1}{m}\sum_{j=1}^{m}(a_j - 1)\right\}\left\{\frac{1}{n}\sum_{t=1}^{n}(\varepsilon_t - \sigma^2)\right\}.$$

The first term in brackets is proven to be $O(1)$ in Robinson (1995a), whereas the second term in brackets is $o(1)$ because 2.57 holds under the present conditions. The contribution to 3.8 from the second term in 3.9 is given by

$$2\frac{1}{m}\sum_{j=1}^{m}(a_j - 1)\frac{1}{n}\sum_{\substack{t>s \\ 1}}^{n} \varepsilon_t\varepsilon_s \cos(s-t)\lambda_j.$$

Call $B_{st} = \sum_{j=1}^m (a_j - 1)\cos(s-t)\lambda_j$. We need to prove that

$$\sum_{\substack{t>s \\ 1}}^{n} \varepsilon_s\varepsilon_t B_{st} = o_p(nm).$$

Now

$$E\left(\sum_{\substack{t>s \\ 1}}^{n} \varepsilon_s\varepsilon_t B_{st}\right) = \sum_{\substack{t>s \\ 1}}^{n} E(\varepsilon_s^2\varepsilon_t^2)B_{st}^2 + \sum{}' E(\varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_v)B_{st}B_{uv} \qquad (3.10)$$

where the summation $\sum'$ is over s, t, u, v such that $s < t$ and $u < v$ and $u \neq s$ or $t \neq v$. The first term on the right-hand side of 3.10 is $o(mn^2)$ from 1.43 and the property derived in Robinson (1995a) that

$$\sum_{s<t} B_{st}^2 = o(mn^2). \qquad (3.11)$$

The second term on the right-hand side of 3.10 has possibly non zero contributions from summands such that $t = u$ only, in view of 1.25. Now as shown in the previous section with $A_{st}^{(r)}$ replaced by B_{st},

$$\sum_{t>s>u} E(\varepsilon_t^2\varepsilon_s\varepsilon_u)B_{st}B_{ut} = \sum_{u<s<t} \phi_{t-s}E(\varepsilon_s^3\varepsilon_u)B_{st}B_{ut}$$

which is zero under 2.55 and which is bounded in absolute value under 2.56 by

$$\left|\sum_{t=3}^{n}\sum_{s=2}^{t-1}\sum_{u=1}^{s-1}\phi_{t-s}E(\varepsilon_t^3\varepsilon_u)B_{st}B_{ut}\right|.$$

$|B_{st}| \leq m$, so that the quantity above is bounded by

$$m\max_t E(\varepsilon_t^4)\sum_{t=3}^{n}\sum_{s=2}^{t-1}\sum_{u=1}^{s-1}|\phi_{t-s}B_{ut}|.$$

By 1.43, this is further bounded by

$$\begin{aligned} Km\sum_{t=3}^{n}\sum_{s=2}^{t-1}\sum_{u=1}^{t-s-1}|\phi_s B_{ut}| &\leq Km\sum_{j=0}^{\infty}|\phi_j|\sum_{1\leq s<t}^{n}|B_{st}| \\ &\leq Kmn\left(\sum_{1\leq s<t}^{n}B_{st}^2\right)^{\frac{1}{2}} = O\left(m^{\frac{3}{2}}n^2\right) \end{aligned}$$

in view of 3.11, which serves to establish 3.8 under Assumption D3.

3.4 Asymptotic normality of the local Whittle estimate

The limiting distributional properties of $\hat{d}_x$ rest on stronger conditions than those sufficient for consistency.

Assumption E1 For some $\beta \in (0,2]$, $f(\lambda)$ satisfies 1.74 with $d_x \in [\Delta_1, \Delta_2]$. In addition, in a neighbourhood $(0,\delta)$ of the origin, $\alpha(\lambda)$ is differentiable and

$$\frac{d}{d\lambda}\log\alpha(\lambda) = O\left(\frac{|\alpha(\lambda)|}{\lambda}\right) \quad \text{as } \lambda \to 0^+,$$

where $\alpha(\lambda) = \sum_{j=0}^{\infty}\alpha_j e^{ij\lambda}$.

Assumption E2 As $n \to \infty$

$$\frac{1}{m} + \frac{m^{1+2\beta}(\log m)^2}{n^{2\beta}} + \frac{(m\log m)^2}{n} \to 0, \tag{3.12}$$

and, if 2.56 does not hold, 1.82 holds with d_ε defined by 1.46.

Assumption E3 Assumption A3 holds with 1.15.

Compared to the corresponding assumptions in Robinson (1995a), Assumptions E1 is unchanged (still restricting $f(\lambda)$ only near the origin, such that β indicates the smoothness of $f(\lambda)/G\lambda^{-2d_x}$ there), but Assumptions E2 and E3 trade off the relaxation of constant conditional innovations variances and fourth moments with some strengthening of conditions. The eighth moment condition 2.12 replaces the fourth moment condition of Robinson (1995a), while, when there is long memory in the ε_t^2, extension of 2.55 to the constancy of the first three odd conditional moments 2.13 is again satisfied in case 2.16. The strengthening of moment conditions is a matter both of practical concern, in view of the characteristics of much financial data, and of theoretical concern in view of the results of Engle (1982), Bollerslev (1986), Nelson (1990b), for example. As with Theorem 4, it is likely that Theorem 5 below can be established under a milder moment condition by a more detailed argument. Condition 1.46 strengthens 1.40 while being satisfied in the examples 1.11 and 1.41. d_ε can be arbitrarily close to $\frac{1}{2}$, so that 1.46 in not of great concern in itself, except that 1.82 strengthens 3.12 unless $d_\varepsilon \leq (1-2\beta)/(4\beta+2)$, which is possible only when $\beta < \frac{1}{2}$ is chosen in 3.12, whereas when the levels x_t themselves have fractional noise or ARFIMA long memory (analogous to models 1.10 and 1.11 for ε_t^2), $\beta = 2$ is available in Assumption E1. In 3.12, the requirement $(m \log m)^2/n \to 0$ was not in Robinson (1995a), but it does not bind when $\beta \leq \frac{1}{2}$. Fractional noise and ARFIMA x_t satisfy 1.15, which is consistent with Assumption E1, and also satisfy the quasi-monotonicity assumption on the α_j, which entails 2.17. In fact, we believe that this requirement, and 1.82, could be removed or relaxed by a more detailed proof, but the quasi-monotonicity requirement does not seem very onerous, while 1.82 is also needed when the ε_t^2 have long memory, and there always exists an m sequence satisfying both 3.12 and 1.82.

<u>Theorem 5</u> Under Assumptions E1-E3, 1.75 holds.

<u>Proof</u> Again, the basic structure of the proof of Robinson (1995a) is unchanged, and a number of properties established there are still of use. Again a mean value theorem argument is applied, and the scores approximated by a martingale. The

approximation, and the treatment of second derivatives of $R(d)$, are affected by the changed conditions, but we postpone discussion of this until after we have established the asymptotic normality of the approximating martingale, whose proof is considerably affected.

Let $z_t = \varepsilon_t \xi_t$, where $\xi_t = \sum_{s=1}^{t-1} \varepsilon_s c_{t-s}$ with $c_s = (2/nm^{\frac{1}{2}}) \sum_{j=1}^{m} \nu_j \cos s\lambda_j$, $\nu_j = \log j - m^{-1} \sum_{j=1}^{m} \log j$, so $\sum_{t=2}^{n} z_t$ is a martingale, and, as in Robinson (1995a), we wish to show that as $n \to \infty$, 2.33 and 2.34 hold. By the Schwarz inequality, $E(z_t^4) \le (E\varepsilon_t^8)^{\frac{1}{2}} (E\xi_t^8)^{\frac{1}{2}}$. Because ξ_t is a martingale, by Burkholder's inequality (Burkholder (1973)),

$$E\left(\xi_t^8\right) \le KE\Big(\sum_{s=1}^{t-1} c_{t-s}^2 \varepsilon_s^2\Big)^4 \le \max_s E\varepsilon_s^8 \Big(\sum_{s=1}^{n} c_s^2\Big)^4 = O\left((\log m)^8/n^4\right) \qquad (3.13)$$

uniformly in t by (4.22) of Robinson (1995a). Thus,

$$\sum_{t=1}^{n} E\left(z_t^4\right) \le K\frac{(\log m)^4}{n} \to 0 \qquad (3.14)$$

to verify 2.33. To check 2.34, write

$$E\left(z_t^2 | \mathcal{F}_{t-1}\right) = \sigma_t^2 \xi_t^2 = \sigma^2 \xi_t^2 + (\sigma_t^2 - \sigma^2)\xi_t^2. \qquad (3.15)$$

From (4.14) and (4.15) of Robinson (1995a),

$$\sum_{t=1}^{n} \xi_t^2 - \sigma^2 = \sum_{t=1}^{n-1} \chi_t r_{n-t} + \sigma^2 \left\{\sum_{t=1}^{n-1} r_{n-t} - 1\right\} + \sum_{t=2}^{n} \sum_{v \neq s} \varepsilon_v \varepsilon_s c_{t-v} c_{t-s}, \qquad (3.16)$$

writing $\chi_t = \varepsilon_t^2 - \sigma^2$ and $r_t = c_1^2 + \ldots + c_t^2$. The first term on the right has mean zero and variance

$$\sum_{t=1}^{n-1} \sum_{u=1}^{n-1} \gamma_{t-u} r_{n-t} r_{n-u}, \qquad (3.17)$$

where $\gamma_j = \mathrm{cov}(\varepsilon_t^2, \varepsilon_{t+j}^2)$. Now

$$|\gamma_j| \le K\Phi_0\Phi_j \to 0, \quad \text{as } j \to \infty \qquad (3.18)$$

by 2.12 and 1.46 and

$$\sum_{t=1}^{n-1} r_{n-t} \to 1, \quad \text{as } n \to \infty \qquad (3.19)$$

established by Robinson (1995a). It follows from the Toeplitz lemma that 3.17 tends to zero. Clearly, the second term in 3.16 thus tends to zero, whereas the last term has mean zero and variance bounded by

$$2\left(\max_t E\varepsilon_t^4\right)\sum_{t,u=2}^{n}\sum_{\substack{v\neq s\\ 1}}^{\min(t-1,u-1)}\left|c_{t-v}c_{t-s}c_{u-v}c_{u-s}\right|. \tag{3.20}$$

This follows from the corresponding derivation in Robinson (1995a), but upper bounding $E(\varepsilon_t^2\varepsilon_s^2)$ by the Schwarz inequality. The absolute value did not arise in Robinson (1995a) but it is clear from his derivation that the bound established there applies to 3.20, namely $O\left((\log m)^4(n^{-1}+m^{-1/3})\right)\to 0$. It remains to show that

$$\sum_{t=1}^{n}(\sigma_t^2-\sigma^2)\xi_t^2\to_p 0.$$

The left side is

$$\sigma^2\sum_{t=1}^{n}\left(\sigma_t^2-\sigma^2\right)r_{t-1}+\sum_{t=1}^{n}\left(\sigma_t^2-\sigma^2\right)\sum_{s=1}^{t-1}c_{t-s}^2\chi_s+\sum_{t=1}^{n}\left(\sigma_t^2-\sigma^2\right)\sum_{\substack{v\neq s\\ 1}}^{t-1}\varepsilon_v\varepsilon_s c_{t-v}c_{t-s}. \tag{3.21}$$

The first term is

$$\sigma^2\sum_{t=2}^{n}\sum_{j=1}^{\infty}\psi_j\chi_{t-j}r_{t-1}=\sigma^2\left(S_1+S_2\right),$$

where

$$S_1=\sum_{j=1-n}^{n-1}\chi_j\sum_{t=1}^{n-1}r_t\psi_{t-j+1},\quad S_2=\sum_{j=-\infty}^{-n}\chi_j\sum_{t=1}^{n-1}r_t\psi_{t-j+1},$$

and $\psi_j=0,\ j\leq 0$. Now S_1 has mean zero and variance

$$\begin{aligned}\sum_{j,k=1-n}^{n-1}\gamma_{j-k}\sum_{s,t=1}^{n-1}r_s r_t\psi_{s-j+1}\psi_{t-k+1} &\leq Knr_{n-1}^2\Big(\sum_{j=1}^{\infty}|\psi_j|\Big)^2\sum_{j=0}^{2n-2}|\gamma_j|\\ &= O\left(\frac{(\log m)^8}{n^{1-2d}}\right)\to 0,\quad \text{as } n\to\infty,\end{aligned}$$

using 1.42, 2.40 and $r_n=O((\log m)^4/n)$, which was established by Robinson (1995a). On the other hand

$$E\left|S_2\right|\leq K\sum_{t=1}^{n-1}r_t\sum_{j=n}^{\infty}|\psi_j|\to 0,\quad \text{as } n\to\infty,$$

from 1.42 and 2.39, so that the first term in 3.21 is $o_p(1)$. The second term in 3.21 is

$$\sum_{t=2}^{n}\sum_{v=-\infty}^{0}\psi_{t-v}\sum_{s=1}^{t-1}c_{t-s}^2\chi_v\chi_s \tag{3.22}$$

$$+\sum_{t=2}^{n}\sum_{v=1}^{t-1}\psi_{t-v}\sum_{s=1}^{t-1}c_{t-s}^2\chi_v\chi_s. \tag{3.23}$$

The expectation of the absolute value of 3.22 is bounded by

$$K\left(\max_t E\varepsilon_t^4\right)\sum_{t=2}^{n}\sum_{j=t}^{\infty}|\psi_j|r_{t-1}\to 0$$

using 1.42, 2.39 and the Toeplitz lemma. 3.23 includes the component

$$\sum_{t=2}^{n}\sum_{s=1}^{t-1}\psi_{t-s}c_{t-s}^2\chi_s^2,$$

whose absolute value has expectation which likewise tends to zero. The remainder of 3.23 can be written

$$\sum_{t=2}^{n}\sum_{v=1}^{t-1}\psi_{t-v}\sum_{s=1}^{v-1}c_{t-s}^2\chi_v\chi_s+\sum_{t=2}^{n}\sum_{v=1}^{t-1}\psi_{t-v}\sum_{s=v+1}^{t-1}c_{t-s}^2\chi_v\chi_s. \tag{3.24}$$

The first term in 3.24 has mean square

$$\sum_{t,u=2}^{n}\sum_{v=1}^{t-1}\psi_{t-v}\sum_{s=1}^{v-1}c_{t-s}^2\sum_{q=1}^{u-1}\psi_{u-q}\sum_{p=1}^{q-1}c_{u-p}^2E\left(\chi_v\chi_s\chi_q\chi_p\right). \tag{3.25}$$

Now each (v,s,q,p) such that $s<v$, $p<q$ satisfies one of the relations $v=q$, $s\le q<v$, $q<s<v$, $p\le v<q$ or $v<p<q$. The contribution from summands in 3.25 such that $v=q$ is bounded by

$$K\left(\max_t E\chi_t^4\right)\sum_{t,u=2}^{n}\sum_{v=1}^{\min(t-1,u-1)}|\psi_{t-v}\psi_{u-v}|\sum_{s=1}^{v-1}c_{t-s}^2\sum_{p=1}^{v-1}c_{u-p}^2$$
$$\le K\left(\max_t E\varepsilon_t^8\right)r_{n-1}^2n\Big(\sum_{j=1}^{\infty}|\psi_j|\Big)^2=O\left((\log m)^4/n\right)\to 0.$$

Next, for $v>q\ge s$, $p<q$,

$$E\left(\chi_v\chi_s\chi_q\chi_p\right)=E\left\{\sum_{j=-\infty}^{q}\phi_{v-j}\nu_j\chi_s\chi_q\chi_p\right\}, \tag{3.26}$$

because

$$E\left(\chi_v|\mathcal{F}_q\right)=\sum_{j=-\infty}^{q}\phi_{v-j}\nu_j,\quad \text{a.s.},\quad v>q, \tag{3.27}$$

as follows from 1.33 and

$$E\left(\nu_j|\mathcal{F}_q\right)=E\left(\varepsilon_j^2|\mathcal{F}_q\right)-E\left(E\left(\varepsilon_j^2|\mathcal{F}_{j-1}\right)|\mathcal{F}_q\right)=0,\ \text{ a.s.},\quad q<j.$$

Now 3.26 is bounded in absolute value by

$$\begin{aligned}
E\left|\Big(\sum_{j=-\infty}^{q}\phi_{v-j}\nu_j\Big)\chi_s\chi_q\chi_p\right| &\le \left\{E\left(\sum_{j=-\infty}^{q}\phi_{v-j}\nu_j\right)^4\left(\max_t E\chi_t^4\right)^3\right\}^{\frac{1}{4}}\\
&\le K\left\{E\left(\sum_{j=-\infty}^{q}\phi_{v-j}^2\nu_j^2\right)^2\right\}^{\frac{1}{4}}\\
&\le K\Phi_{v-q}^{\frac{1}{2}}\left(\sum_{j=-\infty}^{q}\phi_{v-j}^2E\left(\nu_j^4\right)\right)^{\frac{1}{4}}\\
&\le K\Phi_{v-q},
\end{aligned}$$

where the second inequality employs Burkholder's (1973) inequality and the final one $E\left(\nu_j^4\right)\le 8\left[E\left(\varepsilon_j^8\right)+E\left\{E\left(\varepsilon_j^2|\mathcal{F}_{j-1}\right)\right\}^4\right]\le K$, by 2.12. Considering similarly the three cases $\{p<q<s<v\}$, $\{p\le v<q$ and $s<v\}$ and $\{s<v<p<q\}$, we have

$$|E\left(\chi_v\chi_s\chi_q\chi_p\right)|\le K\left(\Phi_{v-q}+\Phi_{v-s}+\Phi_{q-v}+\Phi_{q-p}\right)$$

whenever $s<v$, $p<q$ and $v\ne q$, where $\Phi_j=0$ for $j<0$. Thus the contribution to 3.25 for $v\ne q$ is bounded in absolute value by

$$K\sum_{t,u=2}^{n}\sum_{v=1}^{t-1}|\psi_{t-v}|\sum_{s=1}^{v-1}c_{t-s}^2\sum_{q=1}^{u-1}|\psi_{u-q}|\sum_{p=1}^{q-1}c_{u-p}^2\left(\Phi_{v-q}+\Phi_{v-s}+\Phi_{q-v}+\Phi_{q-p}\right)$$

$$\begin{aligned}
&\le K\sum_{t,u=2}^{n}\left\{\sum_{v=1}^{t-1}\sum_{q=1}^{u-1}|\psi_{t-v}\psi_{u-q}|\Phi_{v-q}\right\}r_{t-1}r_{u-1}\\
&+K\sum_{j=1}^{\infty}|\psi_j|\sum_{u=2}^{n}r_{u-1}\sum_{t=2}^{n}\sum_{s=1}^{t-1}c_{t-s}^2\left\{\sum_{v=1}^{t-1}|\psi_{t-v}|\Phi_{v-s}\right\}.
\end{aligned}\qquad(3.28)$$

The terms in braces are bounded respectively by

$$\sum_{i,j=0}^{\infty}|\psi_i\psi_{i+j+u-t}|\Phi_j,\quad \sum_{i=1}^{\infty}|\psi_i|\Phi_{t-s-i},$$

which tend to zero as $|u-t|\to\infty$ and $|t-s|\to\infty$ respectively, in view of 1.40 and 1.42 and the Toeplitz lemma. Thus, 1.42, 2.39 and the Toeplitz lemma further imply that 3.28$\to$0 as $n\to\infty$, completing the proof that the first term of 3.24 is $o_p(1)$. The second term of 3.24 can be treated in the same way to conclude that 3.23 is $o_p(1)$. The last term of 3.21 is

$$2\sum_{t=2}^{n}\sum_{j=-\infty}^{t-1}\psi_{t-j}\chi_j\sum_{\substack{v<s\\1}}^{t-1}\varepsilon_v\varepsilon_s c_{t-v}c_{t-s}. \tag{3.29}$$

Now, note that

$$E\left(\chi_j\varepsilon_s\varepsilon_v\chi_k\varepsilon_q\varepsilon_u\right)=0, \quad v<s,\ u<q,\ v\neq u \text{ or } s\neq q.$$

This follows by proceeding recursively using 1.28 and nested conditional expectations, and the fact that $E\left(\varepsilon_t|\mathcal{F}_{t-1}\right)$, $E\left(\varepsilon_t^3|\mathcal{F}_{t-1}\right)$, $E\left(\varepsilon_t^4\varepsilon_u|\mathcal{F}_{u-1}\right)$, $t\geq u$ and $E\left(\varepsilon_t^4\varepsilon_u^2\varepsilon_v|\mathcal{F}_{v-1}\right)$, $t\geq u\geq v$, are all a.s. zero under E3. On the other hand, for all indices,

$$\left|E\left(\chi_j\varepsilon_s\varepsilon_v\chi_k\varepsilon_q\varepsilon_u\right)\right|\leq\max_t E\left(\varepsilon_t^8\right)<\infty$$

by Hölder's inequality. It follows that 3.29 has second moment

$$\begin{aligned}4\sum_{t,u=2}^{n}\sum_{j=-\infty}^{t-1}\psi_{t-j}\sum_{k=-\infty}^{u-1}\psi_{u-k}\sum_{\substack{v<s\\1}}^{\min(t,u)-1}c_{t-v}c_{t-s}c_{u-v}c_{u-s}E\left(\chi_j\chi_k\varepsilon_v^2\varepsilon_s^2\right)&\\ \leq K\sum_{t,u=2}^{n}\sum_{\substack{v<s\\1}}^{\min(t,u)-1}\left|c_{t-v}c_{t-s}c_{u-v}c_{u-s}\right|=O\left(\frac{(\log m)^4}{m^{\frac{1}{3}}}\right)&\end{aligned}$$

as in 3.20, to complete the proof that 3.16 $\to_p 0$ and thus of 2.34.

Application of the remainder of the proof of Robinson (1995a) requires estimation of $U_r-r\sigma^2$ and V_r-U_r, where $U_r=2\pi\sum_{j=1}^{r}I_\varepsilon(\lambda_j)$, and $V_r=\sum_{j=1}^{r}I(\lambda_j)/G\lambda_j^{-2d_x}$, for $1\leq r\leq m$. In Robinson (1995a) it is shown that $U_r-r\sigma^2=O_p(r^{\frac{1}{2}})$. However, to show that (4.7) in Robinson (1995a) is $o_p((\log m)^{-8})$ as required, it is sufficient for $U_r-r\sigma^2$ and V_r-U_r to be $O_p(r^{1-\eta})$, for any $\eta>0$. From Robinson (1995a),

$$U_r-r\sigma^2=\frac{r}{n}\sum_{t=1}^{n}(\varepsilon_t^2-\sigma^2)+\sum_{t=2}^{n}\varepsilon_t\sum_{s=1}^{t-1}\varepsilon_s b_{t-s}, \tag{3.30}$$

where $b_s = \frac{2}{n}\sum_{j=1}^{r}\cos s\lambda_j$. The first term of 3.30 has mean zero and, in view of 1.51, it has variance

$$O\left(\frac{r^2}{n}\sum_{j=1}^{n}|\gamma_j|\right) = O\left(r^2 n^{2d_\varepsilon - 1}\right) = O\left(r^{2(1-\eta)}\frac{r^{2\eta}}{n^{1-2d_\varepsilon}}\right),$$

and this is $O\left(r^{2(1-\eta)}\right)$ under 3.12 on taking $\eta \le \frac{1}{2} - d_\varepsilon$. The second term in 3.30 has mean zero and variance

$$E\left\{\sum_{t=2}^{n}\sigma_t^2\sum_{s=1}^{t-1}\varepsilon_s^2 b_{t-s}^2\right\} + E\left\{\sum_{t=2}^{n}\sigma_t^2\sum_{v\neq s}^{t-1}\varepsilon_s\varepsilon_v d_{t-s}b_{t-v}\right\}.$$

The first term is $O_p\left(n(\max_t E\varepsilon_t^6)\sum_{t=1}^{n} b_t^2\right) = O(r)$ from Robinson (1995a), whereas the second term is zero from 2.13. Thus, $U_r - r\sigma^2 = O_p(r^{1-\eta})$, some $\eta > 0$. As for $V_r - U_r$, we establish below the bound

$$O_p\left(r^{1/3}(\log r)^{2/3} + r^{\beta+1}n^{-\beta} + r^{1/2}n^{-1/4} + rn^{d_\varepsilon - \frac{1}{2}}\right), \tag{3.31}$$

which is indeed $o_p(r^{1-\eta})$, some $\eta > 0$. As shown in Robinson (1995a), to approximate the scores by a suitable martingale it is sufficient that

$$\sum_{j=1}^{m}\nu_j\left(\frac{I(\lambda_j)}{G\lambda_j^{1-2d_x}} - \sigma^2 I_\varepsilon(\lambda_j)\right) = o_p(m^{\frac{1}{2}}), \tag{3.32}$$

and the left side is, by summation by parts and $|\log r - \log(r+1)| \le r^{-1}$, bounded by

$$\sum_{r=1}^{m-1}\frac{1}{r}|V_r - U_r| + 2\log m|V_m - U_m|.$$

We can then invoke 3.12 and 1.82 to establish 3.32, if indeed $V_r - U_r = 3.31$. Now $V_r - U_r$ has second moment equal to

$$\sum_{j,k=1}^{r} E\left(\frac{I_x(\lambda_j)}{G\lambda_j^{-2d_x}} - \frac{2\pi}{\sigma^2}I_\varepsilon(\lambda_j)\right)\left(\frac{I_x(\lambda_k)}{G\lambda_k^{-2d_x}} - \frac{2\pi}{\sigma^2}I_\varepsilon(\lambda_k)\right). \tag{3.33}$$

Robinson (1995a) proves that this is

$$O\left(r^{\frac{2}{3}}(\log r)^{\frac{4}{3}} + r^{2\beta+1}n^{-2\beta} + rn^{-\frac{1}{2}}\right)$$

under assumptions such that $\mathrm{cum}(\varepsilon_w, \varepsilon_s, \varepsilon_t, \varepsilon_u) = \kappa$ when $w = s = t = u$ and zero otherwise. Under the present assumptions, 1.62-1.64 also contribute. The complete fourth cumulant contribution to 3.33 is the following:

$$\frac{1}{(2\pi n)^2}\sum_{j,k=1}^{r}\frac{1}{G^2(\lambda_j\lambda_k)^{-2d_x}}\sum_{w,s,t,u=1}^{n}\text{cum}(x_w,x_s,x_t,x_u)e^{i(w-s)\lambda_j-i(t-u)\lambda_k} \tag{3.34}$$

$$+\frac{1}{(2\pi n)^2}\sum_{j,k=1}^{r}\left(\frac{2\pi}{\sigma^2}\right)^2\sum_{w,s,t,u=1}^{n}\text{cum}(\varepsilon_w,\varepsilon_s,\varepsilon_t,\varepsilon_u)e^{i(w-s)\lambda_j-i(t-u)\lambda_k} \tag{3.35}$$

$$-\frac{1}{(2\pi n)^2}\sum_{j,k=1}^{r}\frac{1}{G\lambda_k^{-2d_x}}\frac{2\pi}{\sigma^2}\sum_{w,s,t,u=1}^{n}\text{cum}(\varepsilon_w,\varepsilon_s,x_t,x_u)e^{i(w-s)\lambda_j-i(t-u)\lambda_k} \tag{3.36}$$

$$-\frac{1}{(2\pi n)^2}\sum_{j,k=1}^{r}\frac{1}{G\lambda_j^{-2d_x}}\frac{2\pi}{\sigma^2}\sum_{w,s,t,u=1}^{n}\text{cum}(x_w,x_s,\varepsilon_t,\varepsilon_u)e^{i(w-s)\lambda_j-i(t-u)\lambda_k} \tag{3.37}$$

Now, applying 1.1,

$$\begin{aligned}
\text{cum}(x_w,x_s,x_t,x_u) &= \sum_{p=-\infty}^{w}\sum_{q=-\infty}^{s}\sum_{l=-\infty}^{t}\sum_{v=-\infty}^{u}\alpha_{w-p}\alpha_{s-q}\alpha_{t-l}\alpha_{u-v}\text{cum}(\varepsilon_p,\varepsilon_q,\varepsilon_l,\varepsilon_v)\\
&= \kappa\sum_{p=-\infty}^{n}\alpha_{w-p}\alpha_{s-p}\alpha_{t-p}\alpha_{u-p}\\
&+ \sum_{\substack{p\neq q\\-\infty}}^{n}\gamma_{p-q}\left(\alpha_{w-p}\alpha_{s-p}\alpha_{t-q}\alpha_{u-q}+\alpha_{w-p}\alpha_{s-q}\alpha_{t-p}\alpha_{u-q}\right.\\
&\qquad\left.+\alpha_{w-p}\alpha_{s-q}\alpha_{t-q}\alpha_{u-p}\right)
\end{aligned}$$

in view of 1.61-1.64 and with the convention that $\alpha_j = 0,\ j < 0$. In the same way,

$$\begin{aligned}
\text{cum}(\varepsilon_w,\varepsilon_s,x_t,x_u) &= \sum_{p=-\infty}^{t}\sum_{q=-\infty}^{u}\alpha_{t-p}\alpha_{u-q}\text{cum}(\varepsilon_w,\varepsilon_s,\varepsilon_p,\varepsilon_q)\\
&= \kappa\delta_{ws}\alpha_{t-w}\alpha_{u-w}+\delta_{ws}\sum_{p=-\infty}^{\min(t,u)}\gamma_{w-p}\alpha_{t-p}\alpha_{u-p}\\
&+ \gamma_{w-s}(\alpha_{t-w}\alpha_{u-s}+\alpha_{t-s}\alpha_{u-w}),
\end{aligned}$$

and a symmetric expression can be written for $\text{cum}(x_w,x_s,x_t,x_u)$. As pointed out above, the contributions to 3.33 from $\text{cum}(x_w,x_s,x_t,x_u)$ are proved in Robinson (1995a) to be

$$O\left(r^{\frac{2}{3}}(\log r)^{\frac{4}{3}}+r^{2\beta+1}n^{-2\beta}+rn^{-\frac{1}{2}}\right).$$

The contribution to 3.33 from 1.62-1.64 is the following:

$$\frac{1}{(2\pi nG)^2}\sum_{j,k=1}^{r}(\lambda_j\lambda_k)^{2d_x}\sum_{w,s,t,u=1}^{n}\sum_{\substack{p\neq q\\-\infty}}^{n}\gamma_{p-q}\left(\alpha_{w-p}\alpha_{s-p}\alpha_{t-q}\alpha_{u-q}\right.$$

$$+\alpha_{w-p}\alpha_{s-q}\alpha_{t-p}\alpha_{u-q} + \alpha_{w-p}\alpha_{s-q}\alpha_{t-q}\alpha_{u-p}\big) e^{i(w-s)\lambda_j - i(t-u)\lambda_k} \quad (3.38)$$

$$+ \frac{n^{-2}}{\sigma^4}\sum_{j,k=1}^{r}\sum_{\substack{w \neq s \\ 1}}^{n}\gamma_{w-s}\left(1+e^{i(w-s)(\lambda_j+\lambda_k)}+e^{i(w-s)(\lambda_j-\lambda_k)}\right) \quad (3.39)$$

$$- \frac{2n^{-2}}{2\pi G\sigma^2}\sum_{j,k=1}^{r}\lambda_k^{2d_x}\sum_{w,s,t,u=1}^{n}\gamma_{t-u}\left(\alpha_{w-t}\alpha_{s-u}+\alpha_{w-u}\alpha_{s-t}\right)e^{i(w-s)\lambda_j-i(t-u)\lambda_k} \quad (3.40)$$

$$- \frac{2n^{-2}}{2\pi G\sigma^2}\sum_{j,k=1}^{r}\lambda_k^{2d_x}\sum_{w,s,t=1}^{n}\sum_{p=-\infty}^{\min(t,u)}\gamma_{w-p}\alpha_{t-p}\alpha_{u-p}e^{-i(t-u)\lambda_k}. \quad (3.41)$$

3.39 is $O(r^2 n^{-1}\sum_{j=1}^{n}|\gamma_j|) = O(r^2 n^{2d_\epsilon - 1})$ as desired. 3.38 contains three terms of the form

$$\frac{1}{(2\pi n G)^2}\sum_{j,k=1}^{r}(\lambda_j\lambda_k)^{2d_x}\sum_{\substack{l \neq u \\ -\infty}}^{n}\gamma_{l-u}\alpha_u(\lambda_j)\alpha_l(-\lambda_j)\alpha_l(\lambda_k)\alpha_u(-\lambda_k), \quad (3.42)$$

where $\alpha_u(\lambda) = \sum_{t=1}^{n}\alpha_{t-u}e^{it\lambda}$ and $\alpha_t = 0,\ t < 0$. When $u < 0$ such that $(-u)^{-1} = O(|\lambda|)$ we have by summation by parts, 1.15 and 2.17, that

$$\begin{aligned}|\alpha_u(\lambda)| &\le \sum_{t=1-u}^{n-u-1}|\alpha_t - \alpha_{t+1}|\left|\sum_{s=1-u}^{t}e^{is\lambda}\right| + |\alpha_{n-u}|\left|\sum_{s=1-u}^{n-u}e^{is\lambda}\right| \\ &= \sum_{t=1-u}^{n-u-1}\left|\frac{\alpha_t-\alpha_{t+1}}{\sin\lambda/2}\right| + |\alpha_{n-u}|\left|\frac{\alpha_{n-u}}{\sin\lambda/2}\right| \\ &\le \sum_{t=1-u}^{n-u-1}|\alpha_t-\alpha_{t+1}|\left|\frac{\sin(t+1)\lambda/2}{\sin\lambda/2}\right| + |\alpha_{n-u}|\left|\frac{\sin(n+1)\lambda/2}{\sin\lambda/2}\right|. \quad (3.43)\end{aligned}$$

As $(-u)^{-1} = O(|\lambda|)$ as $\lambda \to 0$, $\lambda \neq 0(\pi)$, we have, by 2.17, that 3.43 is bounded by

$$K\left\{\sum_{t=1-u}^{n-u-1}\frac{|\alpha_t|}{t|\lambda|} + \frac{|\alpha_{n-u}|}{|\lambda|}\right\} \le K\frac{(1-v)^{d_x-1}}{|\lambda|} = O\big(|\lambda|^{-d_x}\big), \quad (3.44)$$

where the inequality uses 1.15. When $u < 1$ such that $-u = O(1/|\lambda|)$, we have

$$|\alpha_u(\lambda)| \le \sum_{t=1-u}^{1-u+s}|\alpha_u| + \left|\sum_{t=1-u+s}^{n-u}\alpha_t e^{is\lambda}\right| \quad (3.45)$$

for $1 \le s < n$. Applying summation by parts in the same way as above to the second term of 3.45 indicates that it is $O((1-v+s)^{d_x-2}/|\lambda|)$, while the first term is $O((1-v+s)^{d_x})$. Choosing s such that $1-v+s \sim 1/|\lambda|$ indicates that 3.45 is also $O(|\lambda|^{-d_x})$. When $1 \le u \le n$, we have, by summation by parts,

$$|\alpha_u(\lambda)| \le \sum_{t=1}^{s}|\alpha_{t-u}| + \left|\sum_{t=s+1}^{n}\alpha_{t-u}e^{it\lambda}\right|.$$

Applying summation by parts to the second term on the right-hand side, and choosing $s \sim 1/\lambda$, we find again that $\alpha_u(\lambda) = O(|\lambda|^{-d_x})$. Therefore, 3.42 is is $O(r^2 n^{-1} \sum_{j=1}^{n} |\gamma_j|) = O(r^2 n^{2d_\epsilon - 1})$ as desired. The two remaining terms 3.40 and 3.41 are treated in precisely the same way. This completes the proof that the fourth cumulant contribution to $V_r - U_r$ is $O_p(rn^{d_\epsilon - \frac{1}{2}})$. We have of course not assumed 2.56 in the above, but if we do then $\sum_{j=0}^{\infty}|\gamma_j| < \infty$, so it is easily seen that 3.42 is $O(r^2 n^{-1})$, whence 1.82 is not required.

3.5 Finite sample comparison

While the asymptotic properties of $\hat{d}_x$ which we have established are highly desirable, and reassuring in applications to long financial series, it is of interest to examine their relevance to series of more moderate length. For example, conditional heteroscedasticity might worsen the normal approximation in 1.75, and if there is considerable persistence, of the ARCH or GARCH type or especially of the long memory type which our asymptotic theory also permit, the variance of $\hat{d}_x$ might differ considerably from $1/4m$. It is also of interest to consider robustness to departures from the moment conditions of Theorems 4 and 5. Finite sample performance of $\hat{d}_x$ was examined under the presumption of no conditional heteroscedasticity by Robinson (1995a), and compared with that of a version of the log-periodogram estimate, while Taqqu and Teverovsky (1995a) include such estimates in a more comprehensive simulation study, but again restricting to conditionally homoscedastic environments.

We present here results of a Monte Carlo study of the local Whittle estimate applied to simulated series x_t following an ARFIMA(0,d_x,0) parametric version of 1.1 with conditionally Gaussian innovations ε_t (see 2.16) satisfying the same five models (i) to (v) for the conditional variance σ_t^2 as in chapter 2. The model specification 1.28 adopted here for the conditional variance σ_t^2 does not allow for asymmetric response of conditional variances to positive and negative returns. This effect is reported in the empirical finance literature as the leverage effect. The local Whittle estimate of long memory is nonetheless applied to series x_t following an ARFIMA(0,d_x,0) parametric version of 1.1 with conditionally Gaussian innovations following a specific form of Nelson's EGARCH, which models the leverage effect, and which will be denoted model (vi).

(vi) EGARCH: $\varepsilon_t = \sigma_t z_t$, z_t are independent standard normal variables, and $\ln \sigma_t^2 = -.5 + .9 \ln \sigma_{t-1}^2 - .5 z_{t-1} + .5|z_{t-1}|$. The coefficient of z_{t-1} induces a strong leverage effect, i.e. volatility rises in response to unexpectedly low returns. In case of unexpectedly high returns, the volatility behaves as in a simple first order autoregressive stochastic volatility model, with autoregressive coefficient calibrated on typical values in the empirical literature on financial volatilities. The innovations ε_t have finite unconditional moments of arbitrary order.

So far as the ARFIMA(0,d_x,0) model for x_t is concerned, so that in relation to 1.18, $\sum_{j=0}^{\infty} \alpha_j L^j = (1-L)^{-d_x}$, we consider:

(a) "Antipersistence": d_x=-.25,

(b) "Short memory": d_x=0,

(c) "Moderate long memory": d_x=.25,

(d) "Very long memory": d_x=.45.

We study each of (i)-(vi) with (a)-(d), covering a range of short/long/negative memory in ε_t and a range of short/long memory in ε_t^2.

Tables 1-4, 5-8, 9-12 and 13-16 deal respectively with each of the four d_x values (a)-(d). In each case the results are based on n=64, 128 and 256 observations, with bandwidths m= $n/16$, $n/8$, $n/4$, and 10000 replications, as in the Monte Carlo study of Robinson (1995a) with conditionally homoscedastic ε_t. In each group of tables we report, for the conditional variance specifications (i)-(vi), Monte Carlo bias of the local Whittle estimate; Monte Carlo root mean squared error; 95% coverage probabilities based on the N(d_x, $1/4m$) approximation 1.75 for $\hat{d}_x$; and also the

Monte Carlo BIASES for the local Whittle estimate of long memory applied to an ARFIMA(0, −.25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.060	0.014	-0.001	-0.006	-0.011	-0.004	-0.028	-0.017	-0.004
ARCH	0.062	0.010	-0.001	-0.003	-0.016	-0.007	-0.028	-0.016	-0.006
GARCH	0.065	0.020	0.005	-0.004	-0.010	-0.003	-0.026	-0.018	-0.006
LMARCH	0.064	0.012	0.002	-0.001	-0.012	-0.004	-0.022	-0.014	-0.003
VLMARCH	0.064	0.018	0.001	-0.002	-0.010	-0.004	-0.020	-0.013	-0.004
EGARCH	-0.107	-0.054	-0.039	-0.033	-0.012	-0.017	-0.002	-0.002	-0.007

Table 3.1: Local Whittle biases with antipersistence

Monte Carlo ROOT MEAN SQUARED ERRORS for the local Whittle estimate of long memory applied to an ARFIMA(0, −.25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.34	0.24	0.16	0.23	0.16	0.11	0.16	0.11	0.07
ARCH	0.34	0.23	0.17	0.23	0.16	0.12	0.16	0.11	0.08
GARCH	0.34	0.25	0.19	0.24	0.19	0.14	0.18	0.14	0.11
LMARCH	0.34	0.24	0.17	0.24	0.16	0.12	0.16	0.12	0.08
VLMARCH	0.34	0.25	0.18	0.24	0.17	0.13	0.17	0.13	0.10
EGARCH	0.37	0.26	0.18	0.25	0.17	0.13	0.17	0.11	0.08

Table 3.2: Local Whittle RMSEs with antipersistence

efficiency of the log-periodogram estimate relative to the local Whittle estimate, that is the ratio of the Monte Carlo mean squared errors, and we can compare this with the ratio of the asymptotic standard deviations $\sqrt{6}/\pi \simeq .78$.

The series were simulated in the same way as in chapter 2 and for each series $\hat{d}_x$ computed using a simple gradient algorithm. The optimisation was constrained to the compact set $[-.499, .499]$ (chosen values for Δ_1 and Δ_2 respectively) and for selected replications, $R(d)$ was plotted on the interval $[-1.5, 1.5]$ and was always found to be very smooth with a single relative minimum.

We make the comparison with the log periodogram estimate (the version in Robinson

95% COVERAGE PROBABILITIES for the local Whittle estimate of long memory applied to an ARFIMA(0, −.25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.85	0.90	0.84	0.91	0.84	0.89	0.83	0.88	0.91
ARCH	0.85	0.90	0.82	0.92	0.84	0.85	0.84	0.88	0.85
GARCH	0.84	0.88	0.75	0.90	0.76	0.76	0.77	0.77	0.74
LMARCH	0.84	0.90	0.82	0.91	0.83	0.85	0.82	0.86	0.86
VLMARCH	0.85	0.89	0.79	0.91	0.79	0.80	0.79	0.81	0.80
EGARCH	0.81	0.86	0.80	0.88	0.83	0.84	0.84	0.88	0.86

Table 3.3: Local Whittle 95% coverage probabilities with antipersistence

RELATIVE EFFICIENCY of the log periodogram compared to the local Whittle estimate of long memory applied to an ARFIMA(0, −.25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.56	0.68	0.73	0.68	0.76	0.78	0.76	0.80	0.78
ARCH	0.57	0.67	0.74	0.67	0.74	0.79	0.75	0.79	0.81
GARCH	0.57	0.67	0.74	0.66	0.74	0.80	0.73	0.80	0.84
LMARCH	0.57	0.68	0.74	0.67	0.75	0.80	0.76	0.81	0.81
VLMARCH	0.56	0.68	0.75	0.67	0.75	0.81	0.75	0.82	0.83
EGARCH	0.56	0.67	0.73	0.67	0.74	0.80	0.75	0.80	0.81

Table 3.4: Log periodogram relative efficiencies with antipersistence

Monte Carlo BIASES for the local Whittle estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	-0.035	-0.029	-0.025	-0.027	-0.026	-0.013	-0.020	-0.013	-0.008
ARCH	-0.034	-0.030	-0.021	-0.030	-0.024	-0.016	-0.021	-0.015	-0.009
GARCH	-0.033	-0.034	-0.019	-0.037	-0.022	-0.018	-0.026	-0.019	-0.012
LMARCH	-0.031	-0.034	-0.020	-0.032	-0.021	-0.013	-0.019	-0.011	-0.009
VLMARCH	-0.032	-0.032	-0.025	-0.033	-0.024	-0.016	-0.022	-0.015	-0.007
EGARCH	-0.030	-0.036	-0.031	-0.031	-0.025	-0.020	-0.018	-0.015	-0.010

Table 3.5: Local Whittle biases with short memory

Monte Carlo ROOT MEAN SQUARED ERRORS for the local Whittle estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.37	0.27	0.18	0.27	0.18	0.11	0.18	0.11	0.07
ARCH	0.36	0.27	0.19	0.27	0.18	0.13	0.17	0.11	0.08
GARCH	0.36	0.29	0.21	0.28	0.20	0.15	0.20	0.15	0.11
LMARCH	0.37	0.28	0.19	0.27	0.18	0.12	0.18	0.12	0.08
VLMARCH	0.37	0.28	0.20	0.28	0.19	0.13	0.19	0.13	0.10
EGARCH	0.36	0.27	0.19	0.27	0.18	0.13	0.17	0.11	0.09

Table 3.6: Local Whittle RMSEs with short memory

(1995b), but with no trimming) because it has been popularly used, but we do not otherwise report the results for this estimate.

As for the averaged periodogram estimate of long memory investigated in Chapter 2, the most striking feature of the results is the poor performance of $\hat{d}_x$ and of the normal inference rule 1.75 provided by Theorem 5 in the GARCH case, relative to the others. Out of the 36 d_x, m, n combinations, the GARCH bias is largest in 18 cases, while its MSE ties largest in 3 cases and is outright largest in 28. Moreover the deviation of 95% coverage probabilities from their normal values ties largest 3 times and is outright largest 28 times, for GARCH. Relative efficiency to the log periodogram estimate are also most out of line with their asymptotic values for the

95% COVERAGE PROBABILITIES for the local Whittle estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.63	0.76	0.84	0.77	0.84	0.89	0.83	0.88	0.92
ARCH	0.65	0.77	0.81	0.77	0.83	0.84	0.85	0.88	0.86
GARCH	0.65	0.72	0.76	0.74	0.77	0.75	0.79	0.77	0.74
LMARCH	0.64	0.75	0.81	0.76	0.82	0.85	0.82	0.86	0.87
VLMARCH	0.64	0.75	0.79	0.75	0.80	0.81	0.80	0.81	0.81
EGARCH	0.65	0.77	0.80	0.78	0.84	0.84	0.85	0.88	0.86

Table 3.7: Local Whittle 95% coverage probabilities with short memory

RELATIVE EFICIENCY of the log periodogram compared to the local Whittle estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.60	0.78	0.82	0.78	0.84	0.80	0.84	0.82	0.77
ARCH	0.60	0.77	0.80	0.78	0.83	0.82	0.83	0.82	0.82
GARCH	0.60	0.76	0.81	0.77	0.84	0.84	0.84	0.86	0.85
LMARCH	0.60	0.78	0.82	0.78	0.84	0.82	0.84	0.83	0.81
VLMARCH	0.60	0.76	0.82	0.78	0.83	0.83	0.84	0.85	0.84
EGARCH	0.61	0.78	0.82	0.79	0.83	0.82	0.83	0.81	0.82

Table 3.8: Log periodogram relative efficiencies with short memory

Monte Carlo BIASES for the local Whittle estimate of long memory applied to an ARFIMA(0, .25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	-0.108	-0.050	-0.027	-0.040	-0.012	-0.010	-0.004	0.001	-0.007
ARCH	-0.112	-0.053	-0.031	-0.035	-0.014	-0.015	-0.003	-0.004	-0.005
GARCH	-0.113	-0.057	-0.033	-0.043	-0.020	-0.020	-0.014	-0.007	-0.006
LMARCH	-0.110	-0.051	-0.026	-0.038	-0.013	-0.011	-0.005	0.001	-0.006
VLMARCH	-0.104	-0.052	-0.034	-0.044	-0.015	-0.010	-0.005	-0.004	-0.006
EGARCH	-0.107	-0.054	-0.039	-0.033	-0.012	-0.017	-0.002	-0.002	-0.007

Table 3.9: Local Whittle biases with moderate long memory

Monte Carlo ROOT MEAN SQUARED ERRORS for the local Whittle estimate of long memory applied to an ARFIMA(0, .25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.38	0.26	0.17	0.26	0.17	0.11	0.17	0.11	0.07
ARCH	0.37	0.26	0.18	0.25	0.17	0.12	0.16	0.11	0.08
GARCH	0.37	0.28	0.20	0.27	0.20	0.15	0.19	0.14	0.11
LMARCH	0.38	0.27	0.18	0.26	0.17	0.12	0.17	0.12	0.08
VLMARCH	0.37	0.27	0.19	0.27	0.18	0.13	0.18	0.13	0.10
EGARCH	0.37	0.26	0.18	0.25	0.17	0.12	0.17	0.11	0.08

Table 3.10: Local Whittle RMSEs with moderate long memory

95% COVERAGE PROBABILITIES for the local Whittle estimate of long memory applied to an ARFIMA(0, .25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.80	0.86	0.83	0.87	0.84	0.88	0.84	0.89	0.91
ARCH	0.81	0.86	0.80	0.88	0.84	0.85	0.85	0.88	0.86
GARCH	0.80	0.84	0.75	0.86	0.76	0.76	0.79	0.77	0.75
LMARCH	0.80	0.85	0.81	0.87	0.83	0.85	0.82	0.86	0.87
VLMARCH	0.80	0.85	0.79	0.86	0.80	0.81	0.80	0.82	0.81
EGARCH	0.81	0.86	0.80	0.88	0.83	0.84	0.84	0.88	0.86

Table 3.11: Local Whittle 95% coverage probabilities with moderate long memory

RELATIVE EFFICIENCY of the log periodogram compared to the local Whittle estimate of long memory applied to an ARFIMA(0, .25, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.61	0.75	0.79	0.74	0.78	0.79	0.79	0.81	0.79
ARCH	0.62	0.75	0.78	0.74	0.79	0.80	0.78	0.82	0.80
GARCH	0.60	0.74	0.79	0.74	0.79	0.81	0.80	0.82	0.83
LMARCH	0.61	0.76	0.78	0.74	0.80	0.80	0.79	0.81	0.81
VLMARCH	0.61	0.75	0.80	0.74	0.79	0.81	0.79	0.82	0.81
EGARCH	0.61	0.75	0.80	0.74	0.79	0.81	0.78	0.80	0.80

Table 3.12: Log periodogram relative efficiencies with moderate long memory

Monte Carlo BIASES for the local Whittle estimate of long memory applied to an ARFIMA(0, .45, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	-0.201	-0.102	-0.059	-0.087	-0.044	-0.027	-0.035	-0.015	-0.013
ARCH	-0.190	-0.107	-0.070	-0.085	-0.047	-0.033	-0.034	-0.017	-0.018
GARCH	-0.210	-0.132	-0.088	-0.110	-0.073	-0.053	-0.060	-0.043	-0.037
LMARCH	-0.210	-0.117	-0.076	-0.101	-0.060	-0.043	-0.052	-0.030	-0.024
VLMARCH	-0.218	-0.121	-0.081	-0.112	-0.064	-0.047	-0.056	-0.037	-0.032
EGARCH	-0.187	-0.105	-0.070	-0.084	-0.046	-0.034	-0.034	-0.017	-0.017

Table 3.13: Local Whittle biases with very long memory

Monte Carlo ROOT MEAN SQUARED ERRORS for the local Whittle estimate of long memory applied to an ARFIMA(0, .45, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.38	0.23	0.14	0.22	0.13	0.09	0.12	0.08	0.06
ARCH	0.37	0.23	0.16	0.21	0.14	0.10	0.12	0.08	0.07
GARCH	0.38	0.25	0.17	0.23	0.15	0.11	0.14	0.10	0.08
LMARCH	0.38	0.23	0.15	0.21	0.13	0.09	0.13	0.08	0.06
VLMARCH	0.38	0.24	0.16	0.22	0.14	0.10	0.13	0.09	0.07
EGARCH	0.37	0.23	0.16	0.21	0.13	0.10	0.12	0.08	0.07

Table 3.14: Local Whittle RMSEs with very long memory

95% COVERAGE PROBABILITIES for the local Whittle estimate of long memory applied to an ARFIMA(0, .45, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.80	0.86	0.89	0.88	0.91	0.93	0.93	0.94	0.95
ARCH	0.81	0.86	0.87	0.89	0.91	0.91	0.93	0.94	0.92
GARCH	0.81	0.85	0.85	0.87	0.88	0.87	0.90	0.90	0.86
LMARCH	0.81	0.86	0.88	0.89	0.91	0.91	0.92	0.94	0.93
VLMARCH	0.80	0.86	0.87	0.87	0.90	0.89	0.91	0.92	0.89
EGARCH	0.82	0.87	0.87	0.89	0.92	0.91	0.93	0.94	0.92

Table 3.15: Local Whittle 95% coverage probabilities with very long memory

RELATIVE EFFICIENCY of the log periodogram compared to the local Whittle estimate of long memory applied to an ARFIMA(0, .45, 0) series with five specified innovation structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.61	0.65	0.64	0.62	0.61	0.63	0.57	0.59	0.65
ARCH	0.59	0.67	0.67	0.62	0.62	0.63	0.57	0.59	0.66
GARCH	0.62	0.67	0.65	0.63	0.61	0.61	0.57	0.57	0.60
LMARCH	0.61	0.65	0.64	0.61	0.59	0.61	0.56	0.54	0.59
VLMARCH	0.62	0.65	0.65	0.61	0.60	0.60	0.56	0.56	0.62
EGARCH	0.61	0.67	0.68	0.61	0.61	0.64	0.57	0.58	0.66

Table 3.16: Log periodogram relative efficiencies with very long memory

GARCH: it ties with the largest discrepancy 12 times and has the outright largest 10 times. To further investigate this relatively poor performance of the local Whittle estimate in case of GARCH errors, empirical distributions of $2\sqrt{m}(\hat{d}_x - d_x)$ are plotted for all values of d_x (-.25 corresponding to antipersistence, 0 corresponding to short memory, .25 to moderate long memory and .45 to very long memory) on one graphic alongside the standard normal distribution for comparison.

Three graphics are presented in figures 3.1 to 3.3, for three different choices of the pair (n, m), $n = 64$ and $m = 4$, $n = 128$ and $m = 16$, $n = 256$ and $m = 64$. These empirical distributions are truncated because the estimate is restricted to the interval of admissible values $[-0.499, 0.499]$. In the case where $n = 64$ and $m = 4$, the empirical distributions are highly leptokurtic and a high proportion of estimated values for d_x hit one of the boundaries of the interval of admissible values. When n and m increase, the tails become thinner.

Looking at the other error specifications, VLMARCH leads to a slightly worse performance than LMARCH, but with no reliable evidence that this is significantly worse than ARCH, or indeed IID. Failure of the moment conditions 1.43 and 2.12 has no evident effect. In our series of modest length, the relatively poor behaviour under GARCH may be better explained by the impact of a near unit root; for much larger values of n, LMARCH and VLMARCH would presumably do worse than

GARCH, but in such samples this is unlikely to be a matter of great practical concern. In absolute terms, even GARCH does not perform so badly for us to question the usefulness of the asymptotic robustness results in moderate sample sizes. The local Whittle estimate performs has identical root mean squared errors and 95% coverage probabilities in case of EGARCH errors and in case of ARCH errors. In case of EGARCH errors, Monte Carlo biases are typically larger when there is antipersistence and smaller in case of very long memory. Finally, out of the 36 d_x, m, n combinations, relative efficiency of the log periodogram estimate ties largest in 8 cases and is outright largest in 4 cases when the error structure is EGARCH.

As expected, MSE decreases monotonically, as n, and m, decrease. The decay in bias in n is less noticeable, while the typical decay in bias in m is broadly in line with results of Robinson (1995a), in case of fractional Gaussian noise (which has similar spectral shape to that of the ARFIMA(0,d_x,0)). As in the no-ARCH finite sample results of Robinson (1995a), coverage probabilities are markedly sensitive to choice of m.

Finally, the effect of heavy-tailed conditional distributions for ε_t is investigated in tables 17-20 and 21-24 in case of short memory levels ($d_x = 0$).

Monte Carlo biases, root MSEs, coverage probabilities and relative efficiencies of the log periodogram estimate are reported as before for models (i) to (v) only with $\varepsilon_t = \sigma_t \eta_t$, where the η_t are i.i.d. t_2 in tables 17-20 and i.i.d. t_4 in tables 21-24 so that η_t has respectively infinite second moment and infinite fourth moment.

Relative efficiency of the log periodogram estimate seems unaffected by heavy-tailedness. However, when there is no conditional heteroscedasticity, $\hat{d}_x$ on the whole performs better when η_t is t_4 than when it is normal, and better still when it is t_2, in terms of Monte Carlo bias, MSE and coverage probability. Conditional heteroscedasticity produces a reverse picture. The results for t_4 η_t are better than those for normal η_t in only 7 cases in respect of bias, 4 in respect of MSE and 2 in respect of coverage probability. The results for t_2 are better than those for t_4 in only 1 case in respect of bias, 4 in respect of MSE and 4 in respect of coverage

Monte Carlo BIASES for the local Whittle estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	-0.018	-0.027	-0.019	-0.024	-0.018	-0.010	-0.017	-0.009	-0.006
ARCH	-0.043	-0.047	-0.042	-0.042	-0.039	-0.037	-0.036	-0.032	-0.034
GARCH	-0.047	-0.042	-0.035	-0.051	-0.048	-0.040	-0.055	-0.047	-0.038
LMARCH	-0.036	-0.038	-0.032	-0.040	-0.034	-0.028	-0.047	-0.038	-0.028
VLMARCH	-0.042	-0.036	-0.037	-0.052	-0.043	-0.037	-0.054	-0.048	-0.037

Table 3.17: Local Whittle biases with t_2 errors

probability. Moreover, these exceptions are mostly for the $n = 64$, $m = 8$ combination, and frequently the deterioration produced by extreme heavy-tailedness is substantial. And although bias and MSE typically decrease with increasing n and m for t-distributed η_t, suggesting that consistency of $\hat{d}_x$ is maintained, there is some tendency for coverage probabilities to actually worsen (become smaller) especially for t_2, so that not only is the heavy-tailedness reflected in the distribution of $\hat{d}_x$ but there is evidence that the limit distribution of Theorem 5 may not hold under this violation of the moment conditions.

Overall the results suggest that the possibility of conditional heteroscedasticity can be a cause for concern in moderate sample sizes, especially for IGARCH-like behaviour and when the conditional distribution of ε_t has heavy tails. On the other hand, some forms of conditional heteroscedasticity appear to have little effect and in these circumstances, use of $\hat{H}$ and the associated large sample inference rules of Robinson (1995a) seems warranted at least for reasonably large samples, though as is typically the case with smoothed nonparametric estimation, reporting results for a range of bandwidths is a wise precaution.

Monte Carlo ROOT MEAN SQUARED ERRORS for the local Whittle estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.35	0.25	0.16	0.26	0.16	0.10	0.16	0.10	0.07
ARCH	0.33	0.26	0.23	0.24	0.21	0.20	0.17	0.17	0.19
GARCH	0.35	0.31	0.26	0.31	0.28	0.24	0.28	0.26	0.23
LMARCH	0.36	0.29	0.23	0.30	0.25	0.21	0.28	0.24	0.20
VLMARCH	0.35	0.30	0.25	0.31	0.27	0.23	0.29	0.26	0.22

Table 3.18: Local Whittle RMSEs with t_2 errors

95% COVERAGE PROBABILITIES for the local Whittle estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.68	0.81	0.87	0.80	0.88	0.91	0.87	0.91	0.93
ARCH	0.74	0.78	0.71	0.83	0.78	0.65	0.86	0.76	0.56
GARCH	0.68	0.67	0.62	0.66	0.57	0.53	0.59	0.50	0.42
LMARCH	0.65	0.71	0.70	0.67	0.65	0.62	0.59	0.55	0.50
VLMARCH	0.68	0.68	0.65	0.66	0.61	0.56	0.58	0.50	0.45

Table 3.19: Local Whittle 95% coverage probabilities with t_2 errors

RELATIVE EFICIENCY of the log periodogram compared to the local Whittle estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.61	0.76	0.79	0.78	0.80	0.78	0.83	0.80	0.77
ARCH	0.64	0.75	0.79	0.77	0.81	0.81	0.81	0.83	0.83
GARCH	0.62	0.73	0.78	0.73	0.80	0.83	0.78	0.82	0.84
LMARCH	0.60	0.74	0.80	0.74	0.81	0.84	0.80	0.85	0.86
VLMARCH	0.61	0.73	0.80	0.73	0.80	0.84	0.78	0.83	0.85

Table 3.20: Log periodogram relative efficiencies with t_2 errors

Monte Carlo BIASES for the local Whittle estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	-0.028	-0.031	-0.020	-0.026	-0.022	-0.011	-0.021	-0.011	-0.005
ARCH	-0.033	-0.041	-0.035	-0.028	-0.030	-0.022	-0.025	-0.020	-0.019
GARCH	-0.041	-0.043	-0.027	-0.042	-0.037	-0.027	-0.043	-0.029	-0.024
LMARCH	-0.035	-0.030	-0.027	-0.031	-0.023	-0.016	-0.021	-0.022	-0.013
VLMARCH	-0.031	-0.036	-0.028	-0.029	-0.029	-0.019	-0.030	-0.024	-0.019

Table 3.21: Local Whittle biases with t_4 errors

Monte Carlo ROOT MEAN SQUARED ERRORS for the local Whittle estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.37	0.27	0.17	0.28	0.17	0.11	0.17	0.11	0.07
ARCH	0.35	0.26	0.21	0.25	0.18	0.16	0.17	0.13	0.13
GARCH	0.36	0.30	0.24	0.30	0.25	0.21	0.26	0.22	0.18
LMARCH	0.36	0.28	0.20	0.28	0.20	0.15	0.22	0.16	0.11
VLMARCH	0.36	0.29	0.22	0.29	0.22	0.17	0.24	0.19	0.15

Table 3.22: Local Whittle RMSEs with t_4 errors

95% COVERAGE PROBABILITIES for the local Whittle estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.64	0.76	0.85	0.77	0.85	0.89	0.85	0.89	0.91
ARCH	0.69	0.78	0.76	0.80	0.82	0.76	0.86	0.84	0.72
GARCH	0.66	0.68	0.69	0.69	0.65	0.61	0.63	0.58	0.53
LMARCH	0.65	0.74	0.78	0.73	0.78	0.77	0.74	0.75	0.74
VLMARCH	0.65	0.72	0.74	0.72	0.72	0.72	0.69	0.67	0.64

Table 3.23: Local Whittle 95% coverage probabilities with t_4 errors

RELATIVE EFICIENCY of the log periodogram compared to the local Whittle estimate of long memory applied to white noise with five specified error structures.

	n=64			n=128			n=256		
MODEL	m=4	m=8	m=16	m=8	m=16	m=32	m=16	m=32	m=64
IID	0.60	0.77	0.80	0.78	0.81	0.78	0.83	0.81	0.77
ARCH	0.61	0.77	0.80	0.78	0.83	0.82	0.82	0.83	0.83
GARCH	0.60	0.74	0.81	0.74	0.82	0.85	0.81	0.85	0.88
LMARCH	0.60	0.77	0.82	0.77	0.84	0.84	0.84	0.86	0.85
VLMARCH	0.60	0.76	0.83	0.76	0.83	0.85	0.83	0.87	0.86

Table 3.24: Log periodogram relative efficiencies with t_4 errors

Figure 3.1: Empirical distributions of the local Whittle estimate with GARCH errors

$n = 64$, $m = 4$

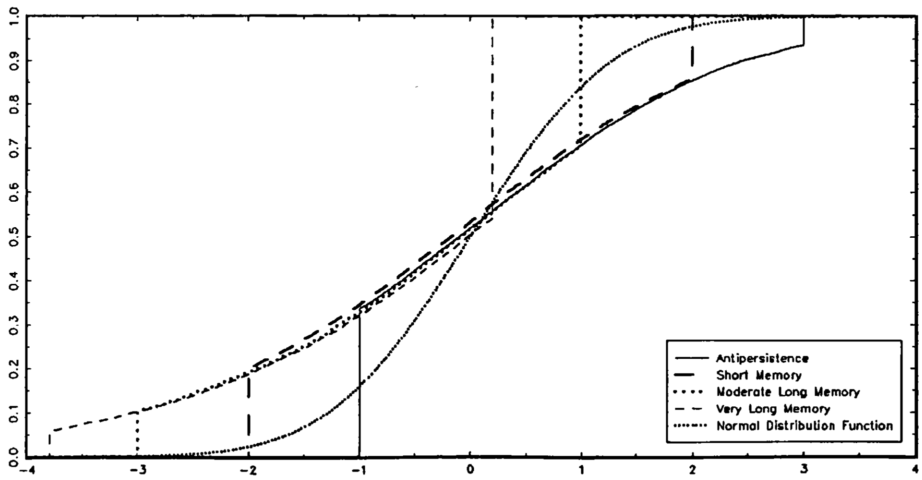


Figure 3.2: Empirical distributions of the local Whittle estimate with GARCH errors

$n = 128$, $m = 16$

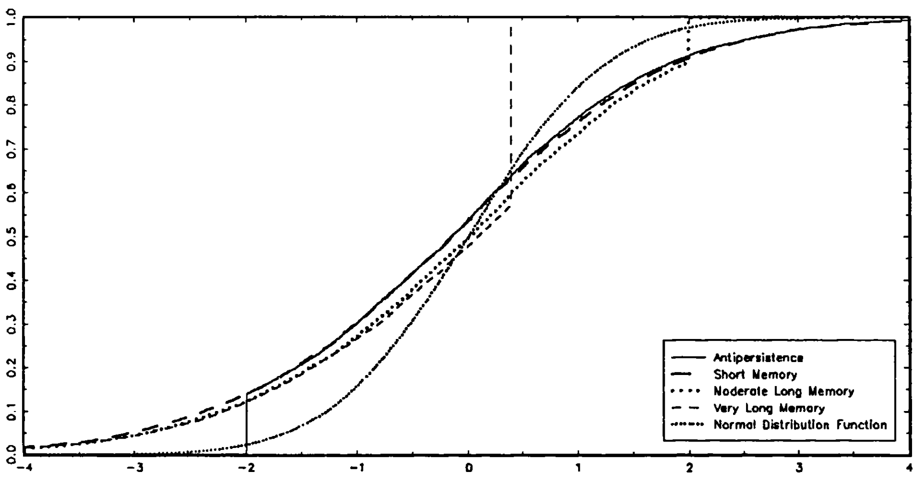
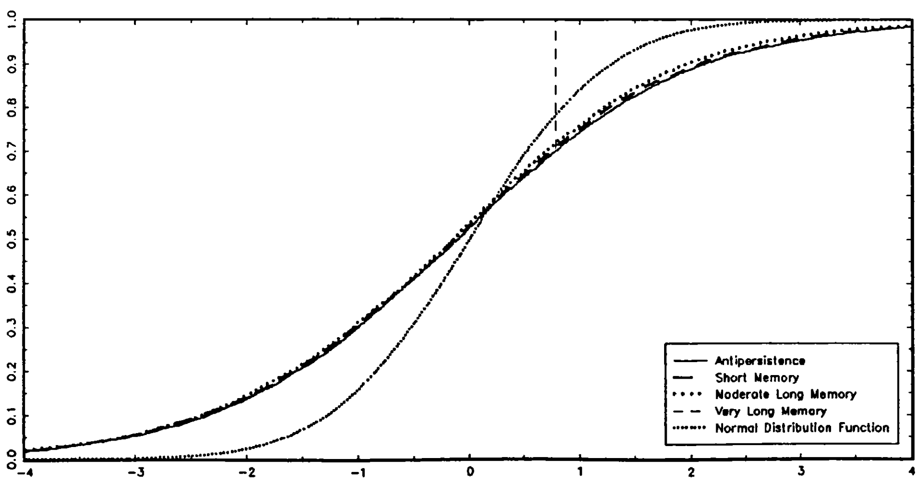


Figure 3.3: Empirical distributions of the local Whittle estimate with GARCH errors

$n = 256$, $m = 64$



3.6 Conclusion

The local Whittle estimate can be used at an initial stage in the analysis of a series x_t, perhaps to test for a specific value of d_x such as 0 (as in Lobato and Robinson (1998)), or to create a fractionally differenced series $\Delta^{\hat{d}_x} x_t$, where Δ is the differencing operator. This represents an asymptotically valid aproximation to an $I(0)$ series without any parametric assumption on the autocorrelation of the underlying $I(0)$ process $\Delta^{d_x} x_t$, so we might then proceed to identify the order of a parametric model such as an ARMA on the basis of the $\Delta^{\hat{d}_x} x_t$, possibly then carrying out estimation of the ARFIMA model for x_t by a parametric Gaussian method. This permits to distinguish between a trend stationary and a difference stationary series as in the derivations of Smith and Chen (1996) and Deo and Hurvich (1998). A question that then arises is whether the innovations in the model (equivalent to our ε_t) have conditional heteroscedasticity, and if so, what is the nature and extent of it. This is of interest whether or not x_t has long memory, and even if x_t is a martingale difference, $x_t = \varepsilon_t$. If 1.33 is parameterized, say by 1.41 or 1.38, then we can estimate the unknown parameters by applying the conditional Gaussian loglikelihood underlying the LM tests developed by Robinson (1991b), though asymptotic properties of the parameter estimates remain to be established in the long memory case, and indeed in many short memory ones. However such a procedure carries the disadvantage that even the memory parameter d_ε will be inconsistently estimated if the short memory dynamics of the squares is misspecified, while we may in any case prefer an exploratory approach at the initial stage.

One may thus consider applying a semiparametric procedure for estimating d_ε to the ε_t^2, or their proxies. For example, the local Whittle method appears to be a candidate, because, although the ε_t^2 cannot be Gaussian, Gaussianity of x_t was not assumed by Robinson (1995a), or in the current thesis. However, while some of the analysis of these papers will be relevant, and 1.33 represents ε_t^2 as a linear filter of martingale differences ν_t, not only do the ν_t have conditional heteroscedasticity but their odd conditional moments are perforce stochastic, so that no conditions analogous to 2.55 or 2.13 can be imposed. The form of the limiting distribution of the local Whittle estimate of d_ε, as well as its derivation, are thus open questions.

Chapter 4

Optimal bandwidth choice

Joint work with Peter Robinson

4.1 Introduction

This chapter is concerned with the optimal choice of bandwidth, or number of periodogram ordinates used in the estimation of long memory in time series. The semiparametric estimates that are considered are those for which asymptotic theory was provided. Namely, this chapter is concerned with bandwidth choice in local Whittle estimation described in chapter 3, in log periodogram estimation and in averaged periodogram estimation described in chapter 2.

The asymptotic normality result

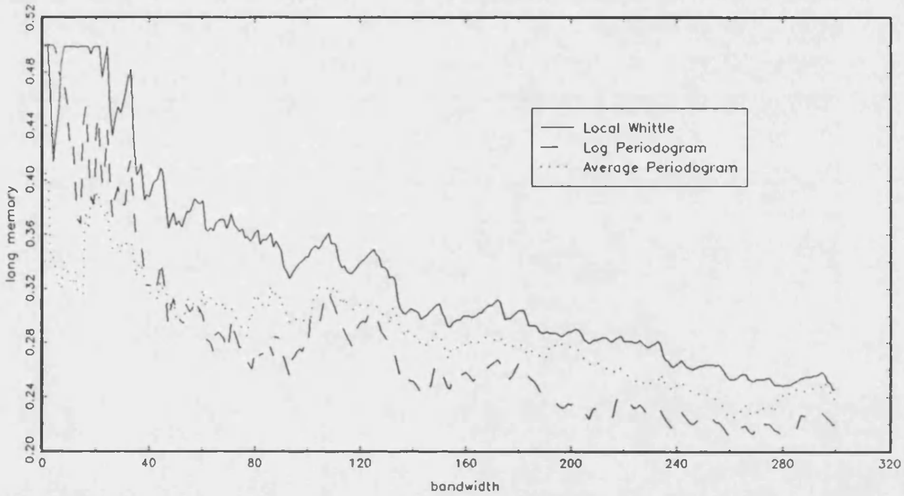
$$\sqrt{m}(\hat{d}_x - d_x) \to_d N(0, V(\hat{d}_x)) \tag{4.1}$$

which holds for the local Whittle estimate (hereafter LW) for $\Delta_1 < d_x < \Delta_2$ for any Δ_1, Δ_2 such that $-\frac{1}{2} < \Delta_1 < \Delta_2 < \frac{1}{2}$ and with variance $V(\hat{d}_x) = \frac{1}{4}$, for the log periodogram estimate in the modified form of Robinson (1995b) (hereafter LP) for $0 < d_x < \frac{1}{2}$ with variance $V(\tilde{d}_x) = \pi^2/24$, and for the averaged periodogram estimate (hereafter AP) for $0 < d_x < \frac{1}{4}$ with variance $V(\hat{d}_{xq}) = (1+q^{-1}-2q^{-2d_x})(\frac{1}{2}-d_x)^2/[(\log^2 q)(1-4d_x)]$ (see Lobato and Robinson (1996)) for any choice of $q \in (0,1)$

shows the crucial role the bandwidth m plays in the precision of the estimate. For the LW, 4.1 is derived under 1.76 which implies a rate of convergence for $\hat{d}_x$ of $n^{-r}M_n$ where $r = \beta/(1+2\beta)$ and $(\log m)^{-1/(1+2\beta)}M_n = o(1)$ as $n \to \infty$. For the LP and AP, 4.1 is derived under 1.77 which implies a rate of convergence of $n^{-r}M_n$ where M_n diverges arbitrarily slowly.

All the results mentioned above are derived under conditions of "oversmoothing"; in other words, under conditions on the bandwidth which ensure that the asymptotic bias of the estimate is of small order of magnitude with regards to its asymptotic variance. However, it appears clearly that a better precision would be achieved by these estimates with choices of m outside the scope of the results expounded above. In fact, Giraitis, Robinson, and Samarov (1997) show that for long memory processes with spectral density satisfying 1.84, the best attainable rate of convergence for an estimate of d_x is n^{-r}. They further show that such a rate is attained in the case of the LP estimate in the modified form of Robinson (1995b). Once this optimal rate is achieved, the problem of choosing bandwidth optimally remains one of balancing bias and imprecision. Taqqu and Teverovsky (1996) propose a graphical method for the determination of bandwidth which consists in plotting a series of estimated values for $\hat{d}_x$ against n/m and choosing the appropriate estimate for d_x in the following way: *Starting at large values of m (small values of n/m), we would expect to find a range where the estimates of d_x are incorrect because of the short range effects. Then, as m decreases (n/m increases), the short range effects should disappear and the value of d_x obtained should represent the true long memory dependence. Thus there should be a period of relative stability, where the estimates of d_x are approximately constant. Then, if we move to smaller m's, we will get into a region where the estimates of d_x are very scattered and unreliable because there are not enough frequencies left to have an accurate regression. Thus we should expect to see a flat region somewhere in the middle of the plot of the estimates of d_x and we can estimate an overall d_x from that region.* Examples they give for electronic data seem graphically convincing. However, for the three semiparametric estimates LW, LP and AP, applied to the Nile River data (described in chapter 1), plotting estimates of d_x against bandwidth

Figure 4.1: Long memory function of bandwidth in the Nile river data



(as in figure 1) does not produce anything resembling a flat region in the plot wherein the estimate of d_x might be selected.

Figures 4.2 to 4.5 are plots of the long memory estimates against bandwidth for a simulated ARFIMA($1,d_x,0$) series of length $n = 1000$ with autoregressive coefficient $a = .5$ and for values of the long memory parameter $d_x = -.25$, $d_x = 0$, $d_x = .25$ and $d_x = .45$. Each graphic corresponds to one true value of d_x, and the estimated values for d_x using the local Whittle, the log periodogram and the averaged periodogram are plotted against bandwidth. In all four figures, there is a region -approximately the first 40 values of bandwidth- where the estimates are extremely erratic, and a region -approximately for $m > 100$- where estimates of long memory increase continuously with bandwidth, because of the increasing influence of the short range structure which spuriously inflates estimates of long range dependence. One would therefore wish to choose a bandwidth somewhere between these regions.

Figure 4.2: Long memory estimation in an ARFIMA(1,-.25,0) series

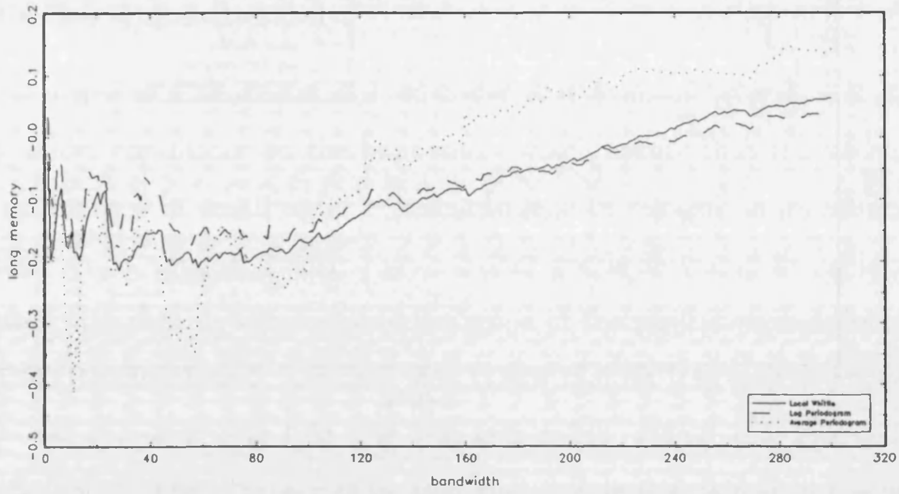


Figure 4.3: Long memory estimation in an ARMA(1,0) series

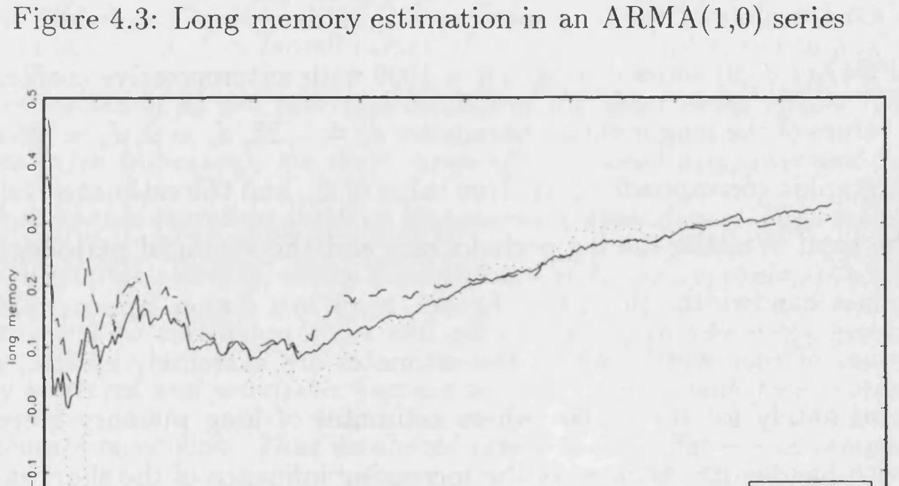
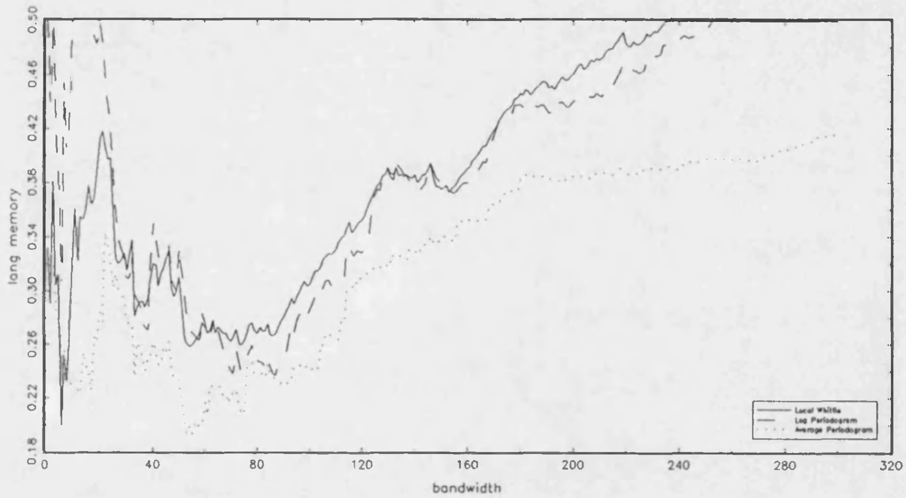


Figure 4.4: Long memory estimation in an ARFIMA(1,.25,0) series

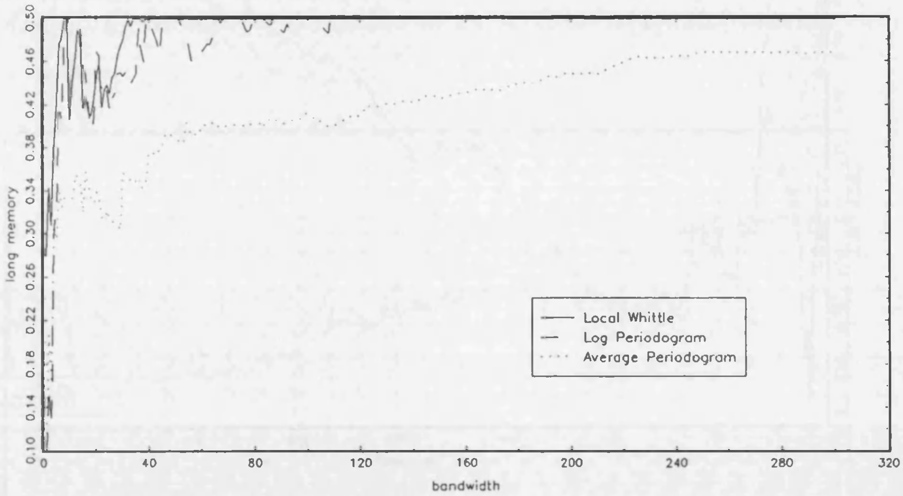


Here, we are interested in theoretical optimal bandwidths for LW, LP and AP, and feasible approximations thereof. Grenander and Rosenblatt (1966) propose a theoretical criterion for optimal choice of bandwidth with the minimisation of the estimate's mean squared error which suitably balances asymptotic bias and asymptotic variance. The mean squared error of a generic estimate $\hat{d}_x$ of d_x is:

$$E|\hat{d}_x - d_x|^2 = V(\hat{d}_x) + (E\hat{d}_x - d_x)^2. \tag{4.2}$$

Robinson (1994b) proposed an optimal bandwidth theory for the AP based on an analogue of the mean squared error of smooth spectral density estimates. Delgado and Robinson (1996) assessed feasible approximations to this optimal bandwidth, and the following section of this chapter shows that such optimal bandwidth formulae and their approximations remain valid when the process x_t follows 1.1 with innovations displaying (possibly long memory) conditional heteroscedasticity.

Figure 4.5: Long memory estimation in an ARFIMA(1,.45,0) series



Section 4.3 of this chapter derives an expression for the mean squared error and the corresponding optimal bandwidth for the LW estimate of long memory and gives feasible approximations to it. Section 4.4 gives a small sample assessment of automatic bandwidth selection procedures in semiparametric estimation of long memory in an extensive Monte Carlo study. The case of the LP is treated in Hurvich, Deo, and Brodsky (1998) for processes satisfying the fractional representation 1.83. The bias behaves asymptotically as

$$-\frac{2\pi^2}{9}\frac{f^{*''}(0)}{f^*(0)}\frac{m^2}{n^2} \tag{4.3}$$

and the variance as $\pi^2/24m$, whence a mean squared error minimising bandwidth is derived,

$$m_{LP} = \Big(\frac{81(2f^*(0))^2}{96\pi^2(f^{*''}(0))^2}\Big)^{\frac{1}{5}} n^{\frac{4}{5}} \tag{4.4}$$

which will be applied for comparison in the small sample assessments of the performance of the three semiparametric estimates of long memory under optimal bandwidth choice.

4.2 Bandwidth selection for the averaged periodogram

Grenander and Rosenblatt (1966) proposed the mean squared error criterion for optimal bandwidth selection in weighted periodogram estimation of the spectral density of a weakly dependent process. for the simple discretely averaged periodogram, this criterion is equivalent to the minimization of

$$E\left[\frac{\hat{F}(\lambda_m)}{\lambda_m} - f(0)\right]^2 = \text{var}\left(\frac{\hat{F}(\lambda_m)}{\lambda_m}\right) + \left(E\left[\frac{\hat{F}(\lambda_m)}{\lambda_m}\right] - f(0)\right)^2.$$

Calling

$$g(\lambda) = L(\lambda)\lambda^{-2d_x}$$

and

$$G(\lambda) = \int_0^\lambda g(\theta)d\theta,$$

Robinson (1994b) proposes an analogue to this mean squared error in case the spectrum of the process x_t is singular at zero frequency and follows 1.18:

$$MSE = E\left[\frac{\hat{F}(\lambda_m)}{G(\lambda_m)} - 1\right]^2.$$

To describe the bias component in this MSE, $f(\lambda)$ is specified in the following way:

<u>Assumption F1</u> $f(\lambda)$ follows 1.84. In addition, f is differentiable, and

$$f'(\lambda) = O\left(\frac{g(\lambda)}{\lambda}\right) \text{ as } \lambda \to 0^+.$$

Other assumptions for the determination of optimal bandwidth are added below:

<u>Assumption F2</u> Bandwidth m satisfies 1.69.

<u>Assumption F3</u> Assumption B3 holds with the additional requirement that the α_j are quasi monotonically convergent, and

$$\alpha_j \sim \zeta_j \quad \text{or} \quad \alpha_j \sim -\zeta_j \quad \text{as } j \to \infty, \tag{4.5}$$

where

$$\zeta_j = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \Gamma(1-d_x)\cos((\frac{1}{2}-d_x)\pi)L^{\frac{1}{2}}(\frac{1}{j})j^{d_x-1}, \tag{4.6}$$

and

$$d_\epsilon \leq \frac{\beta}{(2\beta+1)}. \tag{4.7}$$

Quasi monotonic convergence of the α_j entails 2.17 for all sufficiently large j. It is satisfied if the α_j are eventually decreasing, and is satisfied in case of autoregressive fractionally integrated moving average. 4.5 and 4.6 are also satisfied in that case, and so is 1.84 with $\beta = 2$. 4.5 implies 1.15 which is required for the treatment of fourth cumulants. Optimal bandwidths formulae proposed by Robinson (1994b) for processes under the conditional homogeneity condition 1.26 continue to hold under assumptions F1-F3:

<u>**Theorem 6**</u> Under Assumptions F1-F3 and $0 < d_x < 1/4$,

$$M\hat{S}E \sim 4(\frac{1}{2}-d_x)^2\left(\frac{1}{(1-4d_x)m}+\left(\frac{E_{\beta d_x}}{1-2d_x-\beta}\right)\lambda_m^{2\beta}\right), \tag{4.8}$$

and a mean squared error minimising bandwidth is given by:

$$\hat{m} = \left(\frac{(1-2d_x+\beta)^2}{2\beta E_{\beta d_x}^2(2\pi)^{2\beta}(1-4d_x)}\right)^{\frac{1}{2\beta+1}} n^{\frac{2\beta}{2\beta+1}}. \tag{4.9}$$

When $1/4 < d_x < 1/2$, $f(\lambda)$ is no longer square integrable in a neighbourhood of the origin. Therefore, as Robinson (1994b) notes, a global Lipschitz condition of degree $1-2d_x$ is imposed on the spectral density via quasi monotone convergence of the autocovariances.

<u>**Theorem 7**</u> If Assumptions F1-F3 and $1/4 < d_x < 1/2$ hold, and the autocovariances of the process x_t are quasi monotonically convergent, then a mean squared error minimising bandwidth $\hat{m}$ is given by

$$(\frac{L(\lambda_{\hat{m}})}{L(\frac{1}{n})})^{\frac{1}{1-2d_x+\beta}}\hat{m} \sim \frac{\frac{\beta}{2\pi}}{n^{1-2d_x+\beta}}(\frac{D_{d_x}(1-2d_x+\beta)}{4\beta}[\frac{2d_x-1+\beta}{E_{\beta d_x}(2d_x)}$$

$$+ \quad \frac{1}{|E_{\beta d_x}|}\Big\{\frac{(1-2d_x+\beta)^2}{(d+\frac{1}{2})^2(2d_x)^2} + 16\beta(\frac{1}{2}-d_x)\Big(\frac{1}{(4d_x-1)(2d_x)}$$

$$- \quad \frac{1}{(d_x+\frac{1}{2})^2(2d_x)^2} - \frac{4\Gamma^2(2d_x)}{\Gamma(4d_x+2)}\Big)\Big\}^{\frac{1}{2}}\Big]\Big)^{\frac{1}{1-2d_x+\beta}}, \qquad (4.10)$$

where $D_{d_x} = 2\Gamma(2(d_x-\frac{1}{2}))\cos((\frac{1}{2}-d_x)\pi)$.

Proof Assumption F3 differs from Assumption 7 in Robinson (1994b) only to the extent that Robinson (1994b) assumed that $\text{cum}(\varepsilon_r, \varepsilon_t, \varepsilon_s, \varepsilon_u) = 0$ unless $r = t = s = u$. Therefore, the contributions to the mean squared error described in Theorem 4 of Robinson (1994b) are unchanged except for the contribution of fourth cumulants of ε_t to the variance of $\hat{F}(\lambda_m)$. Therefore, the proof of Theorem 4 of Robinson (1994b) still applies except for the proof that the additional term in the MSE due to non-Gaussianity, namely $\hat{K}_m/G(\lambda_m^2)$ with

$$\hat{K}_m = \frac{1}{n^4}\sum_{j,k=1}^{m}\sum_{q,r,s,t=1}^{n} \text{cum}(x_q, x_r, x_s, x_t)e^{i(q-r)\lambda_j - i(s-t)\lambda_k}$$

is of small order of magnitude with respect to other terms in the MSE, and therefore does not influence asymptotic optimal bandwidth choice. More precisely, we need to prove that

$$\hat{K}_m = o\left(\frac{G(\lambda_m)^2}{m}\right) \quad \text{as} \quad n \to \infty. \qquad (4.11)$$

Now, applying 1.1,

$$\begin{aligned}\text{cum}(x_q, x_r, x_s, x_t) &= \sum_{j=-\infty}^{q}\sum_{k-\infty}^{r}\sum_{l=-\infty}^{s}\sum_{u=-\infty}^{t} \alpha_{q-j}\alpha_{r-k}\alpha_{s-l}\alpha_{t-u}\text{cum}(\varepsilon_j, \varepsilon_k, \varepsilon_l, \varepsilon_u) \\ &= \kappa\sum_{l=-\infty}^{n} \alpha_{q-l}\alpha_{r-l}\alpha_{s-l}\alpha_{t-l} \\ &+ \sum_{\substack{l\neq u \\ -\infty}}^{n} \gamma_{u-l}\left(\alpha_{q-u}\alpha_{r-u}\alpha_{s-l}\alpha_{t-l} + \alpha_{q-u}\alpha_{r-l}\alpha_{s-u}\alpha_{t-l}\right. \\ &\qquad\qquad \left. + \alpha_{q-u}\alpha_{r-l}\alpha_{s-l}\alpha_{t-u}\right)\end{aligned}$$

in view of 1.61-1.64 and with the convention that $\alpha_j = 0$, $j < 0$. Therefore,

$$\begin{aligned}\hat{K}_m &= \frac{1}{n^4}\sum_{j,k=1}^{m}\sum_{q,r,s,t=1}^{n}\kappa\sum_{l=-\infty}^{n}\alpha_{q-l}\alpha_{r-l}\alpha_{s-l}\alpha_{t-l}e^{i(q-r)\lambda_j - i(s-t)\lambda_k} \qquad (4.12) \\ &+ \frac{1}{n^4}\sum_{j,k=1}^{m}\sum_{q,r,s,t=1}^{n}\sum_{\substack{l\neq u \\ -\infty}}^{n}\gamma_{u-l}\left(\alpha_{q-u}\alpha_{r-u}\alpha_{s-l}\alpha_{t-l} + \alpha_{q-u}\alpha_{r-l}\alpha_{s-u}\alpha_{t-l}\right.\end{aligned}$$

$$+\alpha_{q-u}\alpha_{r-l}\alpha_{s-l}\alpha_{t-u}\big)\, e^{i(q-r)\lambda_j - i(s-t)\lambda_k} \tag{4.13}$$

The proof of Lemma 12 of Robinson (1994b) still holds to show that 4.12 is $O(G(\lambda_m)^2/n)$ which is $o(G(\lambda_m)^2/m)$ under 1.69 as required. 4.13 contains three terms of the form

$$\frac{1}{n^4}\sum_{j,k=1}^{m}\sum_{\substack{l \neq u \\ -\infty}}^{n} \gamma_{l-u}\alpha_u(\lambda_j)\alpha_l(-\lambda_j)\alpha_l(\lambda_k)\alpha_u(-\lambda_k), \tag{4.14}$$

where $\alpha_u(\lambda) = \sum_{t=1}^{n} \alpha_{t-u}e^{it\lambda}$ and $\alpha_t = 0,\ t < 0$. When $u < 0$ such that $(-u)^{-1} = O(|\lambda|)$ we have by summation by parts, 1.15 and 2.17, that

$$\begin{aligned}
|\alpha_u(\lambda)| &\leq \sum_{t=1-u}^{n-u-1} |\alpha_t - \alpha_{t+1}| \left|\sum_{s=1-u}^{t} e^{is\lambda}\right| + |\alpha_{n-u}| \left|\sum_{s=1-u}^{n-u} e^{is\lambda}\right| \\
&= \sum_{t=1-u}^{n-u-1} \left|\frac{\alpha_t - \alpha_{t+1}}{\sin\lambda/2}\right| + |\alpha_{n-u}| \left|\frac{\alpha_{n-u}}{\sin\lambda/2}\right| \\
&\leq \sum_{t=1-u}^{n-u-1} |\alpha_t - \alpha_{t+1}| \left|\frac{\sin(t+1)\lambda/2}{\sin\lambda/2}\right| + |\alpha_{n-u}| \left|\frac{\sin(n+1)\lambda/2}{\sin\lambda/2}\right|. \qquad (4.15)
\end{aligned}$$

As $(-u)^{-1} = O(|\lambda|)$ as $\lambda \to 0$, $\lambda \neq 0(\pi)$, we have, by 2.17, that 4.15 is bounded by

$$K\left\{\sum_{t=1-u}^{n-u-1} \frac{|\alpha_t|}{t|\lambda|} + \frac{|\alpha_{n-u}|}{|\lambda|}\right\} \leq K\frac{(1-v)^{d_x-1}}{|\lambda|} = O\big(|\lambda|^{-d_x}\big), \tag{4.16}$$

where the inequality uses 1.15. When $u < 1$ such that $-u = O(1/|\lambda|)$, we have

$$|\alpha_u(\lambda)| \leq \sum_{t=1-u}^{1-u+s} |\alpha_u| + \left|\sum_{t=1-u+s}^{n-u} \alpha_t e^{is\lambda}\right| \tag{4.17}$$

for $1 \leq s < n$. Applying summation by parts in the same way as above to the second term of 4.17 indicates that it is $O((1-v+s)^{d_x-2}/|\lambda|)$, while the first term is $O((1-v+s)^{d_x})$. Choosing s such that $1-v+s \sim 1/|\lambda|$ indicates that 4.17 is also $O(|\lambda|^{-d_x})$. When $1 \leq u \leq n$, we have, by summation by parts,

$$|\alpha_u(\lambda)| \leq \sum_{t=1}^{s} |\alpha_{t-u}| + \left|\sum_{t=s+1}^{n} \alpha_{t-u}e^{it\lambda}\right|. \tag{4.18}$$

Applying summation by parts to the second term on the right-hand side, and choosing $s \sim 1/\lambda$, we find again that $\alpha_u(\lambda) = O(|\lambda|^{-d_x})$. Therefore, 4.14 is

$$O\left(\frac{1}{n^4}\left(\sum_{j,k=1}^{m}(\lambda_j\lambda_k)^{-2d_x}\right) n\sum_{j=1}^{n}|\gamma_j|\right).$$

Now from the proof of Theorem 1 of Robinson (1994b), $G(\lambda) \sim K\lambda^{1-2d_x}$ as $\lambda \to 0^+$, so that 4.14 is $O\left(G(\lambda_m)^2 n^{-1} \sum_{j=1}^{n}\right)$ which is $O\left(G(\lambda_m)^2 n^{2d_\epsilon - 1}\right)$ by 1.46. When d_ϵ satisfies 4.7, 4.14 is $o(G(\lambda_m)/m)$ which is negligible with respect to other components of the MSE. We have of course not assumed 2.56 in the above, but if we do then $\sum_{j=0}^{\infty}|\gamma_j| < \infty$, so it is easily seen that 4.14 is $O(G(\lambda_m)^2/m)$, whence 4.7 is not required. Therefore, fourth cumulant contributions to the MSE are $o(1/m)$, which is negligible with respect to other components of strict order of magnitude $1/m$.

Delgado and Robinson (1996) proposed feasible approximations to this optimal bandwidth, noting that its rate of convergence is free of d_x only when $0 < d_x < 1/4$. The validity of the approximations to the optimal bandwidth follows from the validity of the theoretical optimal bandwidths.

4.3 Bandwidth choice for the local Whittle estimate

The estimate considered in this section is the LW described in chapter 3 which minimises $R(d)$ defined in 3.2. Hereafter, the dependence in m of R will be referred to explicitly by writing $R(m, d)$. For the sake of asymptotic bias determination, the spectral density of the process x_t needs to be specified with 1.84 with $0 < L(\lambda) = G < \infty$ and writing $E_{\beta d_x}$ explicitly as a function of d_x as $E_\beta(d_x)$. As was noted in chapter 1 and the introduction to the present chapter, Giraitis, Robinson, and Samarov (1997) showed that there exists a lower bound for the rate of convergence of semiparametric estimates on a class of spectral densities including the above specified. They showed, moreover, that this lower bound is attained by the LP. We suggest that the same property holds for the local Whittle estimate $\hat{d}_x$.

As is familiar from other uses of smoothed nonparametric estimates, this optimal rate of convergence lies outside the asymptotic normality range for the local Whittle estimate $\hat{d}_x$ considered here, as M_n is not free to diverge arbitrarily slowly. Nonetheless, heuristically, the asymptotic variance remains equal to $1/4m$ and the optimal bandwidth is the one that yields that same rate of convergence for the squared bias.

More precisely, the Mean Value Theorem yields:

$$\hat{d}_x - d_x = \frac{\frac{\partial R(m,d_x)}{\partial d}}{\frac{\partial^2 R(m,\tilde{d})}{\partial^2 d}} \quad \text{where} \quad |\tilde{d} - d_x| \leq |\hat{d}_x - d_x|. \tag{4.19}$$

Now, as shown in Robinson (Robinson 1995a), $\frac{\partial^2 R(m,\tilde{d})}{\partial d^2} \xrightarrow{p} 4$. The first two moments of $\frac{\partial R(m,d)}{\partial d}$ can be treated heuristically as follows. As $n \to \infty$,

$$\frac{\partial R}{\partial d} \sim_p \frac{2}{m}\sum_{j=1}^{m} \nu_j [\frac{I_x(\lambda_j)}{G\lambda_j^{-2d_x}} - 1] \quad \text{where} \quad \nu_j = \log j - \frac{1}{m}\sum_{k=1}^{m} \log j, \tag{4.20}$$

the notation $\sim_p$ meaning that the ratio of left and right sides tends to 1 in probability. We thus suggest that the expectation of $\frac{\partial R}{\partial d}$ (assuming it exists) can be approximated by

$$\frac{2}{m}\sum_{j=1}^{m} \nu_j [\frac{f(\lambda_j)}{G\lambda_j^{-2d_x}} - 1], \tag{4.21}$$

so under 1.84 the bias may be approximated by

$$\frac{1}{2m} E_\beta(d_x) \sum_{j=1}^{m} \nu_j \lambda_j^\beta \sim \frac{1}{2}(\frac{2\pi}{n})^\beta E_\beta(d_x) \frac{\beta}{(\beta+1)^2} m^\beta \tag{4.22}$$

$$= \frac{\beta}{2(\beta+1)^2} E_\beta(d_x) \lambda_m^\beta. \tag{4.23}$$

Likewise we suggest that the variance of $\frac{\partial R}{\partial d}$ can be approximated by

$$\frac{4}{m^2}\sum_{j=1}^{m} \nu_j^2 E[\frac{I_x(\lambda_j)}{G\lambda_j^{-2d_x}} - \frac{E[I_x(\lambda_j)]}{G\lambda_j^{-2d_x}}]^2 \sim \frac{4}{m^2}\sum_{j=1}^{m} \nu_j^2 \sim \frac{4}{m} \tag{4.24}$$

Thus we suggest, from 4.23 and 4.24, that the mean squared error $E(\hat{d}_x - d_x)^2$ is dominated by

$$\frac{1}{4}[\frac{1}{m} + E_\beta(d_x)^2 \frac{\beta^2}{(\beta+1)^4} \lambda_m^{2\beta}], \tag{4.25}$$

from which an "optimal bandwidth" as n tends to infinity can be derived by straightforward calculus:

$$m_{\text{opt}} = \left[\frac{(\beta+1)^4}{2\beta^3 E_\beta(d_x)^2 (2\pi)^{2\beta}}\right]^{\frac{1}{1+2\beta}} n^{\frac{2\beta}{1+2\beta}}. \tag{4.26}$$

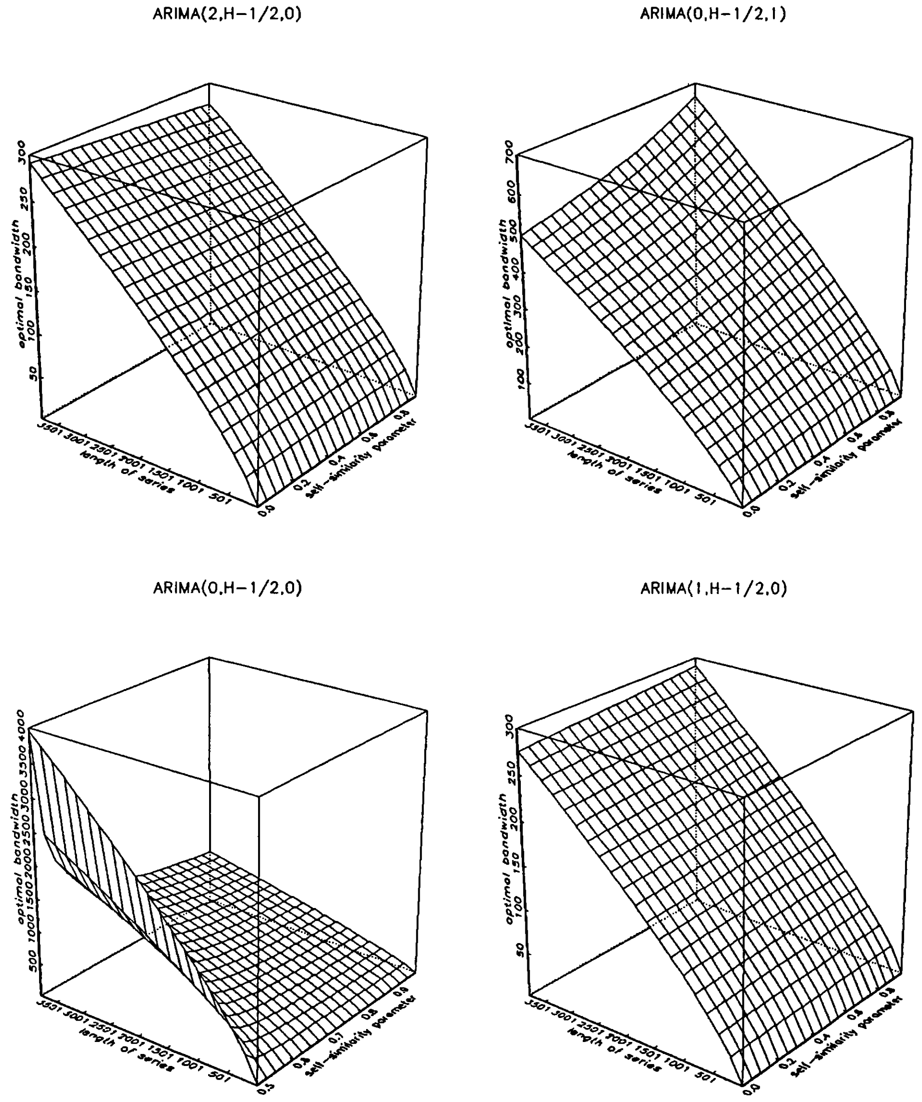
This optimal bandwidth may be compared to ones relevant to the averaged periodogram estimate of d_x in 4.9 and 4.10.

When we consider the smoothest specification, i.e. $\beta = 2$, the heuristic optimal spectral bandwidth becomes: $m_{\text{opt}} = (\frac{3n}{4\pi})^{\frac{4}{5}} |E_2(d_x)|^{-\frac{2}{5}}$. In a fractional differencing representation of the form 1.83, $E_2(d_x)$ can be approximated by $\frac{f^{*\prime\prime}(0)}{2f^*(0)} + \frac{1}{12}(d_x)$, as shown by Delgado and Robinson (1996). Figure 4.6 plots the optimal bandwidth as a function of n and d_x and four models for the levels: ARFIMA(0,d_x,0), ARFIMA(1,d_x,0), ARFIMA(0,d_x,1) and ARFIMA(2,d_x,0) with autoregressive coefficients $a_1 = .5$ and $a_2 = .2$ and moving average coefficient $b = -.45$. When there is no short range dependence structure, the optimal bandwidth displays very little dependence in d_x except in a close neighbourhood of $d_x = 0$, where the optimal bandwidth is singular. It is therefore truncated at $m = [(n-1)/2]$ which corresponds to parametric Whittle estimation. The optimal bandwidth is even less variable in d_x and exhibits no singularity in the case of an ARFIMA(p,d_x,q) for $\max(p,q) > 0$. As might be expected for series exhibiting autoregressive moving average features, the optimal bandwidth is far smaller than in the previous case, fewer harmonic frequencies being used to avoid flawing the estimates. Note that in the ARFIMA(1,d_x,0) case, $\frac{f^{*\prime\prime}(0)}{2f^*(0)} = -\frac{a}{(1-a)^2}$, see Delgado and Robinson (1996), so that the optimal bandwidth tends to zero as the autoregressive coefficient a tends to one. For ARFIMA(1,d_x,0) series of length $n = 1000$, with autoregressive parameter $a = .5$, the theoretical optimal bandwidth is $m_{\text{opt}} = 60$ for all values of d_x, which indeed falls into the region which seemed optimal upon inspection of figures 2 to 5.

4.4 Approximations to the optimal bandwidths

In this section, infeasible and feasible approximations to the optimal bandwidths are proposed for the local Whittle, the log periodogram and the averaged periodogram estimates of long memory. We consider the smoothest specification in the class above, namely $\beta = 2$. An approximation to the optimal bandwidth $m_{\text{opt}} = (\frac{3n}{4\pi})^{\frac{4}{5}} \; |E_2(d_x)|^{-\frac{2}{5}}$ relies on a preliminary approximation of the unknown

Figure 4.6: Optimal bandwidth for the local Whittle estimate of long memory



E_2. The latter's dependence on the long memory parameter naturally points to an iterative procedure whereby

$$\hat{d_x}^{(k)} = \text{argmin}_{d\in[\Delta_1,\Delta_2]}\ R(\hat{m}^{(k)}, d), \tag{4.27}$$

$$\hat{m}^{(k+1)} = (\frac{3n}{4\pi})^{\frac{4}{5}}\ |E_2(\hat{d_x}^{(k)})|^{-\frac{2}{5}}, \tag{4.28}$$

starting from an ad hoc value $\hat{m}^{(0)} = n^{\frac{4}{5}}$ and with $\Delta_1 = -0.499$ and $\Delta_2 = 0.499$.

Consider the fractional representation 1.83. In that case, as pointed out in the previous section, $E_2(d_x)$ can be approximated by $\tau^* + \frac{d_x}{12}$ where $\tau^* = \frac{f^{*''}(0)}{2f^*(0)}$. In this framework, all three semiparametric estimates can have their optimal bandwidth formulae stated as follows:

<u>LW</u> For $-\frac{1}{2} < d_x < \frac{1}{2}$,

$$m_{\text{LW}} = \left(\frac{3}{4\pi}\right)^{\frac{4}{5}} \left(\tau^* + \frac{d_x}{12}\right)^{-\frac{2}{5}} n^{\frac{4}{5}} \tag{4.29}$$

<u>LP</u> For $0 < d_x < \frac{1}{2}$,

$$m_{\text{LP}} = \left(\frac{81}{96\pi^2}\right)^{\frac{1}{5}} \tau^{*-\frac{2}{5}} n^{\frac{4}{5}} \tag{4.30}$$

<u>AP$_1$</u> For $0 < d_x < \frac{1}{4}$,

$$m_{\text{AP}_1} = \left(\frac{(3-2d_x)^2}{4(2\pi)^4(1-4d_x)}\right)^{\frac{1}{5}} \left(\tau^* + \frac{d_x}{12}\right)^{-\frac{2}{5}} n^{\frac{4}{5}} \tag{4.31}$$

<u>AP$_2$</u> For $\frac{1}{4} < d_x < \frac{1}{2}$,

$$\begin{aligned} m_{\text{AP}_2} = \frac{n^{\frac{2}{3-2d_x}}}{2\pi} &\left\{ \frac{2\Gamma(1-2d_x)\cos((\frac{1}{2}-d_x)\pi)(3-2d_x)}{8} \left(\frac{2d_x-3}{2d_x E_2(d_x)}\right.\right. \\ &+ \frac{1}{|E_2(d_x)|}\left[\frac{(3-2d_x)^2}{(2d_x)^2(d_x+\frac{1}{2})^2} + 32\left(\frac{1}{2}-d_x\right)\left\{\frac{1}{2d_x(4d_x-1)}\right.\right. \\ &\left.\left.\left.\left. - \frac{1}{(4d_x+1)(d_x+\frac{1}{2})^2} - \frac{4\Gamma(2d_x)^2}{\Gamma(4d_x+2)}\right\}\right)^{\frac{1}{2}}\right]\right\}^{\frac{1}{3-2d_x}} \end{aligned} \tag{4.32}$$

First an infeasible procedure is considered in which τ^* is taken as known. We consider three models for the levels: ARFIMA(0,d_x,0), ARFIMA(1,d_x,0) and ARFIMA(1,d_x,1) with autoregressive coefficient $a = .5$ and moving average coefficient $b = -.45$ when applicable. Delgado and Robinson (1996) show that in the ARFIMA case,

$$\tau^* = \frac{b}{(1-b)^2} - \frac{a}{(1-a)^2} \tag{4.33}$$

so that the true values are $\tau^* = 0$ for the ARFIMA(0,d_x,0), $\tau^* = -2$ for the ARFIMA(1,d_x,0), and $\tau^* = -3.49$ for the ARFIMA(0,d_x,1). Series of length n = 1000 are simulated with the six error structures corresponding to models (i) to (vi) in chapter 3 and with long memory parameter values $d_x = -.2$ corresponding to antipersistence, $d_x = 0$ corresponding to short memory, $d_x = .2$ corresponding to moderate long memory and $d_x = .45$ corresponding to very long memory. For $d_x = -.2$, only 4.29 is theoretically applicable, but 4.30 and 4.31 are used nonetheless for the LP and the AP estimates respectively for comparison. $d_x = .2$ is within the range of applicability of 4.29, 4.30 and 4.31, whereas $d_x = .45$ is within the range of applicability of 4.29, 4.30 and 4.32. Series are simulated in the same way as in chapter 2 and optimal bandwidths are derived together with corresponding estimates of d_x using the recursive procedure defined by 4.27 and 4.28 for the LW and by

$$\hat{d_x}^{(k)} = d(\hat{m}^{(k)}), \tag{4.34}$$

$$\hat{m}^{(k+1)} = m_{\text{opt}}(\hat{d_x}^{(k)}, \tau^*) \tag{4.35}$$

for the LP and the AP. τ^* are taken as known.

A feasible approximation to the optimal bandwidth is then proposed following the lines of Delgado and Robinson (1996). It is based on an expansion of the semiparametric spectral density $f(\lambda) = |1 - \exp(i\lambda)|^{-2d_x} f^*(\lambda)$. $f^*(0)$ and $f^{*\prime\prime}(0)$ are taken to be respectively the first and last coefficient in the least squares regression of the periodogram $I_x(\lambda_j)$ against $|1 - \exp(i\lambda_j)|^{-2\hat{d_x}^{(0)}}$ $(1, \lambda_j, \frac{\lambda_j^2}{2})$ for $j = 1$ to $\hat{m}^{(0)}$. The results are significantly worse when the approximations to τ^* are updated at each iteration, convergence of the selected bandwidth is much slower and often fails altogether, and the results are not reported here. With no updating of the approximation to τ^*, convergence of the selected bandwidth occurs at the second iteration. Table 4.1 presents the feasible and infeasible long memory estimates (with selected

Table 4.1: Automatic estimates of long memory in introductory examples

Infeasible and feasible automatic estimates of long memory (with selected bandwidths in brackets) for the Nile river data and the ARFIMA(1,d_x,0) series presented in figures 4.2-4.5.

MODEL	LW		LP		AP	
	feasible	infeasible	feasible	infeasible	feasible	infeasible
Nile river	0.38(56)	–	0.28(109)	–	0.31(58)	–
$d_x = -.25$	-0.21(69)	-0.20(61)	-0.11(148)	-0.15(116)	-0.21(69)	-0.33(61)
$d_x = 0$	0.07(42)	0.10(61)	0.10(80)	0.17(116)	0.19(42)	0.10(61)
$d_x = .2$	0.35(23)	0.25(61)	0.28(46)	0.32(116)	0.25(55)	0.20(61)
$d_x = .45$	0.44(34)	0.49(61)	0.47(65)	0.49(116)	0.36(40)	0.39(61)

bandwidths) for the Nile river data and the four series simulated in figures 4.2-4.5. For the ARFIMA series, all selected values of bandwidth (feasible and infeasible with all three estimates) fall within the range which was identified on figures 4.2-4.5 as that where variability starts to decrease, but estimates have not yet started to increase steadily under the influence of the short range dynamics.

To investigate the improvement provided by choosing the bandwidth optimally in the local Whittle procedure, figures 4.7 and 4.8 respectively give plots of Monte Carlo biases and root mean squared errors against bandwidth for ARFIMA(1,d_x,0) series of length $n = 1000$ with autoregressive coefficient $a = .5$, d_x taking the values -.25, 0, .2. Monte Carlo RMSEs are compared with the theoretical RMSEs derived in section 4.3. Monte Carlo RMSEs are identical to the theoretical RMSEs and one sees that they are a very smooth function of bandwidth with a unique minimum at $m = 61$. RMSEs vary between 0.08 and 0.12 for bandwidths between 30 and 120. Choosing bandwidth optimally obviously provides a considerable improvement on the efficiency of the estimate and even approximations to that optimal bandwidth which are as low as 30 or as high as 120 will provide a considerable improvement on an estimation with ad hoc bandwidth of 200, for which the RMSE is 0.2. One notices that Monte Carlo biases of the local Whittle estimate of long memory are biased towards zero for low values of the bandwidth, whereas for bandwidths larger than 30, biases are equal for all true values of the long memory parameter, they are

Figure 4.7: Local Whittle biases against bandwidth

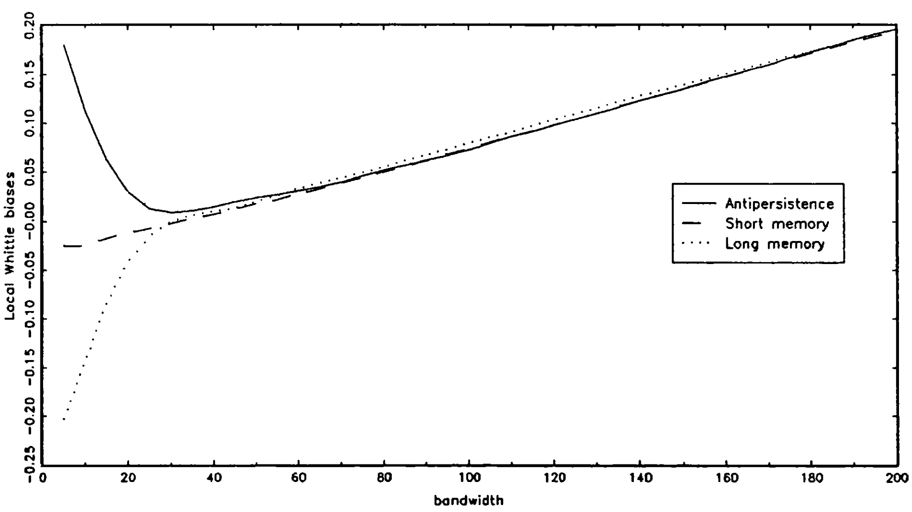
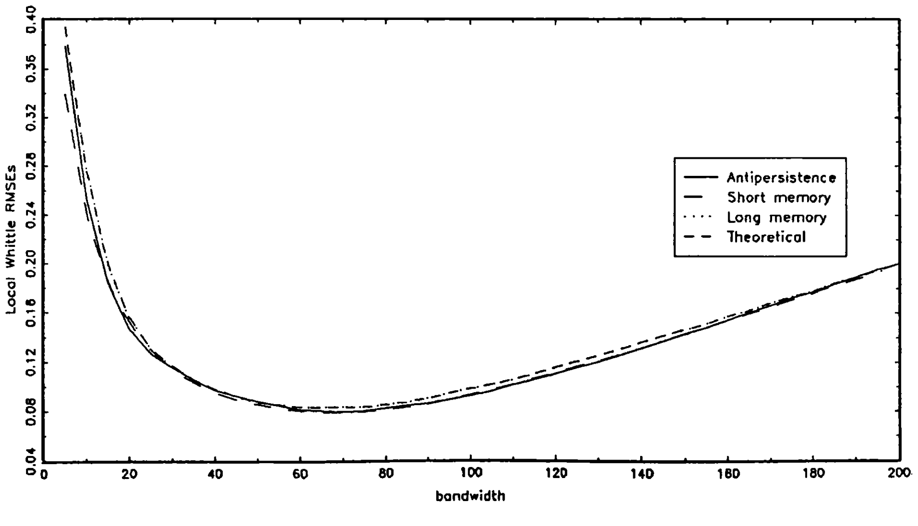


Figure 4.8: Local Whittle RMSEs against bandwidth



slightly positive and increase continuously and linearly with bandwidth.

Table 4.2 presents biases and RMSEs of local Whittle estimates of long memory using both the infeasible and the feasible procedures on series simulated according to the three models ARFIMA(0,d_x,0), ARFIMA(1,d_x,0) and ARFIMA(1,d_x,1), d_x taking the values -.25, 0, .2 and .45. The autoregressive and moving average parameter are $a = .5$ and $b = -.45$ respectively when applicable. The series were simulated 10000 times with sample size 1000 (the simulation techniques were discussed in Chapter 2), and the averaged selected bandwidths in each case are reported for first, second and final iterations. Looking first at results with the infeasible procedure, one notes that the bandwidth selected is always equal to the theoretical optimal bandwidth. Apart from very slight bias towards zero, the true value of d_x seems to have no influence in case of ARFIMA(0.d_x,0) series. In case of ARFIMA(1,d_x,0) series, the bias is more significant and positive. However, it seems to be cancelled by the introduction of an MA term. RMSEs are very similar in case of ARFIMA(1,d_x,0) and ARFIMA(1,d_x,1). The significant edge in terms of bias and RMSEs that is observed for $d_x = .45$ is due to the estimates being censored at .499. We now turn to the results of the feasible procedure. In case of ARFIMA(0,d_x,0) series, automatic bandwidth selection performs worse than the ad hoc choice $m = n^{4/5}$, but this does not affect quality of estimation very much due to the absence of short memory dynamics. Biases are essentially unchanged compared to the results of the infeasible procedure, and although RMSEs are twice as large as with the infeasible procedure, they remain small. The cases of ARFIMA(1,d_x,0) and ARFIMA(1,d_x,1) are very similar, Monte Carlo RMSEs being slightly worse in the latter case, due presumably to the near unit root short memory structure. Compared to the case of the infeasible procedure, the sign of Monte Carlo bias is reversed in 5 out of 8 cases with short memory structure; the bandwidth selected being lower than the optimal, bias is located on the left side of the node which is apparent in figure 4.7. Automatically selected bandwidths are significantly lower than optimal values in all but one of the 12 cases considered. In case $d_x = .2$, they fall as low as half the optimal value, which results in high RMSEs, although this does not eliminate

Table 4.2: Infeasible and feasible automatic local Whittle estimation

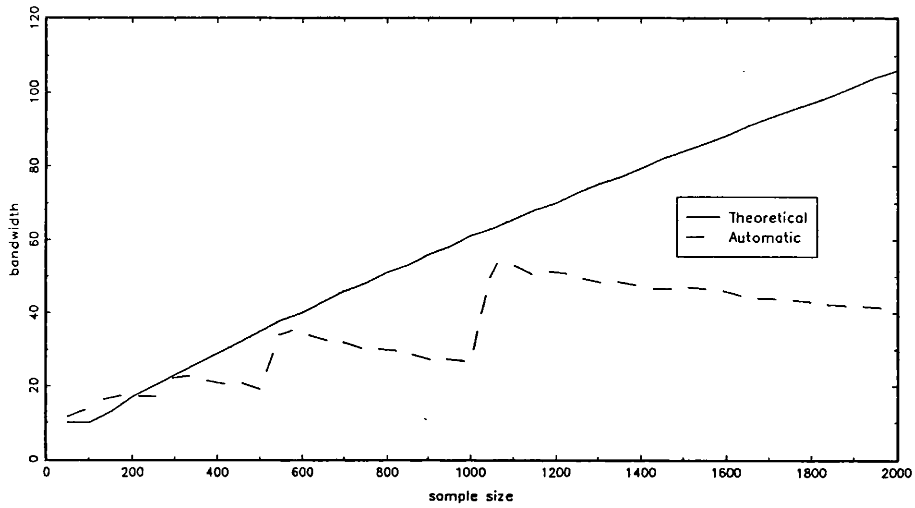
Monte Carlo Biases and Root Mean Squared Errors of local Whittle estimates of series following twelve specified models with bandwidths selected automatically using the infeasible and the feasible procedure respectively.

MODEL		ARFIMA(0,d_x,0)		ARFIMA(1,d_x,0)		ARFIMA(1,d_x,1)	
		infeasible	feasible	infeasible	feasible	infeasible	feasible
$d_x = -.25$	bias	0.013	0.002	0.030	0.050	0.007	0.031
	rmse	0.031	0.046	0.081	0.088	0.077	0.090
$d_x = 0$	bias	-0.001	-0.002	0.027	0.006	0.001	-0.012
	rmse	0.023	0.044	0.080	0.102	0.075	0.132
$d_x = .2$	bias	-0.012	-0.009	0.033	-0.059	0.005	-0.041
	rmse	0.030	0.059	0.083	0.187	0.076	0.204
$d_x = .45$	bias	-0.010	-0.021	0.013	-0.047	0.001	-0.053
	rmse	0.032	0.086	0.045	0.177	0.056	0.182

Averaged bandwidths selected with the infeasible and the feasible procedure in local Whittle estimation of long memory in series following twelve specified models.

MODEL		ARFIMA(0,d_x,0)		ARFIMA(1,d_x,0)		ARFIMA(1,d_x,1)	
		infeasible	feasible	infeasible	feasible	infeasible	feasible
$d_x = -.25$	$n^{(0)}$	256	256	256	256	256	256
	$n^{(1)}$	376	164	61	78	58	50
	$n^{(\infty)}$	376	163	61	77	58	51
$d_x = 0$	$n^{(0)}$	256	256	256	256	256	256
	$n^{(1)}$	498	174	61	44	58	30
	$n^{(\infty)}$	498	174	61	44	58	29
$d_x = .2$	$n^{(0)}$	256	256	256	256	256	256
	$n^{(1)}$	411	146	61	27	58	24
	$n^{(\infty)}$	411	146	61	26	58	23
$d_x = .45$	$n^{(0)}$	256	256	256	256	256	256
	$n^{(1)}$	299	64	61	44	58	44
	$n^{(\infty)}$	299	65	61	44	58	44

Figure 4.9: Automatic and optimal bandwidths for the local Whittle estimate

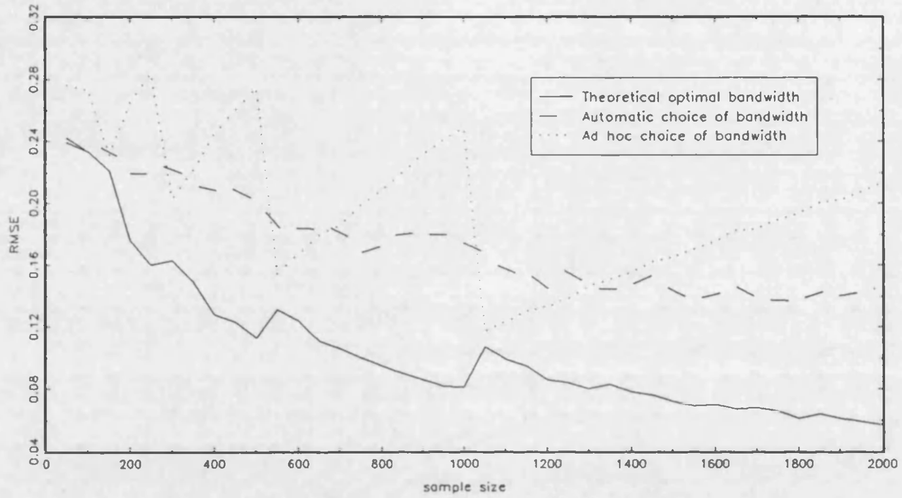


the usefulness of the iterative procedure proposed here, as improvement on ad hoc choice $m = 256$ remains significant. We plotted automatically selected and optimal bandwidths for 40 equally spaced values of sample size between 50 and 2000. They appear on figure 4.9, while figure 4.10 shows a plot of corresponding Monte Carlo RMSEs compared to RMSEs in case of ad hoc choice $m = \left[n^{4/5}\right]$ for bandwidth. Automatic bandwidth remains below optimal bandwidth, and the negative bias seems to increase with sample size in an alarming way. However, for large sample sizes, choice of bandwidth is likely to become a matter of less concern, as long as the optimal rate $n^{4/5}$ is maintained.[1]

Concentrating on the ARFIMA(1,d_x,0) model with autoregressive parameter $a = .5$, in view of the focus of section 4.2, it is of interest to know how the automatic long memory LW, LP and AP estimation procedures compare in the way they are affected by conditional heteroscedasticity in the innovations ε_t. Asymptotic results

[1]Series were simulated 1000 times according to and ARFIMA(1,.2,0) model.

Figure 4.10: RMSEs with optimal, automatic and ad hoc bandwidth choice



of sections 4.2 and 4.3 show that the optimal bandwidth formulae for the LW and AP estimates are not affected. However, their feasible approximations may be affected. Table 4.3 compares biases and mean squared errors of the three estimates using the infeasible and the feasible procedures on ARFIMA(1,.2,0) series simulated with the five error models described in section 2.6. In case of the infeasible procedure, bandwidth selection is unaffected by conditional heteroscedasticity. The effect of conditional heteroscedasticity on bias is non-existent. The effect of conditional heteroscedasticity on RMSEs is the same across estimates. As was observed in chapters 2 and 3, GARCH innovations lead to the worst performance for all three estimates of long memory. VLMARCH innovations leads to a better performance than GARCH, and LMARCH better still, whereas ARCH innovations lead to identical performances to i.i.d. innovations. Log periodogram performance is worse than local Whittle performance. In fact, the relative performance of the log periodogram equals its asymptotic value in case of normally and identically distributed innovations. The slightly better performance of the averaged periodogram is surprising considering asymptotic relative efficiency of the AP in case of i.i.d. innovations is .75. The feasible Monte Carlo results show that the iterative procedure seems better suited to the averaged periodogram and log periodogram estimates than to

Table 4.3: Sensitivity of automatic procedures to conditional heteroscedasticity

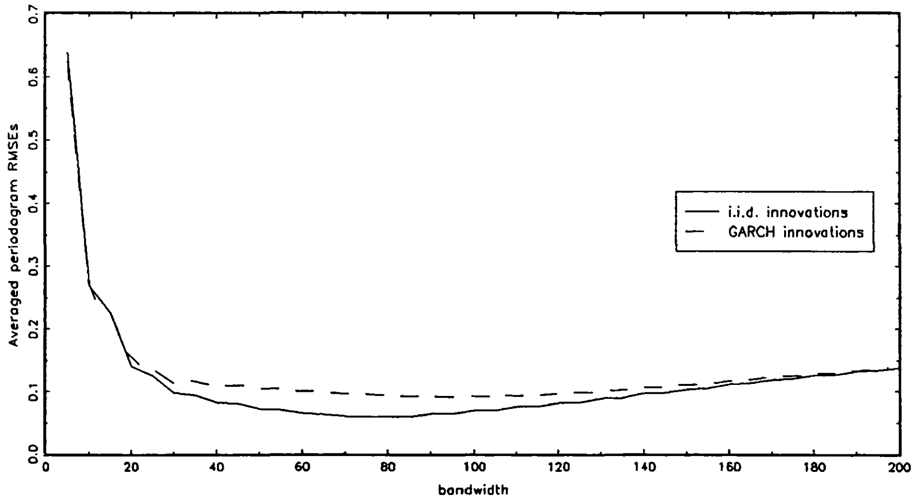
Monte Carlo Biases and Root Mean Squared Errors of local Whittle, log periodogram and averaged periodogram estimates of long memory in ARFIMA(1,.2,0) series with five specified innovation structures and bandwidths selected automatically using the infeasible and the feasible procedure respectively.

ERROR MODEL		LW		LP		AP	
		infeasible	feasible	infeasible	feasible	infeasible	feasible
IID	bias	0.032	-0.048	0.093	0.009	0.004	-0.030
	rmse	0.082	0.173	0.113	0.121	0.078	0.110
ARCH	bias	0.032	-0.044	0.093	0.013	0.004	-0.028
	rmse	0.082	0.171	0.113	0.114	0.076	0.108
GARCH	bias	0.022	-0.051	0.089	0.001	-0.007	-0.033
	rmse	0.121	0.193	0.140	0.146	0.110	0.136
LMARCH	bias	0.032	-0.047	0.094	0.009	0.003	-0.031
	rmse	0.088	0.175	0.117	0.117	0.081	0.116
VLMARCH	bias	0.028	-0.050	0.092	0.003	-0.005	-0.037
	rmse	0.109	0.191	0.128	0.139	0.101	0.136

Automatic bandwidths selected for local Whittle, log periodogram and averaged periodogram estimation of long memory in ARFIMA(1,.2,0) series with five specified innovation structures.

ERROR MODEL		LW		LP		AP	
		infeasible	feasible	infeasible	feasible	infeasible	feasible
IID	$n^{(0)}$	256	256	256	256	256	256
	$n^{(1)}$	61	27	116	60	61	59
	$n^{(\infty)}$	61	26	116	60	61	60
ARCH	$n^{(0)}$	256	256	256	256	256	256
	$n^{(1)}$	61	28	116	59	61	60
	$n^{(\infty)}$	61	26	116	59	61	61
GARCH	$n^{(0)}$	256	256	256	256	256	256
	$n^{(1)}$	61	29	116	59	61	51
	$n^{(\infty)}$	61	30	116	58	61	51
LMARCH	$n^{(0)}$	256	256	256	256	256	256
	$n^{(1)}$	61	26	116	56	61	49
	$n^{(\infty)}$	61	27	116	55	61	49
VLMARCH	$n^{(0)}$	256	256	256	256	256	256
	$n^{(1)}$	61	28	116	57	61	51
	$n^{(\infty)}$	61	29	116	57	61	51

Figure 4.11: Averaged periodogram RMSEs against bandwidth



the local Whittle. For local Whittle and log periodogram estimates, automatic bandwidths are significantly lower than optimal bandwidths, although this is less damaging to the log periodogram than to the local Whittle estimate. In case of average periodogram estimation, automatic bandwidths are equal to optimal bandwidths in case of i.i.d. and ARCH innovations, whereas they are 15% lower in case of GARCH, LMARCH and VLMARCH. To investigate whether the corresponding loss in precision of the averaged periodogram estimate when innovations follow a GARCH model is to be attributed partly to the bandwidth selection procedure, we plot RMSEs functions of bandwidth for ARFIMA(1,.2,0) series of length 1000 with autoregressive coefficient $a = .5$ and with i.i.d. and GARCH innovations.[2] The RMSE function is much flatter around optimal bandwidth (61) in case of GARCH innovations, and it is also significantly above that for i.i.d. innovations. The flatness of the RMSE function in case of GARCH innovations explains the slightly less efficient automatic bandwidth selection but at the same time makes it irrelevant to the precision of the estimate. The reduced efficiency is due to the effect of near unit root conditional heteroscedasticity on RMSEs for all bandwidths.

[2]the Monte Carlo values used to generate the graph were derived with with 10000 replications.

Table 4.4: Automatic local Whittle estimation of long memory in fractional Gaussian noise series

Monte Carlo Biases and Root Mean Squared Errors of automatic local Whittle estimates of long memory in fractional Gaussian noise series with four values of the self-similarity parameter.

H	.25	.5	.7	.95
bias	-0.014	-0.001	0.001	0.006
rmse	0.067	0.030	0.048	0.066

Finally, recalling that the idea for the approximation of $E_2(d_x)$ was based on the fractional representation 1.83, it is of interest to see how the automatic selection procedure described above performs on a fractional Gaussian noise series with autocorrelations given by 1.41. Series of length $n = 1000$, autocorrelations given by 1.41 and variance 1 are simulated using the algorithm of Davies and Harte (1987). Monte Carlo biases and root mean squared errors are reported for $H = .25$, $H = .5$, $H = .7$ and $H = .95$ in table 4.4. Apart from the fact that estimates of long memory in fractional noise series seem to be biased away from zero, whereas they were biased towards zero for ARFIMA(0,d_x,0) series, automatic local Whittle performance does not seem affected by the model chosen for the series. Performance is (if anything) slightly better on fractional noise series than on ARFIMA(0,d_x,0) series.

4.5 Conclusion

This chapter has addressed the question of bandwidth choice for estimates of long memory analysed in chapters 2 and 3, the averaged periodogram and the local Whittle estimates. The need for motivated bandwidth selection is made apparent by looking at sensitivity of the estimates to bandwidth. Existing theory on optimal bandwidth for the averaged periodogram estimate was shown to hold in case of (possibly long memory) conditionally heteroscedastic innovations for the process. An optimal bandwidth formula was derived for the local Whittle estimate with a heuristic justification unaffected by innovation conditional dependence structure.

An automatic iterative procedure was assessed for these estimates, primarily on ARFIMA series with simple short range dynamics, and it was shown to produce excellent results (irrespective of conditional dependence structures of the innovations) in case of the averaged periodogram estimate. Further research into possible theoretical justification for this optimal bandwidth iterative approximation would therefore be needed. The slightly less convincing results from automatic local Whittle estimation do not eliminate the usefulness of the procedure in that case, but the search for a more efficient one may be warranted.

Chapter 5

Analysis of dependence in intra-day foreign exchange returns

Joint work with Richard Payne

5.1 Introduction

In recent years, a vast amount of empirical work has been devoted to the characterisation of temporal dependence in financial time series. Many authors have examined the time series structure in asset returns, trading volumes and, perhaps more extensively, return volatility. Such studies are valuable in that they yield insights into issues such as the discrimination between regular and irregular market activity, the nature of information flows into financial markets, the way in which this information is assimilated into asset prices and the manner in which information is transmitted between markets. This chapter extends the research in this area with an empirical analysis concentrated on modelling the volatility process associated with a year long intra-day sample of three major exchange rates.

Financial returns are now widely recognised to exhibit non linear features such as volatility clustering, leptokurtosis, and various distributional asymmetries. Volatility clustering and leptokurtosis are traditionally accounted for by allowing conditional variance to vary across time, as in 1.27. The central issue, therefore, is the type of process followed by the conditional variance σ_t^2 insofar as it provides a way of determining whether σ_t^2 is stationary, whether shocks are persistent in σ_t^2, and more precisely, of measuring the degree of temporal dependence in σ_t^2.

Analysing issues of temporal dependence in a time series requires obviously as large a sample size as physically possible, all the more so if one wishes to disentangle short run from long run dependencies through use of semiparametric or nonparametric procedures. However, structural breaks may spuriously increase the measure of dependence in data collected over long time spans under the presumption of structural stability. This contradiction may be resolved if one turns to data sampled at higher frequencies (typically within one day) recently made available from news screens such as Reuters or Bloomberg. This approach is advocated by Goodhart and O'Hara (1997). The validity of this approach is supported by Nelson's result on continuous sampling of diffusion processes (Nelson (1990a)). He proves that processes following certain stochastic difference equations converge in distribution to well defined solutions of stochastic differential equations when the time interval tends to zero, and that GARCH(1,1)-M and Exponential ARCH(1) models in particular have diffusion limits. One would therefore expect analysis of high frequency data to yield more accurate results on the nature of the temporal dependence in σ_t^2. In particular, the analysis of long high frequency financial data series warrants the use of a long memory paradigm in the volatility equation, be it in the ARCH or the Stochastic Volatility framework, provided the strong intra-day seasonality characteristics of these series -time of the day effects described in section 2- are controlled for. The use of the long memory paradigm for the volatility process is moreover empirically justified by the findings of correlograms for these intra-day volatility series which decay far more slowly than the exponential decay which is associated with conventional GARCH or Stochastic Volatility models. Finally, the use of the long memory paradigm is highly desirable for the flexible representation it provides and also as a framework to define and test persistence concepts. The ARFIMA model

1.11 for long memory time series is a particularly suitable and flexible alternative to standard ARIMA, permitting a far more general characterisation of the temporal dependencies in a given time series. An ARFIMA(p,d,q), with p and q the orders of the autoregressive and moving average polynomials respectively, satisfying 1.11, is a process which is integrated of order d, labelled I(d). I(0) corresponds to the weakly dependent ARMA process, while I(1) corresponds to a process with a unit root. $d > -\frac{1}{2}$ ensures invertibility of the ARFIMA while $d < \frac{1}{2}$ ensures covariance stationarity. For $d < 1$, the process can be said to be "mean reverting", a concept which is different from return to initial position with probability one, a feature of the random walk. It is mean reverting in the sense that, if x_t is I(d) with $d < 1$, $E(x_{t+1}|x_t > E(x_t)) < x_t$. A finer notion of persistence of innovations on the process (or lack of it) may also be derived with a long memory structure, in that the j-th impulse response coefficient of an I(d) process is of order $O(j^{d-1})$, the larger d, the greater the persistence of shocks on the process.

This chapter proposes a comprehensive methodology for assessing the nature of temporal dependence. The methodology entails the following steps. The first step is to test the order of integration in the process using the methodology presented in Robinson (1994a). The test is based on an underlying ARFIMA structure for the series in question and permits any degree of integration (integer or fraction) as a null hypothesis. Next, we gain a precise estimate of the degree of integration using each of the three robust semiparametric estimates of long memory which have been described in the preceding chapters: the LW, the LP and the AP. At this point, and analogous to the I(1) case, one can filter the long range dependence from the series and fit a covariance stationary ARMA to the residuals using traditional model selection procedures. We go on to fit a fully parametric model to the series. The model is an extended version of the Long Memory in Stochastic Volatility (LMSV) model given in Harvey (1993), allowing for both short and long range dependence, and is sensitive to any short range mis-specification due to its fully parametric nature. The reason for fitting the fully parametric model is to permit one to assess the contributions of the short and long memory components to the overall dependence in the series. This is accomplished via a set of Quasi-Likelihood Ratio statistics for the fully parametric model. Hence, ultimately, we can discriminate between the long and short range dependent features of the process.

The data we examine in this work are the volatilities associated with three intra-day foreign exchange (FX) return series (the exchange rates in question being the DEM/USD, JPY/USD and JPY/DEM.) A pervasive result from previous work on this type of data is that the volatility process can be characterised as non-stationary (see, inter alia, Andersen and Bollerslev (1997b), DeGennaro and Shrieves (1995) and Guillaume (1995).) There are however, some indications that this result may be due to mis-specification of the volatility models employed. First, the temporal aggregation results for GARCH processes do not hold when applied to FX data. The degree of persistence one identifies in daily data, for example, far exceeds that which would be implied by the results of estimations from data sampled at 1 hour intervals.[1] Second, other authors (e.g. Dacorogna, Müller, Nagler, Olsen, and Pictet (1993)) have noted that the correlograms of these intra-day volatility series decay far more slowly than the exponential decay which is associated with conventional GARCH or SV models. The combination of these two points serves as the motivation for our investigation of long-memory in volatility.

A theoretical motivation for the presence of long memory in asset price volatility can be generated by combining the simple mixture of distributions model in Tauchen and Pitts (1983) and the results on aggregation in Robinson (1978b) and Granger (1980). The former demonstrate, in a highly stylised framework, that both the volume and volatility in asset markets inherit the temporal dependencies associated with the latent flow of information into the market. Now assume that information flows are heterogeneous. Specifically, assume that there are an infinity of information arrival processes, each of which follows a stationary autoregression. The heterogeneity is modelled by variation in the AR parameters, which we assume follow a beta distribution. As Robinson (1978b) demonstrates, the aggregate information flow process will then exhibit long range dependence and, hence, so will volatility.

[1]See Andersen and Bollerslev (1997b), for example.

This would imply not only that the long memory component in volatility is a feature inherent to the returns generating mechanism as opposed to the consequence of omitted nonlinearities, but also that sharing the same aggregate information flow process (which is a reasonable assumption to make on the exchange rate market for major currencies such as USD, Yen and DM), several returns generating mechanisms would also share the same long memory component in volatility and hence display fractional cointegration in volatilities. Such a feature is tested for and fractional cointegration estimated according to the methodology described in Robinson (1994c) and chapter 2 of this thesis.

The chapter is set out as follows. Section 2 introduces the data employed in the study. As previously mentioned, we focus on intra-day foreign exchange rate volatilities, which are sampled at ten minute calendar intervals. In section 3, we present a more detailed account of the empirical methodology. Section 4 presents estimation and testing results. We find that the foreign exchange return process is well characterised by an I(0) process, in line with the Efficient Markets Hypothesis. Results for the three volatility series demonstrate that all are covariance stationary and exhibit significant long memory. Further, estimation and testing of the fully parametric model demonstrates that the finding of non stationarity in foreign exchange volatility in previous work is due to mis-specification. When one permits the possibility of long memory in volatility, all specifications strongly indicate covariance stationarity. Finally, investigation of cointegration between two of the three series (to avoid a circularity effect) strongly indicate that the long memory component is shared between the series, a tentative evidence of the validity of the above interpretation.

5.2 The Data

As indicated in the Introduction, the focus of this work is the behaviour of volatility in the intra-day Foreign Exchange (FX) market. We study three sets of FX returns, on the DEM/USD, JPY/USD and JPY/DEM, covering the period from

Table 5.1: Summary statistics for exchange rate returns

Rate	Mean	s.d.	Skew	Kurtosis	ρ_1	ρ_2	ρ_3	$Q(10)$
DEM/USD	6×10^{-6}	0.001	0.16	9.61	-0.076	-0.040	-0.005	306.8
JPY/USD	-4×10^{-4}	0.074	-0.06	13.85	-0.09	-0.015	0.0035	334.8
JPY/DEM	-5×10^{-4}	0.045	-0.25	7.93	0.0066	-0.0004	0.0009	13.5

Notes: the coefficients ρ_1, ρ_2 and ρ_3 represent the first through third sample autocorrelations respectively. $Q(10)$ is the tenth order Box-Ljung test statistic. The Box-Ljung statistic is distributed χ^2_{10} and has critical value 23.2 at 1%.

the beginning of October 1992 to the end of September 1993.[2] These return series are filtered transcriptions of the tick-by-tick quotation series which appear on the Reuters FXFX page. Each quote encompasses a timestamp, bid and ask quotation pair, plus identifiers which allow one to determine the inputting bank and its location. In this study we ignore the identification of the inputting institution, using the tick-by-tick data solely to construct a homogenous time-series in calendar time.

The basic horizon over which we calculate returns is 10 minutes.[3] This yields, for each currency, a time-series with 37583 observations. The basic summary statistics of the returns are shown in table 5.1 .

The above table illustrates the following facts. First, all three return series have a mean which is insignificantly different from zero. A point which conforms with many earlier studies is that there is pronounced excess kurtosis in the returns distribution. This, as pointed out by Bollerslev and Domowitz (1993), is a natural

[2]These data were supplied by Olsen and Associates (Zurich), to whom we are most grateful.

[3]Returns are determined as follows: at each 10-minute observation point the last mid-quote entered into the system is taken as the market price. We then first difference this quote series to obtain returns. At points when no quote is entered in a 10 minute interval, an artificial quote is calculated by linear interpolation between the nearest preceding and succeeding quotes. Finally, all weekend quotes are eliminated from the analysis due to the lack of FX market activity at these times. We define weekends as 21:00 GMT Friday to 21:00 GMT Sunday. Note also that the results presented in this paper carry over to the analysis of percentage returns.

Table 5.2: Summary Statistics for the Logarithm of Squared Returns

Rate	Mean	s.d.	Skew	Kurtosis	ρ_1	ρ_2	ρ_3	$Q(10)$
DEM/USD	-14.67	1.28	0.828	0.072	0.281	0.244	0.220	15948.03
JPY/USD	-7.68	3.36	-1.313	1.24	0.24	0.183	0.166	8433.88
JPY/DEM	-8.66	3.16	-0.99	0.45	0.353	0.29	0.264	20463.21

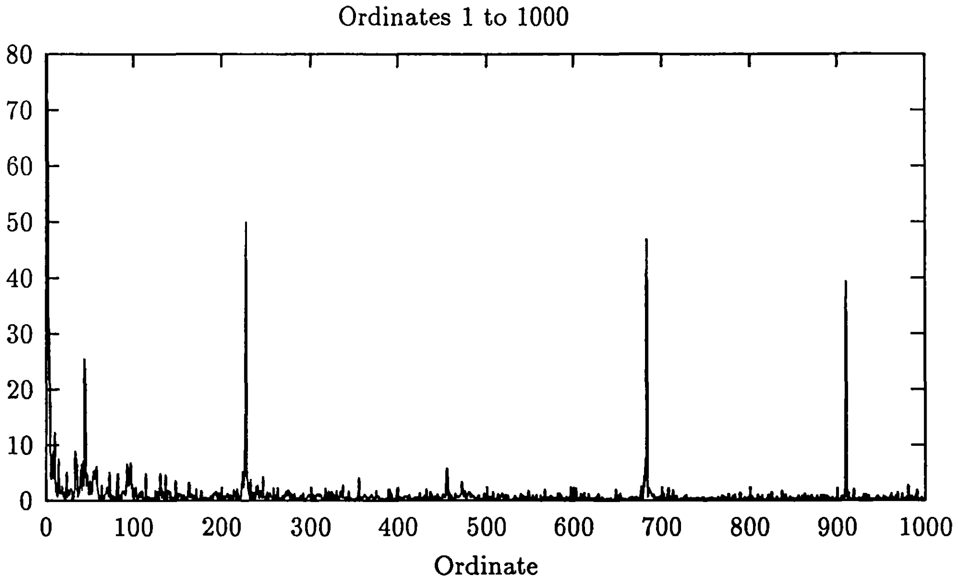
Notes: the coefficients ρ_1, ρ_2 and ρ_3 represent the first through third sample autocorrelations respectively. $Q(10)$ is the tenth order Box-Ljung test statistic. The Box-Ljung statistic is distributed χ^2_{10} and has critical value 23.2 at 1%.

feature of time-series which display conditional heteroskedasticity, although their analysis shows that after correcting for the conditional heteroskedasticity much of the kurtosis remains. Finally, the autocorrelation coefficients show that there is some temporal dependence in the return series, the DEM/USD and JPY/USD demonstrating negative autocorrelation whilst the JPY/DEM displays positive first-order autocorrelation. The significance of these autocorrelation coefficients is confirmed in the Box-Ljung statistics, which demonstrate that one cannot reject the hypothesis of up to tenth order serial correlation.

In table 5.2 we present identical sets of statistics for our volatility proxy. We employ the logarithm of squared returns as our volatility measure, a choice which is motivated by the Long Memory in Stochastic Volatility model which is presented in Section 3. The main feature of these results lies in the correlation structure of volatility. As is visible from comparing tables 5.1 and 5.2, there is far larger dependence in volatility than in returns. The first-order autocorrelations are between 3 and 5 times greater for volatility than for returns, whilst the Box-Ljung statistics are, at least an order of magnitude greater. The characterisation of this temporal dependence is the focus of this work.

In order to clarify the nature of the dependencies in volatility in Figures 5.1 to 5.3 we present the first 1000 periodogram and logged periodogram ordinates for the

Figure 5.1: Periodogram for JPY/USD Log Squared Returns



YEN/USD volatility plus the first 1000 sample autocorrelations.[4]

Examining first the correlogram, one feature which is immediately apparent is the existence of a pronounced daily seasonal in volatility. This seasonal has recently been the subject of many papers, including Dacorogna, Müller, Nagler, Olsen, and Pictet (1993), Andersen and Bollerslev (1997b) and Payne (1996). It is generated by the 24 hour activity in the foreign exchange market and the alterations in market activity which occur as trading shifts from the Far East to Europe to North America and so on. There is also evidence of seasonality at the weekly frequency. In the current context, however, this component is of no intrinsic interest and simply masks the underlying temporal structure of volatility. Hence, when estimating our long memory specifications we filter this component.

In the periodogram of the data this seasonal component is represented by peaks at integer multiples of the fundamental seasonal frequency.[5] A feature of the peri-

[4]Throughout the work we present graphical examples for this currency only as those for the other two currencies are qualitatively similar.

[5]As there are 144 ten minute intervals in one day, the seasonal frequency is $\frac{2\pi}{144}$, corresponding,

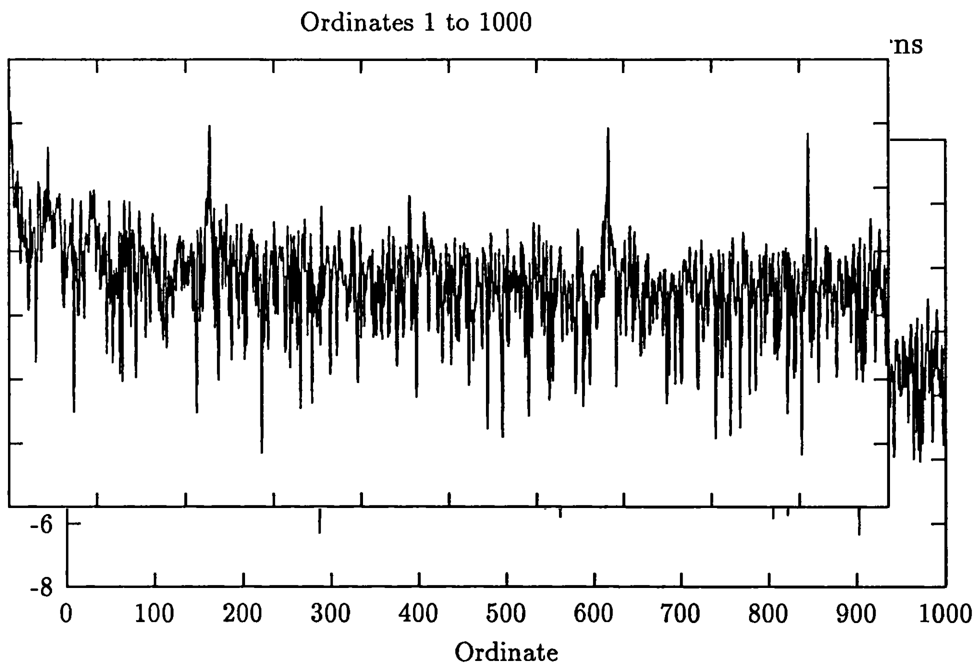
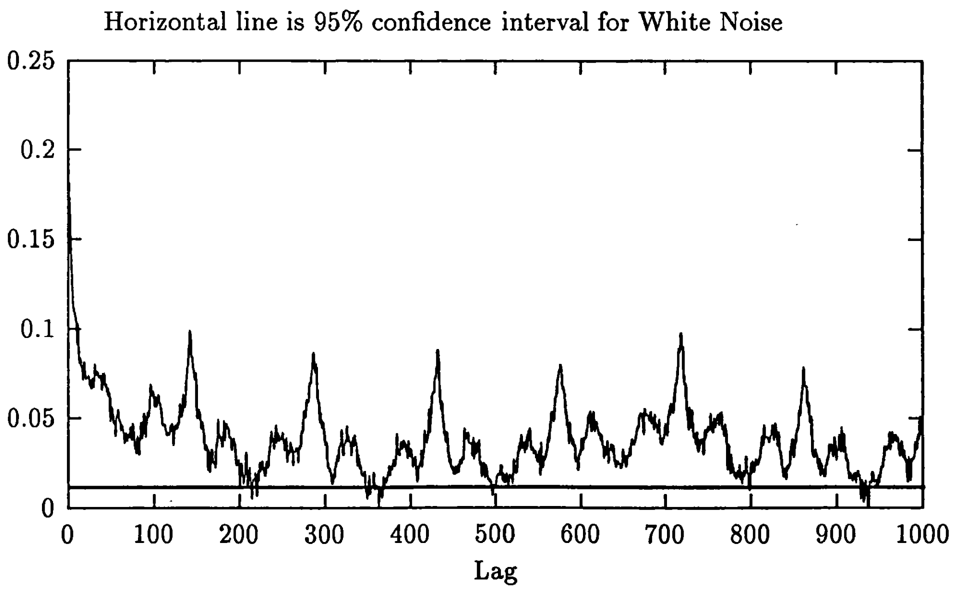


Figure 5.3: JPY/USD Log Squared Returns: Sample Autocorrelations 1 to 1000



odogram which is more relevant to the current study is the behaviour of the periodogram in a neighbourhood of zero frequency, where the peak (visible on Figures 1 and 2) can be viewed as tentative evidence for the presence of long memory in our volatility series.

A salient feature of these data sets is the seasonality in volatility described in more detail in Payne (1996) and Andersen and Bollerslev (1997b). Seasonal components appear in the periodogram as peaks at certain harmonic frequencies. These peaks effect all periodogram based estimation. The parametric estimation becomes invalid, and the efficiency of the robust estimates is significantly reduced.

If one thinks of the spectrum of the process with strong seasonal components as a mixed spectrum, there is a need for spectral estimation methods which remove the Dirac mass points at the seasonal frequencies and smooth out the leakage from these peaks into the neighbouring frequencies. Sachs (1994) proposes a peak insensitive non-parametric procedure to estimate the continuous part of the spectrum, treating the periodic components as outliers (so that it does not permit the estimation of the discrete component in the spectral density). Kooperberg, Stone, and Truong (1995a) propose a fully integrated estimation procedure for both the continuous and the discrete parts of the spectrum. Only related asymptotic results are proposed (see Kooperberg, Stone, and Truong (1995b)), the method is computationally very expensive and there is no indication that it deals with leakage efficiently. The spectral estimate used here is a Double-Window smoother proposed by Priestley (1981) which is designed to remove seasonal components and the leakage around the seasonal frequency. Suppose the volatility series is decomposed into two uncorrelated components $x_t = z_t + \varsigma_t$ where z_t has a continuous spectral density and $\varsigma_t = \sum_{r=1}^{K} A_r \cos(\omega_r t + \phi_r)$. The examination of this mixed spectrum is greatly simplified by the knowledge of the seasonal harmonics ω_r, which correspond to the weekly frequency and multiples of the daily frequency.[6] The amplitudes of the sea-

approximately, to harmonic 228.

[6]Hence we can avoid employing tests to detect harmonic components (Whittle, Bartlett, Hannan or Priestley, in Priestley (1981)).

sonal harmonics are estimated through a regression of x_t against $(\cos(\omega_r t + \phi_r))_{r=1}^{K}$, and the spectrum of z_t is consistently estimated with a Double-Window smoother. The spectral window adopted is the Bartlett-Priestley window

$$W(\theta; M) = \begin{cases} \frac{3M}{4\pi}\left\{1 - \left(\frac{M\theta}{\pi}\right)^2\right\}, & |\theta| \leq \frac{\pi}{M}, \\ \\ 0, & |\theta| \geq \frac{\pi}{M}, \end{cases}$$

where M is the bandwidth.[7] Call $\hat{f}_M(\omega) = \int_{-\pi}^{\pi} I(\theta) W(\omega - \theta; M) d\theta$ the spectral estimate using $W(\theta; M)$. The Double Window spectral estimate is constructed as follows:

$$\hat{f}_{DW}(\omega) = \begin{cases} \hat{f}_m(\omega), & |\omega - \omega_r| > \frac{\pi}{m}, \\ \\ (\hat{f}_l(\omega) - c\hat{f}_m(\omega))/(1-c), & |\omega - \omega_r| \leq \frac{\pi}{m}, \end{cases} \tag{5.1}$$

where $m > l$, $c = W(0; l)/W(0; m)$ and the ω_r's are the harmonics of the seasonal components defined above. A cross-validated likelihood maximising procedure for the determination of both bandwidths (see Hurvich (1985), Beltrão and Bloomfield (1987) and Robinson (1991a) for the asymptotics) proved computationally too expensive and gave poor results. An ad hoc choice of bandwidths $m = \sqrt{n}$ and $l = m/10$ was preferred.

5.3 Methodology

Let $\{r_t\}^n$ be the series of raw returns and define $x_t := \log r_t^2$ as a proxy for the volatility. Assume x_t admits spectral density with representation 1.18 with $0 < L(\lambda) = G < \infty$. Bearing in mind that the LW is justified for all values of d_r in a compact subset of $(-\frac{1}{2}, \frac{1}{2})$, that the LP and AP are justified for $0 < d_r < \frac{1}{2}$, a preliminary test of stationarity and invertibility is required for the log squared returns x_t.

[7]This spectral window is a smoothed version of the Daniell (or rectangular) window and it is chosen for its compact support.

5.3.1 Testing for persistence, long range dependence and stationarity

The testing procedure presented here is fully parametric, and sensitive, as indicated above, to mis-specified short range dynamics. The conclusions of the test need therefore to be confirmed after the model selection stage. This testing procedure relies on efficient tests of long range dependence (Robinson (1994a) and Gil-Alaña and Robinson (1997)) which permit a wide class of null hypotheses. The object is the test of the hypothesis of persistence in foreign exchange volatility. As was mentioned above, most of the available methods for testing for unit roots (see Diebold and Nerlove (1989) for a review on the subject) have non standard limiting distributions and lack Pitman efficiency.[8] Unit root tests against autoregressive alternatives, in particular, are based on the Wald, Likelihood Ratio and Lagrange Multiplier principles, but they lack the sufficient degree of smoothness across the parameter of interest that would yield null χ^2 limiting distributions and Pitman efficiency. Indeed, a process following the autoregression $x_t = \rho x_{t-1} + \epsilon_t$ is weakly dependent for $|\rho| < 1$, non-stationary for $\rho = 1$ (the unit root case) and explosive for $|\rho| > 1$.

Moreover, these tests give only one possible persistence null hypothesis. The testing procedure used here, on the other hand, allows one to postulate any value of d (integer or fraction) as a null hypothesis and possesses efficiency and a null χ^2 limiting distribution. In the fully parametric LMSV model, the volatility satisfies $(1-L)^d(1-\phi L)x_t = \eta_t$ which can be rewritten as $(1-L)^d x_t = u_t$ where u_t is a stationary AR(1), therefore I(0), process. u_t has spectral density

$$f_u(\lambda;\phi) = \frac{\sigma_\eta^2}{2\pi}\left[\frac{1}{1-2\phi\cos\lambda+\phi^2}\right].$$

Suppose we want to test the hypothesis $\mathrm{H}_0 : d = d_0$. Let $I_u(\lambda)$ be the periodogram of the residuals $\tilde{u}_t = (1-L)^{d_0}x_t$. The frequency domain quasi-likelihood is

$$\mathcal{L}(\sigma_\eta^2,\phi) = -\sum_{j=1}^{n-1}\log(2\pi f_{uj}) - \sum_{j=1}^{n-1}\frac{I_u(\lambda_j)}{f_{uj}} \tag{5.2}$$

where $f_{uj} = f_u(\lambda_j;\phi)$. Concentration of this likelihood yields $\sqrt{n}$-consistent estimates $\hat{\phi} = \mathrm{argmin}_\phi \sigma_\eta^2(\phi)$ and $\hat{\sigma_\eta^2} = \sigma_\eta^2(\hat{\phi})$ where $\sigma_\eta^2(\phi) = \frac{2\pi}{n}\sum_{j=1}^{n-1}\frac{I_u(\lambda_j)}{f_{uj}}$.

[8] These tests are improved in Elliot, Stock, and Rothenberg (1994)

The test statistic is constructed on the score principle. Let ω be the $(n-1)\times 1$ vector with j-th element $\log|4\sin^2(\frac{\lambda_j}{2})|$, let $\hat{f}_u$ be the $(n-1)\times 1$ vector with j-th element $f_u(\lambda_j;\hat{\phi})$, and let M be the projector on the space orthogonal to the $(n-1)\times 1$ vector with j-th element $\frac{\partial}{\partial\phi}\log f_u(\lambda_j;\hat{\phi})$. The test statistic is

$$\hat{\mathcal{S}} = -\frac{\pi}{\hat{\sigma}_\eta^2}\frac{\omega'\hat{f}_u}{\|M\omega\|}.$$

Under suitable regularity conditions (Robinson (1994a)), $\hat{\mathcal{S}} \xrightarrow{\mathcal{D}} N(0,1)$ as $n\to\infty$. The resulting testing rules for H_0 are summarized in the table below:

Alternative Hypothesis	Reject H_0 when
$\mathrm{H}_1 : d > d_0$	$\hat{\mathcal{S}} > z_\alpha$
$\mathrm{H}_1 : d < d_0$	$\hat{\mathcal{S}} < -z_\alpha$
$\mathrm{H}_1 : d \neq d_0$	$\hat{\mathcal{S}} > z_{\alpha/2}$

Note: Rules for α-level tests of $H_0 : d = d_0$ against various alternatives. z_α is the quantile of a standard normal variate.

This testing procedure provides us with two efficient tests of persistence: the null $H_0 : d = 1$ against the alternative $H_1 : d < 1$, which is the unit root test, and the null $H_0 : d = 1/2$ against the alternative $H_1 : d < 1/2$, which is a non-stationarity test.[9]

5.3.2 Estimation

The three semiparametric estimates of long memory discussed previously, the LW, the LP and the AP, are applied to the returns r_t and the log squared returns x_t. Standard errors based on asymptotic variances are included when applicable. The semiparametric techniques rely on the specification of the spectral density on a degenerating band of frequencies. They are therefore based on the concentration of the variance of the process in a neighbourhood of frequency zero and are insensitive

[9]I(d) processes can be seen as increasingly nonstationary as d increases from 1/2 to 1.

to any short memory behaviour of the series. Such short range dependent behaviour in the series, if mis-specified in a fully parametric model, will bias the estimation of the long range dependent parameter itself. Thus, the robust method advocated above could serve as a pre-estimation technique and enable us to create a fractionally differenced series $\Delta^{\hat{d}} x_t$ which, as noted in chapter 3 is an asymptotically valid approximation to an I(0) series without any parametric assumption on the autocorrelations of the underlying I(0) process $\Delta^d x_t$. On this differenced series, traditional model selection methods (using the AIC for instance) may be carried out to identify the order of a covariance stationary ARMA model for instance.

Our purpose in using the semiparametric estimates is to yield robust pre-estimates of d which can be compared with those obtained from a fully parametric model which permits both long and short range dependence. If then there are no signs of systematic bias in the estimates obtained from the fully parametric model, we can proceed to compare the contributions made to overall temporal dependence by each of the long and short memory components.

The parametric model we adopt is an extension of the LMSV model of Harvey (1993):

$$\begin{cases} x_t = c + h_t + \xi_t \\ \\ (1-L)^d(1-\phi L)h_t = \eta_t \end{cases} \tag{5.3}$$

where c is a constant, ξ has mean zero and variance $\pi^2/2$, $\eta \sim N(0, \sigma_\eta^2)$ and d lies within the stationarity and invertibility range $(-1/2, 1/2)$.[10] This framework is consistent with the semiparametric specification described above insofar as the spectral density of a process thus specified follows 1.18.

The estimation procedure is a frequency domain quasi log likelihood maximisation[11].

[10]This specification for the log squared returns is derived from the formulation in footnote 3, hence $\xi_t = \log(\epsilon_t^2)$ is distributed as a $\log \chi^2$ variate.

[11]Note that in this case, the maximisation is performed over the whole range of harmonic fre-

Asymptotic distributional results are derived as a special case of the work of Heyde and Gay (1993). Letting $I_x(\lambda)$ denote the periodogram of the process of log squared returns defined as in 1.53, the estimates of the fractional differencing parameter $\hat{d}$ and of the autoregressive parameter $\hat{\phi}$ maximise

$$\mathcal{L}(d,\phi) = -\sum_{j=1}^{n-1} \log(2\pi g_j) - \sum_{j=1}^{n-1} \frac{I(\lambda_j)}{g_j} \tag{5.4}$$

where

$$g_j = \frac{\sigma_\eta^2 |4\sin^2(\frac{\lambda_j}{2})|^{-d}}{2\pi(1-2\phi\cos\lambda_j + \phi^2)} + \sigma_\xi^2.$$

The first order autoregressive short range dependent specification in volatility is chosen for comparison with the traditional stochastic variance specification. Both specifications naturally suffer from the ignored nonlinearity in the ξ_t which is likely to affect x_t. But this new framework allows us to discriminate long memory and strong autoregressive effects with a simple quasi-likelihood ratio test. The AR(1)-LMSV model 5.3 is compared to two nested alternatives

$$(LM) \begin{cases} x_t = c + h_t + \xi_t \\ \\ (1-L)^d h_t = \eta_t \end{cases} \quad \text{and} \quad (AR) \begin{cases} x_t = c + h_t + \xi_t \\ \\ (1-\phi L) h_t = \eta_t \end{cases} \tag{5.5}$$

obtained for $\phi = 0$ and $d = 0$ respectively. The frequency domain likelihood is computed as in 5.4 for each of these nested models and the following tests are performed using the likelihood ratio principle:

$$\begin{gathered} H_0 : d = 0 \quad \text{against} \quad H_a : d > 0. \\ \\ H_0 : \phi = 0 \quad \text{against} \quad H_a : 0 < |\phi| < 1. \end{gathered} \tag{5.6}$$

If the subscript $._u$ denotes the unconstrained estimates, the likelihood ratio statistic is $2\left[\mathcal{L}(\hat{d}_u, \hat{\phi}_u) - \mathcal{L}(\hat{d}, \hat{\phi})\right]$ where $\mathcal{L}$ is the concentrated form of the asymptotically χ^2 quasi-likelihood in 5.4.

quencies. In the semiparametric case, only a degenerate band of harmonic frequencies was used. The present estimate is therefore sensitive to any short range dependent mis-specification. It is likely to be sensitive to the seasonal component in the series discussed in Section 2.4, but it is nonetheless reported before as well as after deseasonalisation for completeness.

5.3.3 Stationary cointegration

The presence of long memory in the log squares of essentially martingale difference returns can be accounted for in a version of the mixture of distributions hypothesis Clark (1973), Epps and Epps (1976) and Tauchen and Pitts (1983) which stipulates that the conditional variance of uncorrelated returns is driven by the aggregation of a number of heterogeneous autoregressive information arrival processes. The aggregation result is given in Robinson (1978b) and Granger (1980) and is restated in precisely the framework adopted here by Andersen and Bollerslev (1997a). The foreign exchange market considered here is sufficiently integrated and global to assume that the information process driving the volatility of one exchange rate is essentially the same as the information process driving the volatility of another. Hence the two volatilities are driven by a unique long memory component. They are therefore cointegrated in the sense defined by Robinson (1994c), namely that a linear combination of the two series has a strictly lesser degree of long memory than either of the individual series. This analysis is made more precise below, and a methodology is given for the analysis of stationary cointegration between volatility series.

Call X_t the vector of log squared returns for several exchanges and suppose it is covariance stationary with absolutely continuous spectral distribution function satisfying the local specification

$$f_X(\lambda) \sim \Lambda G \Lambda \quad \text{as} \quad \lambda \to 0^+ \tag{5.7}$$

where G is a real symmetric $N \times N$ matrix, $\Lambda = \operatorname{diag}\left\{\lambda^{-d_i}\right\}_{i=1}^{N}$ with $0 < d_i < \frac{1}{2}$ for $i = 1, \dots, N$ and $A \sim B$ indicates that the ratios of the corresponding elements of A and B (with identical dimensions) tend to one. Thus the representation is semiparametric in the sense that the spectral density matrix of the squares process is specified only on a degenerating band of frequencies.

The existence of a linear long run relationship between the squared returns entails the existence of a linear combination of X_t with a lesser degree of temporal dependence than the original variates. Namely, there exists a vector β such that βX_t is

I(d') with $0 \leq d' < \min_{1\leq i\leq N} d_i$. In the current framework, this fractional cointegration relationship can be defined by the fact that one (or several) linear combination of X_t has a spectral density with small order of magnitude with respect to the original:

$$F_{\beta' X}(\lambda) = O\left(\lambda^{-2d'}\right) \quad \text{when} \quad \lambda \to 0 \quad \text{with} \quad 0 \leq d' < \min_{1\leq i\leq N} d_i. \tag{5.8}$$

As $F_{\beta' X}(\lambda) = \beta' F_X(\lambda)\beta$, this is realized if $\beta' G \beta = 0$. So the long run components span the same space as the columns of a matrix $\beta_\perp$ such that $G = \beta_\perp \beta'_\perp$. The existence of long run relationships in the squares therefore implies that G is reduced rank. More precisely, the number of long run relationships[12] in the system is $N -$ rank(G) and the number of long memory conditionally heteroskedastic factors in a factor model representation is equal to rank(G).

In a similar analysis to the traditional cointegrating rank eigenvalue test discussed in Johansen (1996), the choice of K is based on the eigenvalues of a spectral estimate for the vector process X_t. Indeed, under specification 5.7 there are exactly K stationary cointegrating relationships if and only if there are exactly $N - K$ of $F_X(\lambda)$'s eigenvalues with small order of magnitude with respect to $\lambda^{-2\min_{1\leq i\leq N} d_i}$ as $\lambda \to 0^+$.

Having determined the number of cointegrating relationships, K stationary cointegrating relationships between the squared returns can be estimated. The matrix of $N - K$ linearly independent normalised stationary cointegrating vectors is estimated with a semiparametric methodology developed by Marinucci and Robinson (1996) and based on set of frequency domain regressions which do not suffer from simultaneous equations bias as time domain ordinary least squares based methods[13] do.

The analysis can be greatly simplified when the degrees of fractional integration in the squared processes are found to be equal as tends to be the case for the

[12]It is also the number of stationary cointegrating vectors.

[13]OLS and state of the art methods described in Johansen (1996) and the references therein are all directed to the I(1) versus I(0) paradigm and are inconsistent in the stationary case.

volatilities examined here. A test for this restriction with standard χ^2 limiting distribution was proposed by Robinson (1995b). It is indispensable to perform this test in case $N = 2$, as fractional cointegration may only arise between two variables with identical degrees of fractional integration.

The procedure advocated is therefore the following (All the methods mentioned are discussed below):

- Estimate semiparametrically the joint degree of fractional integration in the squared process X_t using the local Whittle, log periodogram and averaged periodogram estimates of long memory.
- Test for equality of the degrees of fractional integration in the squared returns process using Robinson (1995b)'s Fischer type statistic with χ^2 limiting distribution. Call d the joint degree of fractional integration.
- Compute the eigenvalues of a spectral estimate for the density matrix of the squares $F_X(\lambda)$ and test for stationary cointegrating rank with a likelihood ratio procedure discussed below. Call K the resulting stationary cointegrating rank, which is also the number of factors to be included in the model.
- Estimate the $N - K$ relevant stationary cointegrating relationships.

These model selection and pre-estimation procedures and the subsequent fully parametric estimation based on a frequency domain approximation of the likelihood are presented below.

First we consider estimation of long memory and testing equality of the degrees of long memory for two series. Let the discrete Fourier tranform of X_{tj} be defined as

$$w_{X_j}(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_{tj} e^{it\lambda} \tag{5.9}$$

A typical element of the multivariate periodogram $I_X(\lambda)$ is

$$I_{X_{jk}} = w_{X_j} w^*_{X_k}$$

where A^* denotes the complex conjugate of A. Testing equality of long memory degrees estimated with the LW, the LP or the AP semiparametric estimates is important for two reasons. First, one might want to simplify the analysis by testing for the equality of the d_i. Moreover, one would need to perform this test prior to identifying a potential fractional cointegration relationship between two variables, in which case it is *sine qua non*. Let $d = (d_i)_{i=1}^N$ the vector of degrees of fractional differencing for the different components of X_t and let $\hat{d}$ be its estimate. Robinson (1995b) proposes a test statistic with χ^2 limiting distribution for a homogeneous restriction

$$H_0 : \; Pd = 0,$$

where P is $H \times N$ with rank $H < N$. It is derived from the log-periodogram regression equation 1.78 for $I_{X_{ii}}(\lambda)$ and is equal to

$$\hat{d}' P' \left[(0, P) \left\{ (Z'Z)^{-1} \otimes \hat{\Omega} \right\} (0, P)' \right]^{-1} P\hat{d}$$

where Z is the matrix of regressors and $\hat{\Omega}$ is the matrix of sample variances and covariances based on the residuals. The next step, when there is more than two series under investigation, is to test for fractional cointegrating rank, or for the number of long run relationships between the series. We consider the following estimate for G:

$$\hat{G}_m = \frac{1}{m} \sum_{j=1}^{m} V_j V_j' \tag{5.10}$$

where

$$V_j' = \{v_1(\lambda_j), \ldots, v_N(\lambda_j)\} \tag{5.11}$$

and

$$v_l(\lambda) = \sqrt{1 - 2d_l} \lambda^{d_l} w_l(\lambda). \tag{5.12}$$

From Theorem 2 of Robinson (1995b), $E(\hat{G}_m) = G + O(\frac{\log m}{m})$ and therefore, $\hat{G}_m$ is asymptotically unbiased. From Theorem 1 of Robinson (1994c), for diagonal elements of G, denoted G_{ll}, we have

$$\left(\frac{2\pi}{n} \sum_{j=1}^{m} I_{X_{ll}}(\lambda_j) \right)^{-1} \left(\frac{\lambda_m^{1-2d_l}}{1 - 2d_l} G_{ll} \right) \to_p 1. \tag{5.13}$$

Proving the same for non diagonal elements G_{gh} of G would yield a consistent estimate of G for $\frac{1}{m}+\frac{m}{n}\to 0$ when $n\to\infty$. Using the asympototic uncorrelatedness of the $v_l(\lambda_j)$ over j and l from Theorem 2 of Robinson (1995b), the problem of finding the limiting distribution of eigenvalues of $\hat{G}_m$ in order to test their significance -thereby testing for cointegrating rank- could be reduced to finding the limiting distribution of

$$\frac{1}{m}\sum_{j=1}^{m}(1-2d_l)\, I_{X_{ll}}(\lambda_j)\lambda_j^{2d_l} \quad \text{for} \quad l=1,\ldots,N. \tag{5.14}$$

Once we have identified the number $N-K$ of stationary cointegrating relationshipswithin the vector of squared returns X_t, the cointegrating vectors can be identified as parameterising linear regressions of one component of X_t, denotes ζ_t, against K others, denoted Z_t, as follows:

$$\zeta_t = Z_t\beta + e_t. \tag{5.15}$$

A direct OLS or GLS estimation would prove inconsistent due to simultaneous equations bias, regardless of the short range specification of the residuals. Marinucci and Robinson (1996) show that a frequency domain analogue on a degenerating band of frequencies is consistent. The shift into frequency domain is perceptible in the following simple transformation. Multiplying each side of 5.15 by $e^{i\lambda_j t}$ and summing over the observations yields the discrete Fourier transform regression

$$w_\zeta(\lambda_j) = W_Z(\lambda_j)\beta + w_e(\lambda_j)$$

where $W_Z = (w_{Z_{jk}})_{j,k=1,\ldots,K}$, $\lambda_j = \frac{2\pi j}{n}$ and w is defined in 5.9, and the latter regression taken over a degenerated band of frequencies yields the estimate

$$\hat{\beta} = \left\{\mathcal{R}e\left(\sum_{j=1}^{m} W_Z(\lambda_j)W_Z(\lambda_j)^*\right)\right\}^{-1} \mathcal{R}e\left(\sum_{j=1}^{m} W_Z(\lambda_j)w_\zeta(\lambda_j)^*\right)$$

which is consistent under 1.69.[14] The convergence rate for each component β_j of β is $(m/n)^{d_j-d'}$ where d' is the degree of fractional integration in the regression residuals. It is conjectured by the authors, moreover, that $(n/m)^{d_j-d'}(\hat{\beta}_j-\beta_j)$ will jointly converge to a normal distribution (when the spectral density of Z is square integrable).

[14]When $m=n-1$, $\hat{\beta}$ is the OLS estimate for the regression.

Table 5.3: Test for long memory on returns

Tests of the significance of the Fractional Differencing Parameter for Raw Exchange Rate Returns

Rate	d=-0.5	d=0	d=0.5
DEM/USD	113.48	-5.02	-18.51
JPY/USD	112.55	-4.68	-19.71
JPY/DEM	122.09	0.46	-18.02

Note: Testing of the value of the fractional differencing parameter is carried out via the test procedure developed in Robinson (1996).

5.4 Results

Our results are presented in two stages. In the first sub-section we document the results of pre-testing for long range dependence, using the procedure developed in Robinson (1994a). We then go on to present the estimations of the fractional differencing parameter, for both returns and volatility, using the LW and the LP semiparametric estimates and the fully parametric AR(1)-LMSV model. Finally, a set of specification tests of the AR(1)-LMSV model is presented.

5.4.1 Testing for Long Range Dependence

In table 5.3 we present the test results for raw exchange rate returns for the DEM/USD, JPY/USD and JPY/DEM. As indicated in Section 3 the test statistic proposed by Robinson (1994a) ($\hat{S}$) has a limiting standard normal distribution under the specified null hypothesis, implying a one-sided rejection region of 2.32 at 1%. Our null hypotheses are formulated as follows. First, standard efficient markets theory indicates that asset prices should follow a random walk, implying that returns should be I(0). This defines one hypothesis as $H_0 : d = 0$. Second, we employ the theoretical bounds for stationarity and invertibility of the fractionally integrated representation for returns as hypotheses, yielding $H_0 : d = -\frac{1}{2}$ and $H_0 : d = \frac{1}{2}$.

Results yield the following observations. For all three currencies one can strongly

Table 5.4: Test of long memory on volatility

Tests of the significance of the Fractional Differencing Parameter for Exchange Rate Volatility ($\log(r^2)$).

Rate	d=0	d=0.5	d=1
DEM/USD	33.23	-13.15	-20.53
JPY/USD	21.12	-14.05	-20.70
JPY/DEM	31.06	-12.40	-20.38

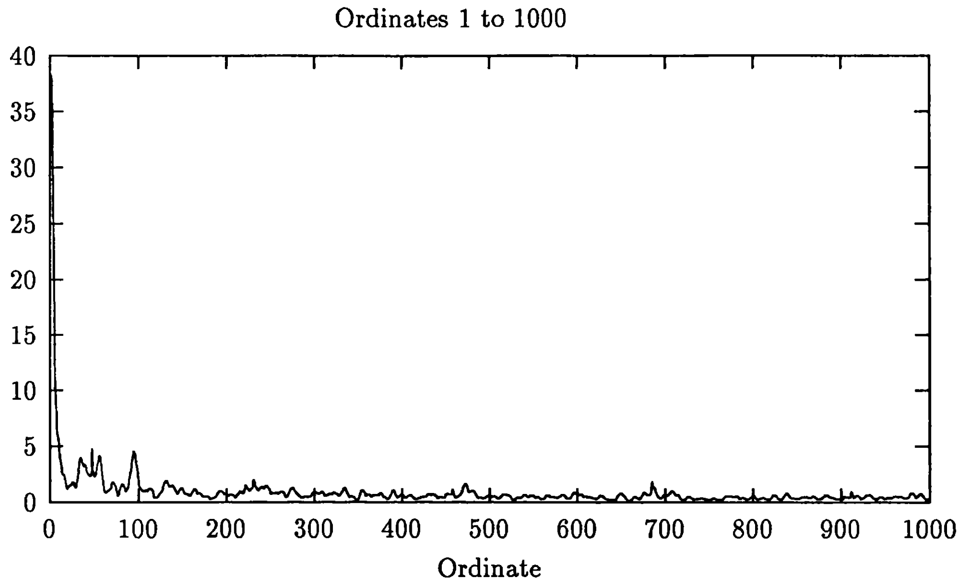
Note: Testing of the value of the fractional differencing parameter is carried out via the test procedure developed in Robinson (1996).

reject the hypothesis that $d = 0.5$ in favour of $d < 0.5$ implying that returns are covariance-stationary. Similarly one can conclude that the return processes are invertible given the sign and magnitude of the test statistics corresponding to the hypothesis that $d = -0.5$. A more interesting result appears when examining the second column of table 5.3. Whereas for the JPY/DEM the test statistic indicates that one cannot reject the hypothesis of FX *quotations* following an I(1) process, for the DEM/USD and JPY/USD there is evidence that the degree of fractional integration in returns is negative. This then implies that the degree of integration for the *quotation* series of these two currencies is between one half and unity, such that these rates are non-stationary but not I(1).

Tables 5.4 and 5.5 present the results from the same testing framework on raw and deseasonalised FX volatility (computed, as indicated in Sections 3 and 4 as $\log(r_t^2)$). As an indication of the efficacy of our deseasonalisation procedure, in figure 5.4 we present the periodogram of our deseasonalised volatility. Comparison with Figure 1 demonstrates that the amplitudes at the seasonal frequencies are greatly reduced, although not completely eliminated.[15]

[15] As an alternative to our deseasonalisation procedure, we also computed all estimations for the volatility of a time-scale transformed series of midquotes. We used the theta-time scale proposed by Dacorogna, Müller, Nagler, Olsen, and Pictet (1993). Results from these estimations were very similar to those for deseasonalised volatility and are available upon request from the authors.

Figure 5.4: Periodogram for Deseasonalised JPY/USD Log Squared Returns



The hypotheses of interest in our examination of volatility are as follows. First, given the many previous studies which have demonstrated that intra-day FX volatility has an IGARCH or almost-integrated SV representation, is there a random walk in volatility? Second, can one characterise volatility as being long range dependent? The former dictates examination of $H_0 : d = 1$ whilst the hypotheses pertinent to the latter are $H_0 : d = 0$ and $H_0 : d = \frac{1}{2}$.

The final column of tables 5.4 and 5.5 presents the evidence pertinent to the hypothesis of a random walk in volatility. There is strong evidence that, for all currencies, this hypothesis can be strongly refuted in favour of a degree of integration in volatility of less than unity. Further, this conclusion is stable across both raw and deseasonalised volatility. There is still, however, the possibility of non-stationarity in volatility if $d \in [0.5, 1]$. Column 2 of the tables indicates that the non-stationarity hypothesis can be rejected also, with the test statistics indicating that $d < 0.5$ for all three currencies. Finally, evidence on the long range dependence of volatility is shown in column 1. From both tables one can draw the conclusion that the true value of d lies between zero and one half, evidence of long memory in the volatility of all currencies.

Table 5.5: Test of long memory on deseasonalized volatility

Tests of the significance of the Fractional Differencing Parameter for Deseasonalised Exchange Rate Volatility.

Rate	d=0	d=0.5	d=1
DEM/USD	22.38	-14.26	-20.62
JPY/USD	18.54	-14.63	-20.28
JPY/DEM	24.26	-13.65	-20.55

Note: Testing of the value of the fractional differencing parameter is carried out via the test procedure developed in Robinson (1996).

Hence, the testing procedure indicates the following. Returns can be characterised as short range dependent, covariance stationary processes, with some indication of negative degrees of fractional integration for the DEM/USD and JPY/USD. The volatility results indicate covariance-stationarity also, with the non-stationarity and I(1) hypotheses convincingly refuted, although there is consistent evidence of long range dependence. The volatility processes are therefore completely mean reverting, a result which is more comfortable than that of I(1) volatility from a theoretical point of view.

5.4.2 Semiparametric Estimations

We first present point estimates for the fractional differencing parameter for the three returns series. The local Whittle and log periodogram estimates of long memory are employed. Table 5.6 gives the results.

Examining first the LP estimates it is quite clear that the negativity of d indicated in the previous subsection is a very minor economic phenomenon, indicating that returns display very small anti-persistent tendencies. For no currency does $\hat{d}$ exceed 0.05 in absolute value. This implies that the *quotation* series may be regarded as following I(1) processes to more-or-less any degree of precision, in line with the efficient markets hypothesis.

Table 5.6: Long memory in returns

Estimation of the Fractional Differencing Parameter ($\hat{d}$) for Exchange Rate Returns.

Rate	$\hat{d}$	
	LW	LP
DEM/USD	-0.38	-0.01
	(0.02)	(0.05)
JPY/USD	-0.02	-0.03
	(0.02)	(0.05)
JPY/DEM	-0.01	-0.01
	(0.02)	(0.05)

Note: Estimation of the long memory models is carried out via the local Whittle and log periodogram procedures. Standard errors in parentheses.

This conclusion becomes less clear when one examines the LW estimates. Whilst the results for the JPY/USD and JPY/DEM are very similar to their LP counterparts, the value of $\hat{d}$ derived for the DEM/USD is now greatly negative. This implies a covariance-stationary and invertible representation for DEM/USD returns which displays non-negligible anti-persistence. Given the confluence between the testing and LP results, however, we are inclined to treat this feature as an anomaly and describe the return generating process as approximately I(0)

The estimation results for the volatility series are presented (for raw and deseasonalised volatility) in tables 5.7 and 5.8. Here we complement the semiparametric procedures used in the analysis of returns with the fully parametric AR(1)-LMSV model.

The results of both tables demonstrate that the testing procedures contained in the previous subsection deliver the correct inferences. Across currencies and estimators there is consistent evidence that the value of $\hat{d}$ for the volatility series is between 0.2 and 0.3. This indicates that volatility can be characterised as covariance-stationary, invertible and long range dependent. The only real difference in estimation results

Table 5.7: Long memory in volatility

Estimation of the Fractional Differencing Parameter ($\hat{d}$) for Exchange Rate Volatility.

Rate	$\hat{d}$		
	LW	LP	LMSV
DEM/USD	0.29	0.21	0.37
	(0.02)	(0.05)	(0.01)
JPY/USD	0.27	0.19	0.24
	(0.02)	(0.05)	(0.01)
JPY/DEM	0.27	0.28	0.32
	(0.02)	(0.05)	(0.01)

Note: Estimation of the long memory models is carried out via the local Whittle and log periodogram procedures, plus the fully parametric Long Memory in Stochastic Volatility model of Harvey (1993) (LMSV). Standard errors in parentheses.

for raw and deseasonalised volatility is that the AR(1)-LMSV estimates tend to be slightly greater for the former and greater than the results delivered by the semiparametric estimators. This is likely to be due to mis-specfication of the short-range dependence in the series i.e. omission of an explicit seasonal in the fully parametric model. As one might expect, the difference in estimated $\hat{d}$ between the semiparametric and fully parametric procedures is far smaller for deseasonalised volatility.

5.4.3 Specification Tests on the Fully Parametric Model

The next step in our empirical methodology involves a series of estimations and specification tests on the fully parametric model outlined in Section 2.2. These tests allow us to examine the relative contributions of short memory and long memory components to the temporal dependence in the volatility process. As indicated in Section 2.2 our fully parametric model nests a pure LMSV model (obtained by setting $\phi = 0$ in equation 5.3) and a standard AR(1)-SV model (obtained by restricting $d = 0$ in 5.3.) By estimating the unrestricted model and these two restricted

Table 5.8: Long memory in deseasonalized volatility

Estimation of the Fractional Differencing Parameter ($\hat{d}$) for Deseasonalised Exchange Rate Volatility.

	$\hat{d}$		
Rate	LW	LP	LMSV
DEM/USD	0.31	0.32	0.26
	(0.02)	(0.05)	(0.01)
JPY/USD	0.32	0.30	0.22
	(0.02)	(0.05)	(0.01)
JPY/DEM	0.30	0.30	0.26
	(0.02)	(0.05)	(0.01)

Note: Estimation of the long memory models is carried out via the local Whittle and log periodogram procedures, plus the fully parametric Long Memory in Stochastic Volatility model of Harvey (1993) (LMSV). Deseasonalisation is carried out via a frequency domain Double-Window Smoother (DWin.). Standard errors in parentheses.

alternatives we can employ a frequency-domain Likelihood Ratio test to gauge the significance of the long memory and AR(1) components. Results of these estimations and the associated tests, for deseasonalised volatility only, are given in table 5.9.

Examining first the results which correspond to the simple AR(1)-SV models one can note the appearance of the common result that the 'underlying' volatility process has an autoregressive parameter very close to unity in all cases. This conforms with the results of many earlier studies. Moving on to the pure LMSV models it is quite clear that the conclusions of the previous estimations still hold and further that the long memory specification gives a far better fit than does the autoregression (as evidenced by the lower minimised log likelihood.)

A comparison with the unrestricted model yields the following observations. First, the improvement in fit of the combined model over the pure LMSV model is marginal, as the comparison of log-likelihoods displays. Second, and more importantly, the

Table 5.9: Parametric testing for long memory in volatility

Extensions of various LMSV specifications and Likelihood Ratio Testing using Deseasonalised Volatility

Rate	DEM/USD			JPY/USD			JPY/DEM		
	LM-AR	LM	AR	LM-AR	LM	AR	LM-AR	LM	AR
d	0.26 (0.01)	0.25 (0.01)	-	0.22 (0.01)	0.24 (0.01)	-	0.26 (0.01)	0.24 (0.01)	-
ϕ	0.12 (0.01)	-	0.96 (0.01)	0.15 (0.01)	-	0.96 (0.01)	0.15 (0.01)	-	0.93 (0.01)
LogL	3097.8	3098.2	3143.2	17228.9	17231.0	17304.8	15645.6	15646.1	15726.4
LR	-	0.85	90.90**	-	4.10*	151.75**	-	0.88	161.50**

Note: Estimation of the long memory models is carried out via the the fully parametric Long Memory in Stochastic Volatility model of Harvey (1993) (LMSV). Columns headed LM-AR present results from the model presented in equation 5.3. Columns headed LM are estimated with the restriction that the autoregressive parameter is zero and columns headed AR are estimates from the model where the fractional differencing parameter is set to zero. The final row of the table gives Likelihood Ratio Statistics relevant to the omission of the given parameter. * denotes the test is significant at 5%, ** denotes significance at 1%. Deseasonalisation is carried out via a frequency domain Double-Window Smoother . Standard errors in parentheses.

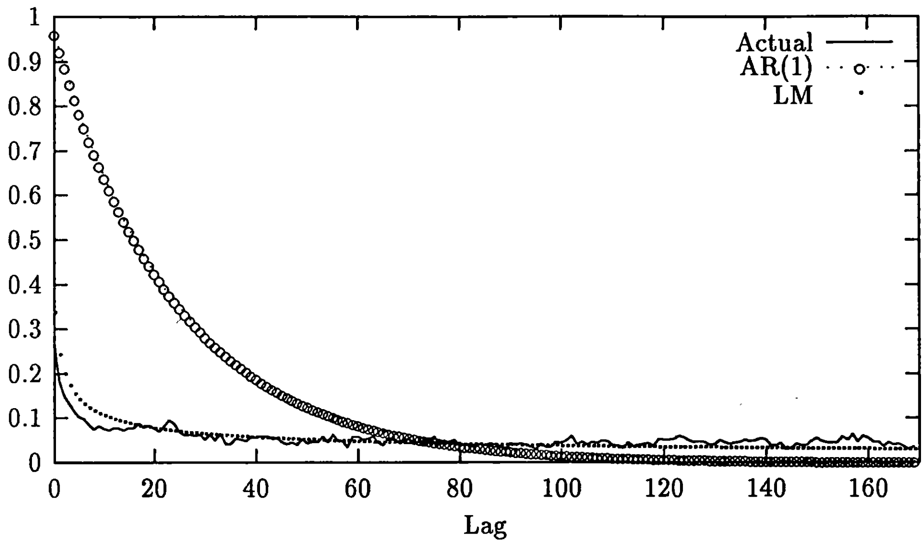
value of the autoregressive parameter is far lower in the unrestricted model, in all cases between 0.1 and 0.2. Given the pure AR(1)-SV results, this may be taken as evidence that the previous findings of near unit-root behaviour in intra-day FX volatility are caused by model mis-specification through the omission of long-range dependent components. This result confirms the evidence from daily data found in Baillie, Bollerslev, and Mikkelsen (1996). These conclusions are reinforced by the quasi-LR test statistics. In the unrestricted model one can convincingly reject the hypothesis that the degree of fractional integration in volatility is zero for all currencies, whereas the hypothesis that the autoregressive parameter is zero cannot be rejected in two of the three cases and is only marginally rejected in the third.

A clear comparison of the AR-SV and LM-SV models, in terms of how well they fit the data, is shown in Figure 5. The figure graphs the actual correlogram of the volatility process for the DEM/USD alongside those implied by the estimated AR-SV and LM-SV models. It is immediately apparent that the long memory specification gives a far better approximation of the true volatility dynamics than does the autoregressive model. The autoregressive model greatly overstates the low order autocorrelations but dies out too quickly to mimic the persistently positive high-order autocorrelations in the DEM/USD data. Intuitively, the estimated autoregressive parameter is driven very close to unity in order to try to approximate the long memory in volatility, but this only results in the low order autocorrelations being far too high whilst the exponential decay in the correlogram ensures that, even if the estimated autoregressive parameter is arbitrarily close to unity, the autoregressive specification cannot match the persistence in volatility exhibited by the data.

5.4.4 Fractional cointegration

In the following section, the DEM/JPY (Deutsch Mark/Japanese Yen) is replaced by another exchange against the US Dollar: GBP/USD (British Pound/US Dollar)

Figure 5.5: Comparison of Actual and Implied Correlograms for DEM/USD



to avoid spurious cointegration due to the circularity effect. We report in table 5.10 the results $\hat{d}_i$ from the local Whittle (LW), the log periodogram (LP) and the averaged periodogram (AP) with automatic bandwidth selection procedures described in chapter 4, and we also report the selected bandwidths m.

The Wald statistic for the test of equality of these long memory parameters is $\hat{W} = 0.184$ which is far lower than the 90% level quantile of a χ^2 distribution with two degrees of freedom. We therefore fail to reject the hypothesis of equality of these long memory parameters.

Let X_t be our vectors of exchange rates against the dollar. We form a spectral estimate of the matrix G as indicated in 5.10 with $m = 857$ and compute its eigenvalues:

$$\begin{aligned} \alpha_1 &= 2.5 \\ \alpha_2 &= 1.1 \\ \alpha_3 &= 0.5 \end{aligned}$$

which does not seem conclusive. If anything, there is evidence of G being rank two. This evidence is reinforced by aberrant regression results when G is assumed to be rank two.

Table 5.10: Automatic estimation of long memory in deseasonalised volatility
Estimation of the Fractional Differencing Parameter ($\hat{d}$) for Exchange Rate Volatilities.

Rate	$\hat{d}$			$\hat{m}$		
	LW	LP	AP	LW	LP	AP
DEM/USD	0.31 (0.02)	0.32 (0.05)	0.35	626	607	551
JPY/USD	0.32 (0.02)	0.30 (0.05)	0.32	657	654	645
GBP/USD	0.35 (0.02)	0.35 (0.05)	0.38	633	619	511

Note: Estimation of the long memory parameters carried out via the local Whittle, log periodogram and averaged periodogram procedures.

In that case, there is exactly one long run relationship between the exchange rates volatilities. In order to identify it, we run the frequency domain analogue of the regression

$$X_t^{DEM} = \beta_1 X_t^{JPY} + \beta_2 X_t^{GBP} + U_t$$

which yields the estimated regression

$$\begin{aligned} \hat{X}_t^{DEM} &= .346 X_t^{JPY} + .555 X_t^{GBP} \\ & \quad (.16) \end{aligned}$$

where the standard errors are computed under the conjecture of normality and the confidence interval on the degree of fractional integration of the residuals includes zero, so that we cannot reject the hypothesis that the residuals are I(0), otherwise interpreted as a hypothesis of full stationary cointegration.

5.5 Conclusion

Hence, to restate the main findings; returns can be characterised as I(0) processes to a great degree of precision; volatility, on the other hand, is best represented by a covariance stationary, long range dependent process; furthermore, there is evidence that the common finding of near-integrated volatility processes is driven by the mis-specification of traditional models which do not permit long range dependence. Finally, there is evidence of stationary cointegration between the volatility series, thus bringing tentative support to the Mixture of Distributions Hypothesis for speculative prices, and paving the way for parsimonious factor representations of vectors of foreign exchange rate returns with fractionally integrated volatilities. However, further research into the process of information filtration through the market would be needed, in particular in view of the evidence of a large discrepancy in persistence on volatilities between shocks related to macroeconomic "news" announcements and other types of shocks. There is indeed a possibility that long memory in the volatility may be generated spuriously by the aggregation of these two kinds of shocks. In this event, however, efficient estimation of even spurious long memory volatilities remains valuable for forecasting purposes and risk analysis.

Bibliography

ADENSTEDT, R. (1974): "On large sample estimation for the mean of a stationary random sequence," *Annals of Statistics*, 2, 1095–1107.

ANDERSEN, T., and T. BOLLERSLEV (1997a): "Heterogeneous information arrivals and return volatility dynamics: uncovering the long run in high frequency returns," *Journal of Finance*, 52, 975–1006.

ANDERSEN, T., and T. BOLLERSLEV (1997b): "Intraday seasonality and volatility persistence in financial markets," *Journal of Finance*, 52, 102–132.

BAILLIE, R., T. BOLLERSLEV, and H. MIKKELSEN (1996): "Fractionally integrated generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, 74, 3–30.

BARTLETT, M. (1950): "Periodogram analysis and continuous spectra," *Biometrika*, 37, 1–16.

BELTRÃO, K., and P. BLOOMFIELD (1987): "Determining the bandwidth of a kernel spectral estimate," *Journal of Time Series Analysis*, 8, 21–38.

BERAN, J. (1986): "Estimation, testing and prediction for self-similar and related processes," Ph.D. Thesis, ETH, Zürich.

BERAN, J. (1993): "Fitting long memory models by generalised linear regression," *Biometrika*, 80, 817–822.

BERAN, J. (1994): *Statistics for long memory processes*, Monographs on Satistics and Applied Probability. Chapman and Hall.

BLOOMFIELD, P. (1973): "An exponential model for the spectrum of a scalar time series," *Biometrika*, 60, 217–226.

BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, 31, 307–327.

BOLLERSLEV, T., R. CHOU, and K. KRONER (1992): "ARCH modeling in finance," *Journal of Econometrics*, 52, 5–59.

BOLLERSLEV, T., and I. DOMOWITZ (1993): "Trading patterns and prices in the interbank foreign exchange market," *Journal of Finance*, 48, 1421–1444.

BOLLERSLEV, T., and R. ENGLE (1993): "Common persistence in conditional variances," *Econometrica*, 61, 167–186.

BOLLERSLEV, T., R. ENGLE, and D. B. NELSON (1994): "ARCH models," in Handbook of Econometrics, Vol. IV, ed. by R. Engle, and D. McFadden. Elsevier, ch. 49, pp. 2959-3038.

BRILLINGER, D. R. (1975): *Time series, Data Analysis and Theory.* Holden-Day: San Francisco.

BROCKWELL, P. J., and R. A. DAVIS (1991): *Time Series: Theory and Methods.* Springer-Verlag: New York.

BROWN, B. (1971): "Martingale central limit theorems," *The Annals of Mathematical Statistics*, 49(1), 59–66.

BURKHOLDER, D. (1973): "Distribution function inequalities for martingales," *Annals of Probability*, 1, 19–42.

CHEUNG, Y. W., and K. S. LAI (1993): "A fractional cointegration analysis of purchasing power parity.," *Journal of the American Statistical Association*, 11, 103–112.

CLARK, P. K. (1973): "A subordinated stochastic process model with fixed variances for speculative prices," *Econometrica*, 41, 135–156.

COMTE, F., and C. HARDOUIN (1995): "Regression on log-regularized periodogram for fractional models at low frequencies," Discussion Paper, CREST.

DACOROGNA, M., U. MÜLLER, R. NAGLER, R. OLSEN, and O. PICTET (1993): "A geographical model for the daily and weekly seasonal volatility in the foreign exchange market," *Journal of International Money and Finance*, 12, 413–438.

DAHLHAUS, R. (1989): "Efficient parameter estimation for self-similar processes," *Annals of Statistics*, 17, 1749–1766.

DANIELSSON, J., L. DE HAAN, L. PENG, and C. G. DE VRIES (1997): "Using a bootstrap method to choose the sample fraction in tail index estimation," preprint.

DAVIES, R., and D. HARTE (1987): "Tests for the Hurst Effect," *Biometrika*, 74, 95–101.

DEGENNARO, R., and R. SHRIEVES (1995): "Public Information Releases, Private Information Arrival, and Volatility in the Foreign Exchange Market," Preprint.

DELGADO, M., and P. M. ROBINSON (1992): "Nonparametric and semiparametric methods for economic research," *Journal of Economic Surveys*, 6, 201–249.

DELGADO, M., and P. M. ROBINSON (1994): "New methods for the analysis of long memory: application to Spanish inflation," *Journal of Forecasting*, 13, 97–107.

DELGADO, M., and P. M. ROBINSON (1996): "Optimal spectral bandwidth for long memory," *Statistica Sinica*, 67, 97–112.

DEO, R., and C. HURVICH (1998): "Trend with long memory," Discussion Paper, New York University.

DIEBOLD, F., and M. NERLOVE (1989): "Unit roots in economic time series: a selective survey," *Advances in Econometrics*, 8, 3–69.

DING, Z., and C. W. J. GRANGER (1996): "Modelling volatility persistence of speculative returns: a new approach," *Journal of Econometrics*, 23, 185–215.

DING, Z., C. W. J. GRANGER, and R. ENGLE (1993): "A long memory property of stock market returns and a new model," *Journal of Empirical Finance*, 1, 83–106.

DOUKHAN, P. (1995): *Mixing properties and examples.* Springer Verlag.

DROST, F. C., and T. E. NIJMAN (1993): "Temporal aggregation of GARCH processes," *Econometrica*, 61, 909–927.

EICKER, F. (1956): "Limit theorems for regression with unequal and dependent errors," in *Fifth Berkeley Symposium in Mathematical Statistics and Probability,* Volume I, ed. by J. Neyman. Berkeley: University of California Press, pp. 59-82.

ELLIOT, G., J. STOCK, and T. ROTHENBERG (1994): "Efficient tests of an autoregressive unit root," preprint.

ENGLE, R. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1007.

EPPS, T. W., and M. L. EPPS (1976): "The stochastic dependence of security price changes and transaction volumes: implications for the mixture of distributions hypothesis," *Econometrica*, 44, 305–321.

FAMA, E. (1965): "The behaviour of stock market prices," *Journal of Business*, 38, 34–105.

FAMA, E. (1991): "Efficient Capital Markets: II," *Journal of Finance*, 46, 1573–1617.

FOX, R., and M. S. TAQQU (1986): "Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series," *Annals of Statistics*, 14, 517–132.

FRANCQ, C., and J.-M. ZAKOÏAN (1997): "Estimating weak GARCH processes," Document de Travail numéro 9740, CREST.

GEWEKE, J., and S. PORTER-HUDAK (1983): "The estimation and application of long memory time series models," *Journal of Time Series Analysis*, 4, 221–238.

GIL-ALAÑA, L., and P. M. ROBINSON (1997): "Testing of unit root and other non-stationary hypotheses in macroeconomic time series," *Journal of Econometrics*, 80, 241–268.

GIRAITIS, L., P. M. ROBINSON, and A. SAMAROV (1997): "Rate optimal semi-parametric estimation of the memory parameter of the Gaussian time series with long range dependence," *Journal of Time Series Analysis*, 18, 49–60.

GIRAITIS, L., and D. SURGAILIS (1986): "Multivariate Appell polynomials and the central limit theorem," in *Dependence in Probability and Statistics*, ed. by E. Eberlein, and M. S. Taqqu, pp. 21–71. Birkhäuser, Boston.

GIRAITIS, L., and D. SURGAILIS (1990): "A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate," *Probability Theory and Related Fields*, 86, 87–104.

GLOSTEN, M., and P. MILGROM (1985): "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders," *Journal of Financial Economics*, 14, 71–100.

GOODHART, C. A. E., and M. O'HARA (1997): "High frequency data in financial markets: issues and applications," *Journal of Empirical Finance*, 1, 83–106.

GOURIÉROUX, C., and I. PEAUCELLE (1991): "Séries codépendentes. Application à l'Hypothèse de Parité du Pouvoir d'Achat," Document de Travail, CREST.

GRANGER, C. W. J. (1966): "The typical spectral shape of an economic variable," *Econometrica*, 34, 150–161.

GRANGER, C. W. J. (1980): "Long memory relationships and the aggregation of dynamic models," *Journal of Econometrics*, 14, 227–238.

GRANGER, C. W. J. (1987): "Two papers: Generalized Integrated Processes and Generalized Cointegration,," Working Paper 87-20, University of California San Diego.

GRANGER, C. W. J., and R. JOYEUX (1980): "An introduction to long memory time series models," *Journal of Time Series Analysis*, 1, 15–29.

GREENE, M., and B. FIELITZ (1977): "Long term dependence in common stock returns," *Journal of Financial Economics*, 4, 339–349.

GRENANDER, U. (1953): "Statistical analysis of time series arising from stationary stochastic processes," *The Annals of Mathematical Statistics*, 24, 519–521.

GRENANDER, U., and M. ROSENBLATT (1966): *Statistical analysis of stationary time series.* Wiley: New York.

GUILLAUME, D. (1995): "On the intra-day performance of GARCH processes," mimeo., Financial Markets Group, London School of Economics.

HANNAN, E. J. (1970): *Multiple time series.* Wiley: New York.

HANNAN, E. J. (1973): "The asymptotic theory of linear time series models," *Journal of Applied Probability*, 10, 130–145.

HANNAN, E. J. (1979): "Non linear time series regression," *Journal of Applied Probability*, 8, 767–780.

HÄRDLE, W. (1990): *Applied nonparametric regression.* Cambridge University Press: Cambridge.

HARVEY, A. (1993): "Long Memory in Stochastic volatility," mimeo, London School of Economics.

HASBROUCK, J. (1991): "Measuring the information content of stock trades," *Journal of Finance*, 46, 179–206.

HE, C. (1997): "The fourth moment structure of the GARCH process," mimeo, Stockholm School of Economics.

HENRY, M., and R. PAYNE (1997): "An investigation of long range dependence in intra-day foreign exchange volatility," Discussion Paper 264, Financial Markets Group, London School of Economics.

HENRY, M., and P. M. ROBINSON (1996): "Bandwidth choice in Gaussian semiparametric estimation of long range dependence," in *Athens Conference on Applied Probability and Time Series Analysis,* Volume II: Time Series Analysis, In memory of E.J. Hannan, ed. by P. M. Robinson, and M. Rosenblatt. Springer Verlag, New York, pp. 220-232.

HEYDE, C. C., and R. GAY (1993): "Smoothed periodogram asymptotics and estimation for processes and fields with possible long range dependence," *Stochastic processes and their applications*, 45, 169–182.

HILL, B. (1975): "A simple approach to inference about the tail of a distribution," *Annals of Statistics*, 3, 1163–1174.

HOSKING, J. (1981): "Fractional differencing," *Biometrika*, 68, 165–176.

HURST, H. E. (1951): "Long term storage capacity of reservoirs," *Trans. Am. Soc. Civil Engineers*, 116, 770–799.

HURVICH, C. (1985): "Data-driven choice of a spectrum estimate: extending the applicability of cross-validation methods," *Journal of the American Statistical Association*, 80, 933–940.

HURVICH, C., and K. BELTRAO (1993): "Asymptotics for the low frequency ordinates of the periodogram of a long memory time series," *Journal of Time Series Analysis*, 14, 455–472.

HURVICH, C., and K. BELTRAO (1994): "Automatic semiparametric estimation of the memory parameter of a long memory time series," *Journal of Time Series Analysis*, 15, 285–302.

HURVICH, C., R. DEO, and J. BRODSKY (1998): "The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series," *Journal of Time Series Analysis*, 19, 19–46.

IBRAGIMOV, I. A. (1959): "Some limit theorems for stochastic processes stationary in the strict sense," *Doklady Academia Nauk, USSR*, 125, 711–14.

IBRAGIMOV, I. A. (1962): "Some limit theorems for stationary processes," *Probability Theory and its Applications*, 7, 349–382.

IBRAGIMOV, I. A. (1963): "On estimation of the spectral function of a stationary Gaussian process," *Theory of Probability and Applications*, 8, 366–401.

IBRAGIMOV, I. A., and Y. V. LINNIK (1971): *Independent and stationary sequences of random variables.* Wolters-Noordhoff: Gröningen.

JANACEK, G. J. (1993): "A spectral approach to long memory time series," in *Dependence in Time Series Analysis, In Memory of Maurice B. Priestley*, ed. by T. Suba Rao. Chapman and Hall, pp. 164-179.

JOHANSEN, S. (1996): *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford University Press.

KOLMOGOROV, A. N. (1940): "Wienersche Spiralen und einige andere interessante Kurven in Hilbertschen Raum," *Doklady Academia Nauk, USSR*, 26, 115–118.

KOOPERBERG, C., C. STONE, and Y. TRUONG (1995a): "Logspline estimation of a possibly mixed spectral distribution," *Journal of Time Series Analysis*, 16, 359–388.

KOOPERBERG, C., C. STONE, and Y. TRUONG (1995b): "Rate of convergence for Logspline spectral density estimation," *Journal of Time Series Analysis*, 16, 389–401.

KUERSTEINER, G. (1997): "Efficient IV estimation for autoregressive models with conditional heterogeneity," mimeo, Yale University.

KÜNSCH, H. (1987): "Statistical aspects of self-similar processes," in *Proceedings of the First World Congress of the Bernouilli Society*, pp. 67–74. VNU Science Press.

KÜNSCH, H. (1996): "Discriminating between monotonic trends and long range dependence," *Journal of Applied Probability*, 23, 1025–1030.

LEE, D., and P. M. ROBINSON (1996): "Semiparametric exploration of long memory in stock prices," *Journal of Statistical Planning and Inference*, 50, 155–174.

LEE, S.-W., and B. E. HANSEN (1994): "Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator.," *Econometric Theory*, 10, 29–52.

LO, A. (1991): "Long term memory in stock market prices.," *Econometrica*, 59, 1279–1313.

LOBATO, I., and P. M. ROBINSON (1996): "Averaged periodogram estimation of long memory," *Journal of Econometrics*, 73, 303–324.

LOBATO, I., and P. M. ROBINSON (1998): "A nonparametric test for I(0)," *Review of Economic Studies*, forthcoming.

LOBATO, I., and N. E. SAVIN (1998): "Real and spurious long memory in stock market data," mimeo Iowa State University.

LUMSDAINE, R. (1996): "Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models.," *Econometrica*, 64, 575–596.

LYONS, R. (1985): "Tests of microstructural hypotheses in the foreign exchange market," *Journal of Financial Economics*, 14, 71–100.

MANDELBROT, B. B. (1963): "The variation of certain speculative prices," *Journal of Business*, 36, 394–419.

MANDELBROT, B. B. (1972): "Statistical methodology for non periodic cycles: from the covariance to the R/S analysis," *Annals of Economic and Social Measurement*, 1, 259–290.

MANDELBROT, B. B. (1973): "le problème de la réalité des cycles lents et le 'syndrome de Joseph'," *Revue d'économie appliquée*, 26, 349–365.

MANDELBROT, B. B., and J. W. V. NESS (1968): "Fractional Brownian motions, fractional noise and applications," *SIAM Review*, 10, 422–437.

MANN, H. B., and A. WALD (1943): "On the statistical treatment of linear stochastic difference equations," *Econometrica*, 11, 173–220.

MARINUCCI, D., and P. M. ROBINSON (1996): "Consistent estimation of fractional cointegration," mimeo, LSE.

NELSON, D. B. (1990a): "ARCH models as diffusion approximations," *Journal of Econometrics*, 45, 7–38.

NELSON, D. B. (1990b): "Stationarity and persistence in the GARCH(1,1) model," *Econometric Theory*, 6, 318–334.

NELSON, D. B. (1991): "Conditional heteroscedasticity in asset pricing: a new approach," *Econometrica*, 59, 347–370.

NEWEY, W. K., and D. G. STEIGERWALD (1994): "consistency of quasi maximum likelihood estimators for models with conditional heteroscedasticity," Discussion Paper, UCSD.

PAYNE, R. (1996): "Announcement effects and seasonality in the intra-day foreign exchange market," Discussion Paper 238, Financial Markets Group, London School of Economics.

PRIESTLEY, M. (1981): *Spectral analysis and time series*, Probability and Mathematical Statistics. Academic Press.

ROBINSON, P. M. (1978a): "Alternative models for stationary stochastic processes," *Stochastic processes and their applications*, 8, 141–152.

ROBINSON, P. M. (1978b): "Statistical Inference for a Random Coefficient Autoregressive Model," *Scandinavian Journal of Statistics*, 5, 163–168.

ROBINSON, P. M. (1983a): "Nonparametric estimation for time series models," *Journal of Time Series Analysis*, 4, 185–208.

ROBINSON, P. M. (1983b): "Review of various approaches to power spectrum estimation," in Handbook of Statistics, Vol. 3, ed. by D. R. Brillinger, and P. R. Krishnaiah. Elsevier, pp. 343-368.

ROBINSON, P. M. (1991a): "Automatic frequency domain inference on semiparametric and nonparametric models," *Econometrica*, 59, 1329–1363.

ROBINSON, P. M. (1991b): "Testing for strong serial correlation and dynamic conditional heteroscedasticity in multiple regression," *Journal of Econometrics*, 47, 67–84.

ROBINSON, P. M. (1994a): "Efficient tests of nonstationary hypotheses," *Journal of the American Statistical Association*, 89, 1420–1437.

ROBINSON, P. M. (1994b): "Rates of convergence and optimal bandwidth in spectral analysis of processes with long range dependence," *Probability Theory and Related Fields*, 99, 443–473.

ROBINSON, P. M. (1994c): "Semiparametric analysis of long memory time series," *Annals of Statistics*, 22, 515–539.

ROBINSON, P. M. (1994d): "Time series with strong dependence," in *Advances in Econometrics*, ed. by C. Sims, vol. 1, pp. 97–107. Cambridge University Press.

ROBINSON, P. M. (1995a): "Gaussian semiparametric estimation of long range dependence," *Annals of Statistics*, 23, 1630–1661.

ROBINSON, P. M. (1995b): "Log periodogram regression of time series with long range dependence," *Annals of Statistics*, 23, 1048–1072.

ROBINSON, P. M., and M. HENRY (1997): "Long and short memory conditional heteroscedasticity in estimating the memory parameters of levels," mimeo, London School of Economics.

ROBINSON, P. M., and C. VELASCO (1996): "Autocorrelation-robust inference.," Discussion Paper EM/96/316, STICERD, London School of Economics.

ROSENBLATT, M. (1956): "On some regression problems in time series analysis," in *Third Berkeley Symposium in Mathematics,* Volume I, ed. by J. Neyman. Berkeley: University of California Press, pp. 165-186.

ROSENBLATT, M. (1981): "Limit theorems for Fourier transforms of functionals of Gaussian sequences.," *Zeitschrift für Wahrscheinlichkeits-theorie und Verwandte Gebiete*, 55, 123–132.

SACHS, R. V. (1994): "Peak-insensitive non-parametric spectrum estimation," *Journal of Time Series Analysis*, 15, 429–452.

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.

SCHUSTER, A. (1898): "On the investigation of hidden periodicities," *Terr. Mag.*, 3, 13–14.

SENTANA, E. (1995): "Quadratic ARCH models," *Review of Economics Studies*, 62, 639–661.

SHEPHARD, N. (1996): "Statistical aspects of ARCH and Stochastic Volatility," in *Time Series Models*, ed. by D. V. H. David R. Cox, and O. E. Barndorff-Nielsen. Chapman and Hall.

SILVERMAN, B. W. (1986): *Density estimation for statistics and data analysis.* Chapman and Hall: London.

SMITH, R., and B. CHEN (1996): "Local Whittle estimation of long memory," in *Athens Conference on Applied Probability and Time Series Analysis,* Volume II: Time Series Analysis, In Memory of E.J. Hannan, ed. by P. M. Robinson, and M. Rosenblatt. Springer Verlag, New York, pp. 220-232.

SOLO, V. (1989): "Asymptotics of quadratic forms for time series," unpublished manuscript.

TAQQU, M. S. (1975): "Weak convergence of fractional Brownian motion to the Rosenblatt process," *Zeitschrift für Wahrscheinlichkeits-theorie und Verwandte Gebiete*, 40, 287–302.

TAQQU, M. S., and V. TEVEROVSKY (1995a): "Estimators for long range dependence: an empirical study," *Fractals*, 3, 785–798.

TAQQU, M. S., and V. TEVEROVSKY (1995b): "Robustness of Whittle-type estimates for time series with long range dependence," preprint.

TAQQU, M. S., and V. TEVEROVSKY (1996): "Graphical methods for bandwidth selection in semiparametric estimation of long memory," in *Athens Conference on Applied Probability and Time Series Analysis,* Volume II: Time Series Analysis, In Memory of E.J. Hannan, ed. by P. M. Robinson, and M. Rosenblatt. Springer Verlag, New York, pp. 220-232.

TAUCHEN, G., and M. PITTS (1983): "The price variability-volume relationship on speculative markets," *Econometrica*, 51, 485–505.

TAYLOR, S. (1980): "Conjectured models for trends in financial prices, tests and forecasts," *Journal of the Royal Statistical Society*, A 143, 338–362.

TOUSSOUN, O. (1925): *Mémoire sur l'Histoire du Nil*, vol. 9 of *Mémoires de l'Institut d'Egypte.* Institut d'Egypte, Le Caire.

TUKEY, J. W. (1950): "Measuring noise colour," unpublished manuscript.

ŽURBENKO, I. G. (1986): *The spectral analysis of time series.* Amsterdam: North Holland.

WEISS, A. A. (1986): "Asymptotic theory for ARCH models: Estimation and testing," *Econometric Theory*, 2, 107–131.

WHISTLER, D. E. N. (1990): "Parametric models of daily and intra-daily exchange rate volatility," Ph.D. Thesis, University of London.

WHITTLE, P. (1962): "Gaussian estimation in stationary time series," *Bulletin of the International Statistical Institute*, 39, 105–129.

WOLD, H. (1938): *A study in the analysis of stationary time series.* Almqvist and Wiksell, Uppsala.

YAJIMA, Y. (1989): "A central limit theorem of Fourier transforms of strongly dependent stationary processes.," *Journal of Time Series Analysis*, 10, 375–383.

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

ZAKOÏAN, J.-M. (1995): "Threshold heteroscedastic models," *Journal of Economic Dynamics and Control*, 18, 931–955.

ZYGMUND, A. (1977): *Trigonometric series.* Cambridge University Press.