Averaged periodogram spectral estimation with long memory conditional heteroscedasticity

Marc Henry[1]

Department of Economics and Center for Applied Probability
Columbia University in the City of New York
420 W 118th Street, New York, NY10027, USA
marc.henry@columbia.edu

1 Abstract

The empirical relevance of long memory conditional heteroscedasticity has emerged in a variety of studies of long time series of high frequency financial measurements. A reassessment of the applicability of existing semiparametric frequency domain tools for the analysis of time dependence and long run behaviour of time series is therefore warranted. To that end, this paper analyses the averaged periodogram statistic in the framework of a generalized linear process with long memory conditional heteroscedastic innovations according to a model specification first proposed by Robinson (1991). It is shown that the averaged periodogram estimate of the spectral density of a short memory process remains asymptotically normal with unchanged asymptotic variance under mild moment conditions, and that for strongly dependent processes, Robinson (1994)'s averaged periodogram estimate of long memory remains consistent.

[1]This research was carried out at the Financial Markets Group of the London School of Economics and Political Science and at the Laboratoire de Finance-Assurance du CREST, and financial support from both institutions is gratefully acknowledged. The author is grateful for considerable help, suggestions and corrections from Peter Robinson, and comments from two anonymous referees and seminar participants at Columbia University, CORE, the Cowles Foundation, CREST, ECARE, the London School of Economics, l'Université de Montréal, New York University and the Sonderforschungsbereich 373. All remaining errors are, of course, his own.

2 Introduction

The issue of temporal dependence and persistence of shocks on financial time series has recently been the focus of considerable attention. Tests of market efficiency (and, conversely, of return predictability) have been based on models aimed at disentangling short run dependencies, seasonal components and long run characteristics of the process under investigation (see Fama (1991) for a survey on the subject). If the focus of interest is, as in this paper, the long run behaviour of a covariance stationary time series, one may study the spectral density in a neighbourhood of frequency zero. To that end, one may want to avoid inconsistency in estimating even low frequency characteristics due to possible misspecification of higher frequency dynamics. This may be achieved through the use of methods which rely on a local specification of the spectral density in a neighbourhood of zero frequency only. The loss of efficiency incurred is presumably of little consequence in the case of long and reliable time series of financial measurements. However, if such semiparametric frequency domain methods are to be valid tools for inference in this field, they need to allow for various forms of conditional heteroscedasticity which is now recognized as a dominant feature of asset and foreign exchange rate returns. The statistic considered in this semiparametric framework is the averaged periodogram statistic which is used in its raw form to estimate the spectral density at zero of a short memory time series, and a transformation of which is used to estimate the degree of long memory in a covariance stationary time series whose spectrum is infinite at zero frequency. The latter estimate, known as the averaged periodogram estimate, was proposed by Robinson (1994).

The contribution of this paper is to show the robustness of these two related estimation methods to (possibly long memory) conditional heteroscedasticity in the Wold innovations of the process under investigation, using a specification first proposed by Robinson (1991) and developed in Robinson and Henry (1999). In the case of short memory time series, the averaged periodogram is the natural statistic to use for nonparametric estimation of the spectrum, and the asymptotic normality result presented is readily extendable to other kernel estimates of the spectral density. As for the estimation of long memory, two other semiparametric estimates are available in the literature: the local Whittle or Gaussian estimate of Robinson (1995a) and the log periodogram estimate of Geweke and Porter-Hudak (1983) and Robinson (1995b). However, unlike the local Whittle estimate, also considered in this long memory conditional heteroscedasticity framework by Robinson and Henry (1999), the averaged periodogram estimate is available in closed form. As for the more popular log periodogram estimate, no consistency result has yet been obtained with time dependent conditional variances.

The next section discusses the nonparametric specification chosen to model conditional heteroscedasticity in the Wold innovations of the weakly stationary times series considered. In that framework, section 4 considers traditional estimation of the spectral density and gives an asymptotic normality result for the averaged periodogram estimate of the spectral density at frequency zero. Section 5 considers Robinson's average periodogram based estimate of long memory (see Robinson (1994)) in this framework, and provides a consistency result and a small sample assessment of the effects of (possibly long memory) conditional heteroscedasticity in the innovations. Section 6 concludes, and proofs of theorems are collected in the appendix.

3 Conditional heteroscedasticity specification

A covariance stationary process is written in infinite moving average form:

$$x_t = E(x_t) + \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \quad \alpha_0 = 1, \quad \sum_{j=0}^{\infty} \alpha_j^2 < \infty, \tag{3.1}$$

and we assume that the ε_t satisfy at least

$$E(\varepsilon_t | \mathcal{F}_{t-1}) = 0 \quad \text{almost surely (a.s.).} \tag{3.2}$$

However, many limit results for generalized linear processes under 3.2, such as Hannan (1979)'s, require the assumption of constant conditional variance

$$E(\varepsilon_t^2 / \mathcal{F}_{t-1}) = \sigma^2 \quad \text{a.s.} \tag{3.3}$$

that many time series 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 3.3 needs to be replaced by

$$E(\varepsilon_t^2 / \mathcal{F}_{t-1}) = \sigma_t^2 \quad \text{a.s.} \tag{3.4}$$

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 Taylor (1980)'s "stochastic volatility" model in which 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 (1991) where σ_t^2 is an infinite sum of lagged values of ε_t^2:

$$\sigma_t^2 = \sigma^2 + \sum_{j=1}^{\infty} \psi_j(\varepsilon_{t-j}^2 - \sigma^2) \quad \text{a.s.} \quad \text{with} \quad \sum_{j=1}^{\infty} \psi_j^2 < \infty. \tag{3.5}$$

This can be reparameterised as

$$\sigma_t^2 = \beta + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j}^2 \tag{3.6}$$

and includes both standard ARCH (when $\psi_j = 0,\ j > p$, for finite p) and GARCH (for which the ψ_j decay exponentially) models. Naturally, as it relies on a specification with a single set of innovations, this specification fails to encompass the stochastic volatility paradigm, and its successfull extension to long memory volatilities by Robinson and Zaffaroni (1998). However, as Robinson (1991) 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,

$$\psi(z) = 1 - \sum_{j=1}^{\infty} \psi_j z^j \tag{3.7}$$

satisfies

$$|\psi(z)| \neq 0, \ |z| \leq 1, \tag{3.8}$$

define

$$\phi(z) = \sum_{j=0}^{\infty} \phi_j z^j = \psi(z)^{-1}, \ \phi_0 = 1. \tag{3.9}$$

Then, Robinson (1991) rewrote 3.5 as

$$\varepsilon_t^2 - \sigma^2 = \sum_{j=0}^{\infty} \phi_j \nu_{t-j}, \tag{3.10}$$

where

$$\nu_t = \varepsilon_t^2 - \sigma_t^2 \tag{3.11}$$

satisfies

$$E(\nu_t | \mathcal{F}_{t-1}) = 0 \quad \text{a.s.}, \tag{3.12}$$

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 3.10 with a suitable choice of filtering weights. To allow for specific types of nonlinearities in the squares, Robinson (1991) also proposed a quadratic version of 3.5:

$$\sigma_t^2 = \Big(\sigma + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j}\Big)^2 \quad \text{a.s.} \tag{3.13}$$

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 (1990)), 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

$$s_{t+l} = \psi_l \chi_t + \sum_{j \neq l}^{\infty} \psi_j \chi_{t-j+l} \quad \text{a.s.} \tag{3.14}$$

Now as under 3.5, $\psi_l \chi_t \to 0$ almost surely when $l \to \infty$, χ is persistent in the volatility according to none of the definitions adopted by Nelson (1990), 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 the "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 3.5 with the $\phi(z)$ polynomial defined as

$$\phi(z) = (1 - z)^{-d_\varepsilon} \frac{b(z)}{a(z)} \tag{3.15}$$

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 3.15 to asset prices with the addition of a drift parameter

$$(1 - L)^{d_\varepsilon} a(L) \varepsilon_t^2 = \mu + b(L) \nu_t. \tag{3.16}$$

Nelson (1990) proves almost sure convergence of the conditional variance σ_t^2 in the short memory case d_ε in 3.15 with $a(z)$ and $b(z)$ of degree one. Apart from 3.15, the requirement

$$0 < \sum_{j=0}^{\infty} \phi_j^2 < \infty \tag{3.17}$$

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

$$\operatorname{corr}\left(\varepsilon_t^2, \varepsilon_{t+j}^2\right) = \frac{\sum_{i=0}^{\infty} \phi_i \phi_{i+j}}{\sum_{i=0}^{\infty} \phi_i^2} = \frac{1}{2}\left\{|j-1|^{2d+1} - |j|^{2d+1} + |j+1|^{2d+1}\right\}. \tag{3.18}$$

Robinson (1991) developed Lagrange multiplier tests for no-ARCH against alternatives consisting of general finite parameterisation of 3.5, specialising to 3.15 and 3.18. In both these cases, the autoregressive weights ψ_j satisfy

$$\sum_{j=1}^{\infty} |\psi_j| < \infty \tag{3.19}$$

Under 3.19 and

$$\max_t E\left(\varepsilon_t^4\right) < \infty, \tag{3.20}$$

it follows that

$$\begin{aligned} E\left(\nu_t^2\right) &\leq E\Big\{\sum_{j=0}^{\infty} \psi_j (\varepsilon_{t-j}^2 - \sigma^2)\Big\}^2 \\ &\leq \Big(\sum_{j=0}^{\infty} |\psi_j|\Big)\Big(\sum_{j=0}^{\infty} |\psi_j| E(\varepsilon_{t-j}^4)\Big) \\ &\leq K \end{aligned} \tag{3.21}$$

where K is a generic constant, so the innovations ν_t in 3.10 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 3.15 or 3.18. Even if 3.20 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 3.17. Both parametric representations 3.15 and 3.18 have the implication that autocorrelations follow

$$\frac{\sum_{i=0}^{\infty} \phi_i \phi_{i+j}}{\sum_{i=0}^{\infty} \phi_i^2} \sim c j^{2d_\varepsilon - 1} \quad \text{as } j \to \infty \tag{3.22}$$

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

$$\phi_j = O(j^{d_\varepsilon - 1}) \quad \text{as } j \to \infty. \tag{3.23}$$

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

$$\gamma_j = \mathrm{cov}(\varepsilon_t^2, \varepsilon_{t+j}^2). \tag{3.24}$$

The rate of convergence of the sample mean is also characteristic of long memory processes when ϕ_j satisfies 3.23. Indeed, 3.12 and 3.21 imply that the partial sums of the squared innovations have variance

$$\begin{aligned} \mathrm{Var}[\sum_{t=1}^{n}(\varepsilon_t^2 - \sigma^2)] &= \sum_{s,t=1}^{n}\sum_{j=0}^{\infty}\phi_j\phi_{j+s-t}E(\nu_{t-j}^2) && (3.25)\\ &= O(n\sum_{t=1}^{n}\Theta_t) && (3.26)\end{aligned}$$

with

$$\Theta_t := \sum_{j=0}^{\infty}|\phi_j\phi_{j+t}| = O(t^{2d_\varepsilon - 1}) \tag{3.27}$$

under 3.23. Therefore, we have the following rate upper bound:

$$\sum_{t=1}^{n}(\varepsilon_t^2 - \sigma^2) = O_p(n^{d_\varepsilon + \frac{1}{2}}). \tag{3.28}$$

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.

4 Averaged periodogram estimation of short memory

In this section, the spectral density of x_t with autocovariances satisfying

$$\mathrm{cov}(x_t, x_{t+j}) = \int_{-\pi}^{\pi} f(\lambda)\cos(j\lambda)d\lambda \quad j = 0, \pm 1, \dots , \tag{4.29}$$

is supposed to be bounded and positive at zero frequency. We are therefore considering short memory time series and proving the asymptotic normality of the averaged periodogram estimate of $f(0)$ in a context of possibly long memory conditionally heteroscedastic innovations. This may increase the reliability of frequency domain autocorrelation consistent variance estimation procedures. Another possible application is in the framework of a Beverage-Nelson decomposition where $f(0)$ is proportional to the variance of the long run component.

For a realization of size n, we consider the discrete Fourier transform of x_t be defined as

$$w_x(\lambda) = (2\pi n)^{-\frac{1}{2}} \sum_{t=1}^{n} x_t e^{it\lambda} \tag{4.30}$$

and the periodogram as

$$I_x(\lambda) = |w_x(\lambda)|^2. \tag{4.31}$$

Define the averaged periodogram by

$$\hat{F}(\lambda) = \frac{2\pi}{n} \sum_{j=1}^{[\lambda n/2\pi]} I_x(\lambda_j) \tag{4.32}$$

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 4.32. $\hat{F}(\lambda)$ is a discrete analogue of the more widely documented continuously averaged periodogram (see Ibragimov (1963)) where 4.31 is replaced by its demeaned version. The estimate

$$\hat{f}(0) = \hat{F}(\lambda_m)/\lambda_m = \frac{1}{m} \sum_{j=1}^{m} I_x(\lambda_j), \tag{4.33}$$

were m is the bandwidth sequence satisfying at least

$$\frac{1}{m} + \frac{m}{n} \to 0 \quad \text{as} \quad n \to \infty, \tag{4.34}$$

was proposed for $f(0)$ by Grenander and Rosenblatt (1966) and is readily generalisable to a wide class of weighted periodogram spectral estimates.

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

$$f(\lambda) = f(0)(1 + E_\beta \lambda^\beta) + o(\lambda^\beta) \quad \text{as} \quad \lambda \to 0^+, \tag{4.35}$$

with

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

and suppose the bandwidth m satisfies

$$\frac{1}{m} + \frac{m^{2\beta+1}}{n^{2\beta}} \to 0 \quad \text{as} \quad n \to \infty. \tag{4.36}$$

Under the conditions above, asymptotic normality of $\hat{f}(0)$ given by 4.33

$$m^{\frac{1}{2}}(\hat{f}(0) - f(0)) \rightarrow_d N(0, f(0)^2) \quad \text{as} \quad n \rightarrow \infty \tag{4.37}$$

occurs under the two following sets of sufficient conditions: Brillinger (1975), Theorem 5.4.3, page 136 assumes summability of all moments and cumulants of x_t; Hannan (1970), Theorem 13, page 224, assumes that x_t follows 3.1 with i.i.d. innovations. Hannan (1970), Theorem 13', page 227 also proves 4.37 under uniform mixing and fourth order stationarity with absolutely summable fourth cumulants. 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).

The following theorem shows that the discretely averaged periodogram given by 4.33 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 4.35. In addition, in a neighbourhood $(0, \delta)$ of the origin, $\alpha(\lambda)$ is differentiable and

$$\frac{d}{d\lambda} \log \alpha(\lambda) = O(\frac{|\alpha(\lambda)|}{\lambda}) \quad \text{as } \lambda \rightarrow 0^+, \tag{4.38}$$

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

Assumption A2 m satisfies 4.36 and

$$\frac{m}{n^{1-2d_\varepsilon}} \rightarrow 0, \quad 0 < d_\varepsilon < \frac{1}{2}. \tag{4.39}$$

Assumption A3 x_t satisfies 3.1, 3.2, 3.4, 3.5, 3.10, 3.17 and 3.19. In addition, 3.23 holds with the same d_ε as in 4.39 and

$$\max_t E\varepsilon_t^8 < \infty, \tag{4.40}$$

$$E\left(\varepsilon_t^4 \varepsilon_u | \mathcal{F}_{t-1}\right) = E\left(\varepsilon_t^4 \varepsilon_u^2 \varepsilon_v | \mathcal{F}_{t-1}\right) = 0, \text{ a.s., } t \geq u \geq v, \tag{4.41}$$

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

$$\alpha_{j+1} < \alpha_j \left(1 + \frac{B}{j}\right). \tag{4.42}$$

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. 4.38 is needed for the treatment of fourth cumulant moments of the errors, to justify the martingale approximation. 4.36 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. 4.39 strengthens 4.36 unless $d_\varepsilon \leq 1/(4\beta+2)$. 4.39 is required for the left-hand side of 4.37 to converge in distribution 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:

$$\sum_{j=0}^{\infty} |\alpha_j| < \infty. \tag{4.43}$$

The requirement 4.41 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

$$\varepsilon_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2). \tag{4.44}$$

Note that 4.40 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 4.44, and of Nelson (1990) under more general distributional assumptions. However, 4.40 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,

$$|\alpha_j - \alpha_{j+1}| \leq K \frac{|\alpha_j|}{j}. \tag{4.45}$$

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 4.39 is also needed when the ε_t^2 have long memory, and there always exists an m sequence satisfying both 4.36 and 4.39.

Theorem 1 Under Assumptions A1-A3, 4.37 holds.

Robinson (1983) gives a survey of the possible applications of spectral estimation. Some of the major applications of 4.33 are documented by Robinson and Velasco (1996). They show how a consistent estimate of the spectral density at frequency zero of a weakly dependent process is instrumental in location inference, linear regression and more complex econometric models. 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. Another obvious application of zero frequency spectral estimation is long run variance estimation in a Beveredge-Nelson type decomposition. This can be particularly useful in the investigation of financial market efficiency.

5 Averaged periodogram estimation of long memory

In this section, long memory in a weakly stationary time series x_t, $t = 0, \pm 1, \ldots$, with autocovariances satisfying 4.29 will be modelled semiparametrically by

$$f(\lambda) \sim L(\frac{1}{\lambda})\lambda^{-2d_x} \quad \text{as} \quad \lambda \to 0^+, \quad \text{with} \quad 0 < d_x < \frac{1}{2}, \tag{5.46}$$

where $L(\lambda) > 0$ is a slowly varying function at infinity defined by

$$\frac{L(t\lambda)}{L(\lambda)} \to 1 \quad \text{as} \quad \lambda \to \infty \quad \text{for all} \quad t > 0. \tag{5.47}$$

Under 5.46, $f(\lambda)$ has a pole at $\lambda = 0$ for $0 < d_x < \frac{1}{2}$ (when there is long memory in x_t). In case $f(\lambda)$ satisfies 5.46, 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). 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

$$E[v(\lambda_j)\bar{v}(\lambda_k)] = \delta_{jk}\sigma^2 + O(\frac{\log j}{k}) \tag{5.48}$$

$$E[v(\lambda_j)v(\lambda_k)] = O(\frac{\log j}{k}). \tag{5.49}$$

for moments of discrete Fourier transforms in a neighbourhood of frequency zero under 5.46, and very weak additional regularity conditions. Their asymptotic distribution under the current framework is still an open question. However, as shown in Robinson (1994), the slope of the spectrum at frequency zero can be consistently estimated using the averaged periodogram statistic providing

$$\frac{\hat{F}(\lambda_m)}{F(\lambda_m)} \to_p 1 \quad \text{as} \quad n \to \infty, \tag{5.50}$$

where $F(\lambda) = \int_0^\lambda f(\theta)d\theta$. This paper proves that 5.50 continues to hold when 3.3 is replaced by 3.4 with 3.5. Robinson (1994) proved 5.50 under minimal moment conditions capable of delivering convergence in probability only, whereas this chapter requires 3.20. This is unfortunate, as 3.20 reduces the scope of GARCH specifications covered by 3.5 and is not always 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 5.50 continues to hold when the error process ε_t displays (possibly long memory) conditional heteroscedasticity.

The following assumptions are introduced:

Assumption B1 $\quad f$ satisfies 5.46, for $0 < d_x < \frac{1}{2}$. In addition, 4.38 holds for $\alpha(\lambda) = \sum_{j=0}^{\infty} \alpha_j e^{ij\lambda}$.

Assumption B2 $\quad m$ satisfies 4.34.

Assumption B3 $\quad x_t$ satisfies 3.1, with 3.2, 3.5-3.10, 3.17, 3.19 and 3.20. In addition either

$$E\left(\varepsilon_t^3 | \mathcal{F}_{t-1}\right) = E(\varepsilon_t^3) \text{ a.s.,} \quad t = 0, \pm 1 \ldots, \tag{5.51}$$

or

$$\sum_{j=0}^{\infty} |\phi_j| < \infty. \tag{5.52}$$

Assumptions B1 and B2 correspond to assumptions A and B in Robinson (1994) with the addition of 4.38 which is used to control the martingale approximation term A.89 below. 4.38 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 (1994)) 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 (1994) is replaced by the fourth moment condition 3.20. Assumption A3 implies Condition C(iv), as

$$\sum_{t=1}^{n}\left(\varepsilon_t^2-\sigma^2\right)=o_p(n) \tag{5.53}$$

Indeed, left side of 5.53 has mean zero and variance

$$\sum_{t,s=1}^{n}\sum_{j,k=0}^{\infty}\phi_j\phi_k E(\nu_{t-j}\nu_{s-k})=\sum_{t,s=1}^{n}\sum_{j=0}^{\infty}\phi_j\phi_{j+s-t}E(\nu_{t-j}^2) \tag{5.54}$$

in view of 3.2, with $\phi_j=0,\ j<0$. In view of 3.21 and the Cauchy inequality, 5.54 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 5.52, thus verifying 5.53.

Again, the requirement 5.51 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 4.44. The alternative requirement 5.52 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.

Theorem 2 Under Assumptions B1-B3, 5.50 holds.

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

$$f(\lambda)=G\lambda^{-2d_x}(1+O(\lambda^{\beta}))\quad \text{as}\quad n\to\infty, \tag{5.55}$$

for some $\beta\in[0,2)$, with $0<d_x<\frac{1}{2}$ and 4.38 holds.

Assumption C2 4.34 holds.

Assumption C3 Assumption B3 holds. In addition, when 5.52 does not hold, 3.23 does with $d_\varepsilon < \frac{1}{2}$.

Assumption C1 is weaker than Condition A' in Robinson (1994) while 3.23 is needed to derive the upper bound 3.28 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 $\delta < (\frac{1}{2} - d_x)/(3 - 2d_x)$.

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

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

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 (1994) and Slutzky's Theorem, it is immediate that

Corollary 1 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

$$m = O(n^\gamma), \quad 0 < \gamma < 1 \tag{5.56}$$

on the bandwidth, and

$$L(\lambda) = GM(\lambda), \quad G > 0, \tag{5.57}$$

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

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

Corollary 2 Under Assumptions C1-C3, 5.56 and 5.57,

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

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

This estimate is particularly convenient for its great computational simplicity compared to the local Whittle estimate proposed by Künsch (1987) 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 (see Robinson (1995a)), it is not free of d_x. When $1/4 < d_x < 1/2$, its asymptotic distribution is influenced by the Rosenblatt process.

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[1], 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 3.1 with innovations ε_t satisfying following five models for the conditional variance σ_t^2 specified and discussed below.

In the first three cases, the ε_t are supposed conditionally normal (as in 4.44) with 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.

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

(ii) GARCH: $\sigma_t^2 = .3 + .2\varepsilon_{t-1}^2 + .5\sigma_{t-1}^2$. The ε_t are GARCH(1,1), with (moderately) strong autocorrelation in the ε_t^2 at "short" lags; they satisfy 3.20 (Bollerslev (1986)).

(iii) LMARCH: $\sigma_t^2 = \left\{1 - (1-L)^{.45}\right\}\varepsilon_t^2$. The ε_t have very long memory ARCH structure satisfying 3.5-3.10 and 3.15 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 (1991), with $d_\varepsilon = .45$, close to the stationarity boundary.

In model (iv), the innovations are simulated from a t_4 distribution with constant conditional variance in order to compare the effect on $\hat{d}$ of thick tails in the innovations distribution to the effect of conditional heteroscedasticity as specified in models (i), (ii) and (iii). It also allows to investigate the effect of a failure of the fourth moment condition 3.20.

As noted in Robinson and Henry (1999), the model specification 3.5 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 3.1 with conditionally Gaussian innovations following a specific form of Nelson's EGARCH, which models the leverage effect, and which will be denoted model (vi).

(v) EGARCH: $\varepsilon_t = \sigma_t z_t$, z_t are independent standard normal variables, and $\ln\sigma_t^2 = -.5 + .9\ln\sigma_{t-1}^2 - .5z_{t-1} + .5|z_{t-1}|$. 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, which are nearly always larger than .9 (see Ghysels, Harvey, and Renault (1996)). 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 3.1, $\sum_{j=0}^{\infty}\alpha_j L^j = (1-L)^{-d_x}$, we consider:

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

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

The choice of d_x=.2 in model (a) is motivated by the standard asymptotic result given in Lobato and Robinson (1996) for $0 < d_x < .25$.

Table 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
GARCH	-0.17	-0.10	-0.07	-0.10	-0.06	-0.04	-0.05	-0.04	-0.03
LMARCH	-0.17	-0.11	-0.08	-0.11	-0.06	-0.05	-0.06	-0.04	-0.03
t_4	-0.17	-0.10	-0.06	-0.09	-0.06	-0.04	-0.05	-0.04	-0.03
EGARCH	-0.17	-0.11	-0.08	-0.10	-0.06	-0.05	-0.06	-0.04	-0.03

Table 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
GARCH	0.42	0.25	0.17	0.26	0.16	0.12	0.16	0.11	0.08
LMARCH	0.41	0.26	0.19	0.26	0.18	0.14	0.18	0.13	0.10
t_4	0.41	0.25	0.16	0.25	0.16	0.11	0.16	0.10	0.07
EGARCH	0.39	0.25	0.18	0.25	0.16	0.13	0.16	0.11	0.09

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 1-2 and 3-4 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 1-2 and 3-4, we report, for the conditional variance specifications (i)-(v), Monte Carlo bias of the averaged periodogram estimate and Monte Carlo root mean squared error.

Table 5 reports relative efficiency of the averaged periodogram estimate of long memory with respect to the local Whittle estimate of Robinson (1995a), discussed in this framework in Robinson and Henry (1999). The relative efficiency is defined as the ratio of Monte Carlo mean squared errors. We make the comparison with the local Whittle estimate for each of the two d_x values (a) and (b) and models (i), (ii) and (iii) for the innovations, because an asymptotic result is given in these cases in Robinson

Table 3: 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	-0.28	-0.20	-0.16	-0.19	-0.14	-0.12	-0.14	-0.11	-0.09
GARCH	-0.29	-0.20	-0.16	-0.19	-0.14	-0.12	-0.14	-0.11	-0.09
LMARCH	-0.29	-0.20	-0.16	-0.19	-0.15	-0.12	-0.15	-0.12	-0.10
t_4	-0.28	-0.20	-0.15	-0.19	-0.14	-0.12	-0.14	-0.11	-0.09
EGARCH	-0.29	-0.20	-0.17	-0.19	-0.15	-0.12	-0.14	-0.11	-0.10

and Henry (1999).

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 2 of Robinson and Henry (1999) 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.

For models (i), (ii), (iii) and (v), the errors ε_t were sampled from a conditionally normal distribution (see 4.44) 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 (iii), truncated to 1000 lags in the long memory cases (iii). In case of (iv), $\varepsilon_t = \sigma_t z_t$, z_t are iid normal, and $\ln\sigma_t^2 = -.5 + .9L(\ln\sigma_t^2) - .5z_{t-1} + .5|z_{t-1}|$. 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.

Table 4: 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	0.42	0.26	0.19	0.25	0.17	0.13	0.17	0.13	0.10
GARCH	0.42	0.27	0.20	0.26	0.18	0.14	0.17	0.13	0.11
LMARCH	0.42	0.27	0.20	0.26	0.19	0.15	0.18	0.14	0.11
t_4	0.41	0.26	0.18	0.25	0.17	0.13	0.17	0.13	0.10
EGARCH	0.42	0.26	0.20	0.25	0.18	0.14	0.17	0.13	0.11

Table 5: Relative efficiencies

RELATIVE EFFICIENCY of the averaged periodogram compared to the local Whittle estimate of long memory applied to an ARFIMA(0, d_x, 0) series with three specified innovation structures, and two values of d_x.

		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
$d_x = .2$	IID	0.86	0.92	1.00	1.00	0.88	1.00	0.88	1.00	1.00
	GARCH	0.85	1.00	1.00	0.93	1.00	1.00	1.00	0.87	0.84
	LMARCH	0.81	1.00	1.00	0.86	0.89	0.86	1.00	1.00	1.00
$d_x = .45$	IID	0.90	0.78	0.60	0.77	0.66	0.56	0.56	0.52	0.44
	GARCH	0.86	0.86	0.65	0.78	0.62	0.53	0.60	0.51	0.44
	LMARCH	0.87	0.79	0.70	0.71	0.67	0.51	0.58	0.47	0.49

Table 6: 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	GARCH	LMARCH
$d_x = .2$	0.62	0.74	0.78
$d_x = .45$	0.16	0.24	0.32

Monte Carlo biases seem relatively unaffected by the model specification for the errors. Biases are all negative. They are identical for all error specifications in 5 cases out of the 18 n, m and d_x combinations, and the largest discrepancy never exceeds 0.03. Monte carlo RMSEs are largest for LMARCH errors in 11 cases and tie largest in 6 cases, which is due to the higher degree of dependence in the errors, but the discrepancy with iid errors never exceeds 0.03. The failure of the fourth moment conditions in the t_4 specification does not seem to have any effect in small samples. t_4 biases are identical to or smaller than iid normal biases in 16 cases, and t_4 RMSEs are identical to or smaller than iid normal RMSEs in all cases. EGARCH results lead to similar observations: EGARCH biases are identical to or smaller than iid normal biases in 8 cases, and EGARCH RMSEs are identical to or smaller than iid normal RMSEs in 11 cases.

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 15 cases. 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 6 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 $n = 1000$.

6 Conclusion

This paper has focused on the simple averaged periodogram statistic insofar as it provides insights into the dependence structure of a time series. It was shown 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 paper has also extended the applicability of Robinson (1994)'s results on the averaged periodogram statistic in the presence of long range dependence. It has been shown that all the results in Robinson (1994) 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.

References

Baillie, R., T. Bollerslev, and H. Mikkelsen (1996): "Fractionally integrated generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, 74, 3–30.

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 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.

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.

Clark, P. K. (1973): "A subordinated stochastic process model with fixed variances for speculative prices," *Econometrica*, 41, 135–156.

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.

Drost, F. C., and T. E. Nijman (1993): "Temporal aggregation of GARCH processes," *Econometrica*, 61, 909–927.

Engle, R. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1007.

Fama, E. (1991): "Efficient Capital Markets: II," *Journal of Finance*, 46, 1573–1617.

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.

Ghysels, E., A. Harvey, and E. Renault (1996): "Stochastic Volatility," in Handbook of Statistics, Vol. 14. Elsevier, New York.

Grenander, U., and M. Rosenblatt (1966): *Statistical analysis of stationary time series.* Wiley: New York.

Hannan, E. J. (1970): *Multiple time series.* Wiley: New York.

Hannan, E. J. (1979): "Non linear time series regression," *Journal of Applied Probability*, 8, 767–780.

He, C. (1997): "The fourth moment structure of the GARCH process," mimeo, Stockholm School of Economics.

Ibragimov, I. A. (1963): "On estimation of the spectral function of a stationary Gaussian process," *Theory of Probability and Applications*, 8, 366–401.

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.

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.

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.

Mandelbrot, B. B. (1973): "Le problème de la réalité des cycles lents et le 'syndrome de Joseph'," *Revue d'économie appliquée*, 26, 349–365.

Nelson, D. B. (1990): "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.

Robinson, P. M. (1983): "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. (1991): "Testing for strong serial correlation and dynamic conditional heteroscedasticity in multiple regression," *Journal of Econometrics*, 47, 67–84.

Robinson, P. M. (1994): "Semiparametric analysis of long memory time series," *Annals of Statistics*, 22, 515–539.

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 (1999): "Long and short memory conditional heteroscedasticity in estimating the memory parameters of levels," *Econometric Theory*, 15, 299–336.

Robinson, P. M., and C. Velasco (1996): "Autocorrelation-robust inference," Discussion Paper EM/96/316, STICERD, London School of Economics.

Robinson, P. M., and P. Zaffaroni (1998): "Nonlinear time series with long memory: a model for stochastic volatility," *Journal of Statistical Planning and Inference*, 68, 359–371.

Rosenblatt, M. (1981): "Limit theorems for Fourier transforms of functionals of Gaussian sequences," *Zeitschrift für Wahrscheinlichkeits-theorie und Verwandte Gebiete*, 55, 123–132.

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.

Taylor, S. (1980): "Conjectured models for trends in financial prices, tests and forecasts," *Journal of the Royal Statistical Society*, A 143, 338–362.

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.

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.

Appendix

Proof of Theorem 1

The proof of Theorem 1 relies on an approximation of the averaged periodogram by a martingale.

$$\frac{\sqrt{m}}{f(0)}\left(\frac{\hat{F}(\lambda_m)}{\lambda_m}-f(0)\right) = \frac{m^{-\frac{1}{2}}}{f(0)}\sum_{j=1}^{m}\left(I_x(\lambda_j)-\frac{2\pi}{\sigma^2}f(\lambda_j)I_\varepsilon(\lambda_j)\right) \tag{A.58}$$

$$+ \frac{m^{-\frac{1}{2}}}{f(0)}\sum_{j=1}^{m}f(\lambda_j)\left(\frac{2\pi I_\varepsilon(\lambda_j)}{\sigma^2}-1\right) \tag{A.59}$$

$$+ \frac{m^{-\frac{1}{2}}}{f(0)}\sum_{j=1}^{m}\left(f(\lambda_j)-f(0)\right). \tag{A.60}$$

First, we prove that the martingale approximation is valid in the present framework. From 4.35,

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

A.58 can be further decomposed into

$$m^{-\frac{1}{2}}\sum_{j=1}^{m}\left(\frac{I_x(\lambda_j)}{f(\lambda_j)}-\frac{2\pi}{\sigma^2}I_\varepsilon(\lambda_j)\right) \tag{A.61}$$

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

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

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 4.35 and 4.36. The second term has first absolute moment bounded by

$$m^{-\frac{1}{2}}\sum_{j=1}^{m}\left|1-\frac{f(\lambda_j)}{f(0)}\right|E\left(I_x(\lambda_j)\right). \tag{A.62}$$

Now,

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

where the second equality is derived from application of 3.1, the third equality is derived from application of 3.2 and the last inequality follows from 4.43. Therefore, A.62 is bounded by

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

from 4.35 and 4.36. A.61 has second moment equal to

$$
m^{-1}\sum_{j,k=1}^{m} E\left(\frac{I_x(\lambda_j)}{f(\lambda_j)}-\frac{2\pi}{\sigma^2}I_\varepsilon(\lambda_j)\right)\left(\frac{I_x(\lambda_k)}{f(\lambda_k)}-\frac{2\pi}{\sigma^2}I_\varepsilon(\lambda_k)\right). \tag{A.63}
$$

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 3.4, with σ_t^2 defined by 3.5, however,

$$
\begin{aligned}
\mathrm{cum}(\varepsilon_r,\varepsilon_s,\varepsilon_t,\varepsilon_u) &= \kappa \quad \text{if} \quad r=s=t=u, && \text{(A.64)}\\
&= \gamma_{r-s} \quad \text{if} \quad r=t\neq s=u, && \text{(A.65)}\\
&= \gamma_{r-t} \quad \text{if} \quad r=s\neq t=u, && \text{(A.66)}\\
&= \gamma_{r-s} \quad \text{if} \quad r=u\neq t=s, && \text{(A.67)}
\end{aligned}
$$

and zero otherwise. The complete fourth cumulant contribution to A.63 is the following:

$$
\begin{aligned}
& \frac{1}{m(2\pi n)^2}\sum_{j,k=1}^{m}\frac{1}{f(\lambda_j)f(\lambda_k)}\sum_{r,s,t,u=1}^{n}\mathrm{cum}(x_r,x_s,x_t,x_u)e^{i(r-s)\lambda_j-i(t-u)\lambda_k} && \text{(A.68)}\\
+ & \frac{1}{m(2\pi n)^2}\sum_{j,k=1}^{m}\left(\frac{2\pi}{\sigma^2}\right)^2\sum_{r,s,t,u=1}^{n}\mathrm{cum}(\varepsilon_r,\varepsilon_s,\varepsilon_t,\varepsilon_u)e^{i(r-s)\lambda_j-i(t-u)\lambda_k} && \text{(A.69)}\\
- & \frac{1}{m(2\pi n)^2}\sum_{j,k=1}^{m}\frac{1}{f(\lambda_k)}\frac{2\pi}{\sigma^2}\sum_{r,s,t,u=1}^{n}\mathrm{cum}(\varepsilon_r,\varepsilon_s,x_t,x_u)e^{i(r-s)\lambda_j-i(t-u)\lambda_k} && \text{(A.70)}\\
- & \frac{1}{m(2\pi n)^2}\sum_{j,k=1}^{m}\frac{1}{f(\lambda_j)}\frac{2\pi}{\sigma^2}\sum_{r,s,t,u=1}^{n}\mathrm{cum}(x_r,x_s,\varepsilon_t,\varepsilon_u)e^{i(r-s)\lambda_j-i(t-u)\lambda_k} && \text{(A.71)}
\end{aligned}
$$

Now, applying 3.1,

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

in view of A.64-A.67 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 A.65-A.67 is the following:

$$\frac{m^{-1}n^{-2}}{(2\pi)^2}\sum_{j,k=1}^{m}\frac{1}{f(\lambda_j)f(\lambda_k)}\sum_{r,s,t,u=1}^{n}\sum_{\substack{p\neq q\\ -\infty}}^{n}\gamma_{p-q}\left(\alpha_{r-p}\alpha_{s-p}\alpha_{t-q}\alpha_{u-q}\right.$$
$$\left.+\alpha_{r-p}\alpha_{s-q}\alpha_{t-p}\alpha_{u-q}+\alpha_{r-p}\alpha_{s-q}\alpha_{t-q}\alpha_{u-p}\right)e^{i(r-s)\lambda_j-i(t-u)\lambda_k} \quad (A.72)$$

$$+\ \frac{m^{-1}n^{-2}}{\sigma^4}\sum_{j,k=1}^{m}\sum_{\substack{r\neq s\\ 1}}^{n}\gamma_{r-s}\left(1+e^{i(r-s)(\lambda_j+\lambda_k)}+e^{i(r-s)(\lambda_j-\lambda_k)}\right) \quad (A.73)$$

$$-\ \frac{2m^{-1}n^{-2}}{2\pi\sigma^2}\sum_{j,k=1}^{m}\frac{1}{f(\lambda_k)}\sum_{r,s,t,u=1}^{n}\gamma_{t-u}\left(\alpha_{r-t}\alpha_{s-u}\right.$$
$$\left.+\alpha_{r-u}\alpha_{s-t}\right)e^{i(r-s)\lambda_j-i(t-u)\lambda_k} \quad (A.74)$$

$$-\ \frac{2m^{-1}n^{-2}}{2\pi\sigma^2}\sum_{j,k=1}^{m}\frac{1}{f(\lambda_k)}\sum_{r,s,t=1}^{n}\sum_{p=-\infty}^{\min(t,u)}\gamma_{r-p}\alpha_{t-p}\alpha_{u-p}e^{-i(t-u)\lambda_k}. \quad (A.75)$$

From 4.43, 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 3.23, and therefore, from 4.39, it is $o(1)$ as required.

There remains to prove that

$$m^{-\frac{1}{2}} \sum_{j=1}^{m} f(\lambda_j) \left(\frac{2\pi I_{\varepsilon j}}{\sigma^2} - 1 \right) \to_p N(0,1). \tag{A.76}$$

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 $n \to \infty$

$$\sum_{t=1}^{n} E(z_t^4) \to 0, \tag{A.77}$$

$$\sum_{t=1}^{n} E\left(z_t^2 | \mathcal{F}_{t-1}\right) \to_p \sigma^4, \tag{A.78}$$

which, following Brown's Martingale Central Limit Theorem (in Brown (1971)), implies A.76. By the Schwarz inequality, $E\left(z_t^4\right) \leq (E\varepsilon_t^8)^{\frac{1}{2}}(E\xi_t^8)^{\frac{1}{2}}$. Because $\xi_t = \sum_{s=1}^{t-1} \varepsilon_s c_{t-s}$ is a martingale, by Burkholder's inequality (Burkholder (1973)),

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

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

$$\begin{aligned}
\sum_{s=1}^{n} c_s^2 &= \sum_{s=1}^{n} 4m^{-1}n^{-2} \left(\sum_{j=1}^{m} f_j \cos s\lambda_j \right)^2 \\
&= 4m^{-1}n^{-2} \sum_{j=1}^{m} f_j^2 \sum_{s=1}^{n} \cos^2 s\lambda_j \qquad \text{(A.79)} \\
&+ 2m^{-1}n^{-2} \sum_{\substack{j \neq k \\ 1}}^{m} f_j f_k \sum_{s=1}^{n} \left[\cos s(\lambda_j + \lambda_k) + \cos s(\lambda_j - \lambda_k)\right].
\end{aligned}$$

A.79 is $O(1/n)$ whereas, using trigonometric identities in Zygmund (1977) p. 49,

$$\begin{aligned}
&\sum_{s=1}^{n} \left[\cos s(\lambda_j + \lambda_k) + \cos s(\lambda_j - \lambda_k)\right] = \\
&\frac{\sin(n+\frac{1}{2})(\lambda_j + \lambda_k)}{2\sin\frac{1}{2}(\lambda_j + \lambda_k)} + \frac{\sin(n+\frac{1}{2})(\lambda_j - \lambda_k)}{2\sin\frac{1}{2}(\lambda_j - \lambda_k)} - 1 = 0
\end{aligned}$$

for $j \neq k$, so that

$$\sum_{s=1}^{n} c_s^2 = O\left(\frac{1}{n}\right), \tag{A.80}$$

which in turn implies

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

to verify A.77. To check A.78, 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),

$$\sum_{t=1}^{n}\xi_t^2 - \sigma^2 = \sum_{t=1}^{n-1}\chi_t r_{n-t} + \sigma^2\left\{\sum_{t=1}^{n-1} r_{n-t} - 1\right\} + \sum_{t=2}^{n}\sum_{r\neq s}\varepsilon_r\varepsilon_s c_{t-r}c_{t-s}, \tag{A.81}$$

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

$$\sum_{t=1}^{n-1}\sum_{u=1}^{n-1}\gamma_{t-u}r_{n-t}r_{n-u}, \tag{A.82}$$

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

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

Robinson (1995a) shows that

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

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

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

$$\sum_{t=1}^{n-1} r_{n-t} \to 1, \quad \text{as } n \to \infty. \tag{A.83}$$

As, moreover,

$$|\gamma_j| \leq K\Phi_0\Phi_j = O\left(j^{2d_\varepsilon-1}\right) \to 0, \quad \text{as } j \to \infty \tag{A.84}$$

by 4.40 and 3.23, it follows from the Toeplitz lemma that A.82 tends to zero. Clearly, the second term in A.81 thus tends to zero, whereas the last term has mean zero and variance bounded by

$$2\left(\max_t E\varepsilon_t^4\right)\sum_{t,u=2}^{n}\sum_{\substack{r\neq s\\1}}^{\min(t-1,u-1)}|c_{t-r}c_{t-s}c_{u-r}c_{u-s}|. \tag{A.85}$$

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 A.85 by

$$K\sum_{t=2}^{n}\sum_{\substack{r\neq s\\1}}^{t-1}c_{t-r}^2c_{t-s}^2+K\sum_{t=3}^{n}\sum_{u=2}^{t-1}\sum_{\substack{r\neq s\\1}}^{u-1}|c_{t-r}c_{t-s}c_{u-r}c_{u-s}|. \tag{A.86}$$

From A.80, the left-hand side is $O(1/n)$. By summation by parts, we have

$$|c_s| \;=\; 2m^{-1/2}n^{-1}\left|\sum_{j=1}^{m-1}(f_j-f_{j+1})\sum_{l=1}^{j}\cos s\lambda_l+f_m\sum_{j=1}^{m}\cos s\lambda_l\right|. \tag{A.87}$$

Now $|\sum_{l=1}^{j}\cos s\lambda_l| = O(n/s)$ for $1\leq j\leq m$ and $1\leq s\leq n/2$ by Zygmund (1977) page 2. Moreover, by 4.35,

$$\sum_{j=1}^{m}|f_j-f_{j+1}| = O\left(\sum_{j=1}^{m}|\lambda_j^{\beta}-\lambda_{j+1}^{\beta}|\right) = O\Big((m/n)^{\beta}\Big). \tag{A.88}$$

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 A.85 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 A.85 is $O\left(m^{-1/3}\right)$ in view of A.80. It remains to show that

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

The remainder of the proof of Theorem 2 of Robinson and Henry (1999) applies to establish this result because the quantities c_t and r_t have tighter upper bounds than the equivalently denoted ones in Robinson and Henry (1999).

Proof of Theorem 2 Calling $I_\varepsilon(\lambda)$ the periodogram of the innovations, Robinson (1994) wrote

$$\hat{F}(\lambda_m) - F(\lambda_m) = \frac{2\pi}{n}\sum_{j=1}^{m}\left(I_x(\lambda_j) - \frac{2\pi}{\sigma^2}f(\lambda_j)I_\varepsilon(\lambda_j)\right) \quad \text{(A.89)}$$

$$+ \frac{2\pi}{n}\sum_{j=1}^{m} f(\lambda_j)\left(\frac{2\pi I_\varepsilon(\lambda_j)}{\sigma^2} - 1\right) \quad \text{(A.90)}$$

$$+ \frac{2\pi}{n}\sum_{j=1}^{m} f(\lambda_j) - F(\lambda_m). \quad \text{(A.91)}$$

The first parts of Propositions 1 and 2 of Robinson (1994) carry through to Assumptions A1 to A3, so that A.91 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 A.89 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 5.46 and lemma 3(ii) in Robinson (1994), A.89 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 A.90 is $o_p(F(\lambda_m))$. The left-hand side of A.90 is proportional to

$$\left\{\frac{1}{n}\sum_{j=1}^{m} f(\lambda_j)\right\}\frac{1}{n\sigma^2}\sum_{j=1}^{n}(\varepsilon_t^2 - \sigma^2) + \frac{2}{n\sigma^2}\sum_{s<t}\varepsilon_t\varepsilon_s\mu_{t-s}, \quad \text{(A.92)}$$

where $\mu_t = \frac{1}{n}\sum_{j=1}^{m} f(\lambda_j)\cos t\lambda_j$. In view of 5.53, it is sufficient to show that

$$\sum_{s<t}\varepsilon_t\varepsilon_s\mu_{t-s} = o_p(nF(\lambda_m)). \quad \text{(A.93)}$$

The left hand side of A.93 has variance

$$\sum_{s<t}E(\varepsilon_t^2\varepsilon_s^2)\mu_{t-s}^2 + 2\sum_{t>s>u}E(\varepsilon_t^2\varepsilon_s\varepsilon_u)\mu_{t-s}\mu_{t-u} \quad \text{(A.94)}$$

from 3.2. Substituting 3.10 in the second term of A.94 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 3.2 for $j > t-s$, and 3.12 for $j < t-s$, whereas the second equality employs 3.11 and nested conditional expectations with 3.2 to verify $E(\sigma_s^2\varepsilon_s\varepsilon_u) = 0$ for $u < s$. Under 5.51, this is identically zero. Under 5.52, it is bounded in absolute value by

$$\begin{aligned} 2n\max_t E(\varepsilon_t^4)\sum_{t>s=1-n}^{n-1}|\phi_s\mu_s\mu_t| &\leq KnF(\lambda_m)\sum_{j=0}^{\infty}|\phi_j|\sum_{s=1}^{n}|\mu_s| \\ &\leq KnF(\lambda_m)\sum_{s=1}^{n}|\mu_s|, \end{aligned} \qquad (A.95)$$

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

Now, following Robinson (1994),

$$\sum_{s=1}^{n}|\mu_s| = O\left(rF(\lambda_m) + n\max_{r<t\leq n/2}|\mu_t|\right) \qquad (A.96)$$

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 (1994),

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

Choosing $r \sim nm^{(1-2d_x)/(2d_x-2)}$, which is indeed larger than n/m, yields the tightest bound for A.96, i.e. $O(nF(\lambda_m)m^{(2d_x-1)/(2-2d_x)})$. It follows that A.95 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 5.54 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.

<u>Proof of Theorem 3</u> The proof of Theorem 3 in Robinson (1994) still applies to show that A.91 is $O_p[((\frac{m}{n})^\beta + m^{-\delta})F(\lambda_m)]$. The proof of Theorem 1 established that

A.89 was $o_p(m^{\eta-\frac{1}{2}})$ for any $\eta > 0$, so that A.89 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 A.90, the first term in A.92 is $O(n^{d_\varepsilon - \frac{1}{2}} F(\lambda_m))$ from 3.28, 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 $d_x \in (0, \frac{1}{2})$.