Generating Schemes for Long Memory Processes: Regimes, Aggregation and Linearity *

James Davidson †
Cardiff Business School, Cardiff University, Colum Drive, Cardiff CF10 3EU, UK.
davidsonje@cardiff.ac.uk.

Philipp Sibbertsen, ‡
Fachbereich Statistik, Universität Dortmund, Vogelpothsweg 87
D-44221 Dortmund,
Germany
sibberts@statistik.uni-dortmund.de

March 2004, forthcoming in *Journal of Econometrics*

JEL Classification: C22
Keywords: Long memory, regime-switching, aggregation, Levy motion.

Abstract

This paper analyses a class of nonlinear time series models exhibiting long memory. These processes exhibit short memory fluctuations around a local mean (regime) which switches randomly such that the durations of the regimes follow a power law. We show that if a large number of independent copies of such a process are aggregated, the resulting processes are Gaussian, have a linear representation, and converge after normalisation to fractional Brownian motion. Alternatively, an aggregation scheme with Gaussian common components can yield the same result. However, a non-aggregated regime process is shown to converge to a Levy motion with infinite variance, suitably normalised, emphasising the fact that time aggregation alone fails to yield a FCLT. Two cases arise, a stationary case in which the partial sums of the process converge, and a nonstationary case in which the process itself converges, the Hurst coefficient falling in the ranges $(\frac{1}{2}, 1)$ and $(0, \frac{1}{2})$ respectively. We comment on the relevance of our results to the interpretation of the long memory phenomenon, and also report some simulations aimed to throw light on the problem of discriminating between the models in practice.

---

*Presented at the NSF/NBER Time Series Conference, Philadelphia, September 2002. We are grateful in particular to Mark Jensen, Richard Davis and Rob Engle for useful comments.

†Research supported by the ESRC under award L138251025..

‡Research undertaken while visiting Cardiff University. The support of Volkswagenstiftung is gratefully acknowledged.

1 Introduction

Autoregressive unit roots are a popular feature of econometric models, not least thanks to the attractive feature that stationarity can be induced by either differencing or forming cointegrating linear combinations of economic time series. However, an often remarked drawback with this approach is that many important series do not seem to fall, logically or empirically, into either of the I(0) (stationary) or I(1) (difference stationary) categories. Their movements may appear mean reverting, for example, yet too persistent to be explained by a stationary, short-memory process. The fractionally integrated class of long memory models provide a seemingly attractive alternative, in which the I(1)/I(0) dichotomy is replaced by a continuum of persistence properties. In this class, a time series x_t has the representation $(1-L)^d x_t = u_t$ for $-\frac{1}{2} < d < \frac{1}{2}$ where u_t is a stationary, weakly dependent, zero mean process. See Granger and Joyeux (1980), Hosking (1981) and Beran (1994) among other well-known references on these models. As detailed in Davidson (2002b), cointegration theory can be adapted straightforwardly to this set-up. Davidson and deJong (2000) show that the normalized partial sums of such series converge to fractional Brownian motion (fBM) under quite general conditions .

However, this approach has its own drawback, that fractional integration cannot be modelled by difference equations of finite order. Thinking of a time series model as describing a representative agent's actions, incorporating hypothesised behavioural features such as adjustment lags and rational expectations, it is natural to see this behaviour as conditioned on the 'recent past', represented by at most a finite number of autoregressive lags. Unless a unit root is involved, all such models exhibit exponentially short memory. It is impossible to generate hyperbolic memory decay from finite order difference equations. Long memory models necessarily involve the infinite history of the observed process, and devising economic models with this structure is, for obvious reasons, a lot harder than constructing finite order models. A series can, of course, be modelled to have long memory characteristics through an error correction model driven by exogenous long memory; but finding a plausible route to endogenous long memory is difficult.

The attempts to devise such mechanisms in the literature have abandoned the representative-agent dynamic framework in favour of some form of cross-sectional aggregation. Since macroeconomic time series are not in fact generated by the behaviour of a fictional representative agent, but represent the net effect of many heterogeneous agents interacting, cross-sectional aggregation is a plausible modelling framework, although it poses some severe conceptual difficulties. The best known example is due to Granger (1980) who, exploiting concepts developed independently by Robinson (1978), pointed out that summing a collection of low-order ARMA processes yields an ARMA process of higher order and, eventually, of infinite order. By arranging for the largest autoregressive roots of the micro-processes to be drawn from a Beta distribution with a concentration of mass close to 1, Granger showed that the resulting moving average coefficients decline hyperbolically, and hence can be closely approximated by a fractional-integration process. This approach has been used by, among others, Ding and Granger (1996) to model conditional heteroscedasticity in financial time series, and Byers, Davidson and Peel (1997, 2000, 2002) to model the dynamics of opinion polling.

More recent contributions have focused on the aggregation of nonlinear processes. Taqqu, Willinger and Sherman (1997), Parke (1999) and Mikosch et. al. (2002) propose similar models, involving the aggregation of persistent shocks whose durations follow a power law distribution. For example, in the context of modelling ethernet traffic, Taqqu et al. aggregate binary processes switching between 0 and 1 where the switch-times are distributed according to a power law. They invoke the central limit theorem 'sideways' to establish Gaussianity of the finite dimensional distributions, and then show that the power law entails the inter-temporal covariance structure of fBM, so that (in a continuous-time framework) this distribution must describe the aggregate process.

Parke's (1999) error duration (ED) model considers the cumulation of a sequence of random variables that switch to 0 after a random delay that again follows a power law. Thus, were the delays of infinite extent the process would be a random walk, and if of zero extent, an i.i.d. process. Controlling the probability of decay allows the model to capture persistence anywhere between these extremes. Parke shows that the ED process has the same covariance structure as the fractionally integrated linear process, and does not consider the question of convergence to fBM. This is an issue we consider in the sequel.

Diebold and Inoue (2000) are concerned with the issue of confusing fractionally integrated processes with processes that are stationary and short memory, but exhibit periodic 'regime shifts', i.e., random changes in the series mean. They show that if such switches occur with a low probability related to sample size (T), then the variance of the partial sums will be related to sample size in just the same way as a fractionally integrated process. Thus, the variance of the partial sums of an I(0) process increase by definition at the rate T, whereas that of a fractional long memory (I(d)) process increase at the rate T^{1+2d}. Diebold and Inoue show that exactly the same behaviour is observed if an independent process is added to a random variable that changes value with a particular low probability. If this probability is $p = O(T^{2d-2})$ for $0 < d < 1$, then the variance of the partial sums grows like T^{1+2d}. Hence, it is argued, such a process might be mistaken for a fractionally integrated process in a given sample. We also comment on this conclusion in the sequel.

The paper is structured as follows. Section 2 describes a class of nonlinear models based on random switches of regime (local mean) with durations following a power law. We establish the basic property of the processes, that the autocorrelations also follow a power law, and describe a simple mechanism for contingent regime shifts which preserves this property. Section 3 then considers processes formed by the cross-sectional aggregation of regime-switching models. Two aggregation schemes are described, one with independent micro-units (Section 3.1) and one where the micro-processes are driven in part by common components (Section 3.2). Section 4 then develops the properties of the aggregate processes, showing that under either model they have a linear representation in the limit, and deriving an invariance principle. It is also shown that different limit processes arise without aggregation. Section 5 extends the analysis to the case of nonstationary processes, in which the mean length of a regime is infinite, although it is shown that the difference processes can be analysed after a rescaling modification. Section 6 relates our findings to the cited literature on these models, and Section 7 reports some simulations of tests of linearity in an ARFIMA framework. Section 8 contains some concluding remarks, commenting on the possible extension to include deterministic components. Section 9 collects the proofs of the main results.

2 A Stochastic Regimes Model

The building blocks of the models we consider in this paper are processes having the form

$$X_t = m_t + \varepsilon_t \tag{2.1}$$

where ε_t is a stationary, short-memory 'I(0)' process with zero mean,[1] and

$$m_t = k_j, \ S_{j-1} < t \leq S_j$$

[1] We define an I(0) process as one whose normalised partial sums converge weakly to regular Brownian motion. See Davidson (2002a) for further details.

where $\{S_j, -\infty < j < \infty\}$ is a strictly increasing, integer-valued random sequence, and $\{k_j, -\infty < j < \infty\}$ is a real zero-mean random sequence, representing the conditional mean of the process during regime j. The duration of the jth regime is the integer-valued random variable

$$\tau_j = S_j - S_{j-1}.$$

The basic assumption is that the tail probabilities of the τ_j follow a power law. A similar model, although omitting our 'noise' component ε_t, has been investigated by Liu (2000). An interesting feature of the model is that the power law can allow relatively long-lasting regimes to arise. Although these are relatively rare events measured in "regime time", they account, by construction, for a significant proportion of calendar time. In other words, the typical observed characteristic of such a series is a 'bunching' of regime changes in calendar time, periods of frequent switching interspersed with periods with periods of quiescence.

The full set of assumptions to be maintained in the sequel are as follows. These are intended to cover as many alternatives as possible, while keeping the proofs of the important properties reasonably compact. They can certainly be extended in various directions to encompass special cases without altering the basic characteristics we are interested in. We will make use of the symbol $\simeq$ as follows: $a_n \simeq b_n$ for $b_n > 0$ if $|a_n|/b_n \to C$ for some unspecified $0 < C < \infty$. This is equivalent to $a_n \sim Cb_n$, where $a_n \sim b_n$ is used to mean $|a_n|/b_n \to 1$.

Assumption 1 **(a)** *The bivariate process* $\{k_j, \tau_j\}_{-\infty}^{\infty}$ *is strictly stationary.*

(b) $P(\tau_0 = c) \simeq c^{-1-\alpha}L(c)$ *as* $c \to \infty$, $1 < \alpha < 2$, *where* $L(.)$ *is slowly varying at* ∞ *and* $\exists$ $\beta > 0$ *such that* $L(c)/\log^{\beta} c \to 0$. [2]

(c) $E(k_0) = 0$, $E(k_0^2) = \sigma_k^2 < \infty$, $E(k_0 k_s) \geq 0$ *for* $s \geq 0$, *and* $\sum_{s=0}^{\infty} E(k_0 k_s) < \infty$.

(d) *Let* $\mathcal{T}$ *denote the* σ*-field generated by* $\{\tau_j, -\infty < j < \infty\}$. *There exists a constant* $0 < B < 1$ *such that for* $s \geq 0$,

$$BE(k_0 k_s) \leq E(k_0 k_s|\mathcal{T}) \leq B^{-1}E(k_0 k_s) \ a.s. \tag{2.2}$$

(e) $\{\varepsilon_t\}_{-\infty}^{\infty}$ *is strictly stationary with* $E(\varepsilon_0) = 0$ *and* $E(\varepsilon_0^2) = \sigma_\varepsilon^2$, $\sum_{s=0}^{\infty} E(\varepsilon_0 \varepsilon_h) < \infty$, *and* $E(m_0 \varepsilon_h) = E(\varepsilon_0 m_h) = 0$ *for all* $h \geq 0$.

Assumption 1(b) implies the key power law property

$$P(\tau_0 > c) \simeq c^{-\alpha}L(c).$$

Note that the expected duration of a regime is given under Assumption 1(b) as

$$E(\tau_0) = \sum_{c=1}^{\infty} cP(\tau_0 = c) < \infty. \tag{2.3}$$

In the sequel, we shall need to calculate the probability that a randomly chosen observation falls in a regime of duration c. Defining $J(t) = \min\{j, t \leq S_j\}$, in other words the index of the regime prevailing at calendar date t, observe that for $c \geq 1$,

$$P(S_{J(t)} - S_{J(t)-1} = c) = \frac{cP(\tau_0 = c)}{E(\tau_0)}. \tag{2.4}$$

[2] In the sequel, the symbol L is used for a generic slowly varying component. For example, if L satisfies the indicated restrictions then so does L^2, which might be represented by writing $L^2 = L$.

Assumption 1(c) controls the dependence of successive regimes in a fairly natural manner. All the restrictions hold if the regimes are serially independent, for example, and also if they are connected by a first-order autoregressive process. They can certainly be relaxed in particular cases, where more specific restrictions can be invoked, but to cover all these cases would complicate the arguments excessively.

Assumption 1(d) controls the dependence between the $\{k_j\}$ and $\{\tau_j\}$ processes by extending the restrictions of part (c) to the conditional distributions. These are essentially mild constraints to ensure that the parameter α is relevant to the memory of the process in the manner to be shown subsequently.

Assumption 1(e) describes the noise process and is likewise mainly simplifying, to rule out awkward cases, and might be relaxed at the cost of more specific restrictions on the behaviour of the noise process. The main problem here is that $\varepsilon_t m_{t+h} = \varepsilon_t m_t$ so long as $t+h$ falls in the current regime, so that summability restrictions on these covariances are tricky to handle.

Under these assumptions, with $1 < \alpha < 2$, the process is covariance stationary (and hence strictly stationary) and long memory. The following theorem is similar to one obtained independently by Liu (2000). However, we extend Liu's result by allowing dependence both over time and between the various stochastic components.

Theorem 2.1 *Under Assumption 1, if* $\gamma_h = E(X_0 X_h)$ *then* $\gamma_h > 0$ *for all* h *and* $\gamma_h \simeq h^{1-\alpha}L(h)$.

A fairly wide class of data generation processes are covered by Assumption 1. In the simplest case, the pair k_j, τ_j are drawn at time S_{j-1}, and are then conditionally fixed for the duration of the jth regime. However, it is more realistic to suppose that switching times can depend on the current state of the process, and the following example shows how this might happen.

Let a random drawing at time S_{j-1} give, not τ_j, but a conditional Bernoulli distribution governing the switch date, under which the *mean* time-to-switch follows the power law. At each date t, an independent binary random variable with values 'switch' and 'don't switch' is drawn. Let p_j denote the switch probability in regime j, so that the probability of a switch after exactly m periods is $(1-p_j)^{m-1}p_j$. Therefore

$$P(m \geq x) = p_j \sum_{m=x}^{\infty} (1-p_j)^{m-1} = (1-p_j)^{x-1} \tag{2.5}$$

and the mean of the distribution is

$$\mu_j = p_j \sum_{m=0}^{\infty} m(1-p_j)^{m-1} = \frac{1}{p_j} - 1$$

so $p_j = 1/(\mu_j + 1)$. Regimes must run for at least one period, so $\mu_j \geq 1$. In the simplest case, this parameter might be drawn from the power law distribution with density

$$f(\mu) = \alpha\mu^{-1-\alpha}. \tag{2.6}$$

Note that this integrates to 1 over $[1, \infty)$, and $P(\mu > x) = x^{-\alpha}$ for $x \geq 1$, as required.

Theorem 2.2 *Let* τ_j *be the number of periods until switching of a regime driven by* μ_j, *a drawing from the distribution in (2.6). Then,* $P(\tau_j > c) \simeq c^{-\alpha}$.

With this set-up, it is more accurate to write the duration as τ_{jt}, a random variable evolving according to the rule: $\tau_{j,t+1} = \tau_{jt} + 1$ with probability $1 - p_j$, and $\tau_{j,t+1} = \tau_{j+1,t+1} = 0$, otherwise. Note that Assumption 1 allows the independent Bernoulli random variable at date t to be dependent on the innovations of the noise process ε_t, and hence a shock hitting the system can precipitate a change of regime. Only the probability of this occurrence (which can be related to the size of shock needed to precipitate the switch) is fixed at time S_{j-1}.

3 Two Models of Cross-Sectional Aggregation

Our interest in these stochastic regimes processes is their relationship with the phenomena of long memory and fractional integration. We assume that such processes govern the behaviour of agents in the economy at the micro level, but that what is observed is the aggregate of their activities. This cross-sectional aggregation is a crucial feature of the analysis. We consider two contrasting models of aggregation, which yield essentially the same distributional result, although from very different premises.

3.1 Independent Aggregation

Our first approach follows essentially that of Taqqu et. al. (1997). Consider the normalised aggregate process

$$F_t^M = M^{-1/2} \sum_{i=1}^{M} X_t^{(i)}$$

where $X_t^{(1)}, \ldots, X_t^{(M)}$ are independent copies of X_t. Note that

$$E(F_t^M F_{t+h}^M) = \gamma_h$$

follows directly from the independence, where γ_h is defined in Theorem 2.1. Let $\{F_t\}_{-\infty}^{\infty}$ denote the limiting random process as $M \to \infty$, defined by the relation

$$(F_{t_1}^M, \ldots, F_{t_K}^M) \xrightarrow{d} (F_{t_1}, \ldots, F_{t_K}) \tag{3.1}$$

where $t_1, \ldots, t_K$ is any finite collection of time coordinates and '$\xrightarrow{d}$' denotes weak convergence. Under the assumptions, note that the limit in (3.1) is multivariate Gaussian, with covariance matrix having elements $\gamma_{|t_j - t_k|}$ for $1 \leq j, k \leq K$. The extension to the infinite-dimensional process $\{F_t\}_{-\infty}^{\infty}$, stationary and Gaussian with autocovariance sequence $\{\gamma_h, h \geq 0\}$, is assured by the Kolmogorov consistency theorem (e.g. Davidson (1994) Th. 12.4). Note that allowing the micro-processes to have heterogeneous distributions, subject to the Lindeberg condition, is an easy extension that we avoid only for the sake of simplicity of exposition.

3.2 Common Components

In this model, we dispense with the assumption that the micro-processes are independent, but impose Gaussianity instead of deriving it. The model of the ith micro-process is

$$\begin{aligned} X_t^{(i)} &= m_t^{(i)} + E_t + \varepsilon_t^{(i)} \\ m_t^{(i)} &= K_{S_{J(t,i)-1}^{(i)}} + k_{J(t,i)}^{(i)} \end{aligned}$$

where $S_j^{(i)}$ is the date of switch j by individual i, and $J(t,i) = \min\{j : t \leq S_j^{(i)}\}$. Here, K_t and E_t are stationary processes representing 'macro' influences that are common to all the 'micro' processes. In other words, at switch date $S_j^{(i)}$, the ith regime mean is equated with the *current* value of the common regime process at that date, plus the idiosyncratic component $k_j^{(i)}$. The regime durations $\tau_j^{(i)}$ are strictly idiosyncratic, however. Formally, we assume the following:

Assumption 2 **(a)** *The bivariate process* $\{K_t, E_t\}_{-\infty}^{\infty}$ *is stationary and Gaussian.*

(b) $E(K_0) = 0$*,* $E(K_0^2) < \infty$*,* $E(K_0 K_h) \geq 0$ *for* $h \geq 0$*, and* $\sum_{h=0}^{\infty} E(K_0 K_h) < \infty$*.*

(c) $E(E_0) = 0$, $E(E_0^2) < \infty$, $\sum_{s=0}^{\infty} E(E_0 E_h) < \infty$, *and* $E(K_0 E_h) = E(E_0 K_h) = 0$ *for all* $h \geq 0$.

Regarded in isolation, note that any one of the individual processes $X_t^{(i)}$, subject to the common components satisfying Assumption 2 and the idiosyncratic ones Assumption 1, itself satisfies Assumption 1. It could not be distinguished from the purely independent case. However, the mechanism of aggregation is quite different.

Note that $K_{S^{(i)}_{J(t,i)-1}}$ is discretely drawn from the set $\{K_{t-1}, K_{t-2}, K_{t-3}, \ldots\}$ with probabilities governed by the power law distribution with parameter α. Define a random variable

$$1_t^{(i)}(r) = \begin{cases} 1, & S^{(i)}_{J(t,i)-1} = t - r \\ 0, & \text{otherwise,} \end{cases}$$

the indicator of the event that the ith process underwent its most recent switch of regime at time $t - r$. Noting that

$$X_t^{(i)} = \sum_{r=1}^{\infty} 1_t^{(i)}(r) K_{t-r} + k^{(i)}_{J(t,i)} + E_t + \varepsilon_t^{(i)}, \tag{3.2}$$

define $F_t^M = M^{-1} \sum_{i=1}^{M} X_t^{(i)}$. Since the switch times are distributed independently in the population, a straightforward application of the law of large numbers (e.g. Khinchine's Theorem) yields

$$F_t^M \xrightarrow{pr} F_t = \sum_{r=1}^{\infty} P(r) K_{t-r} + E_t \tag{3.3}$$

where '$\xrightarrow{pr}$' denotes convergence in probability, and

$$\begin{aligned} P(r) &= E(1_t^{(i)}(r)) \\ &= \sum_{c=r}^{\infty} P(1_t^{(i)}(r) = 1 | P(S_{J(t,i)} - S_{J(t,i)-1} = c) P(S_{J(t,i)} - S_{J(t,i)-1} = c) \\ &= \frac{1}{E(\tau_0)} \sum_{c=r}^{\infty} P(\tau_0 = c) = O(r^{-\alpha} L(r)) \end{aligned} \tag{3.4}$$

Here the third equality makes use of (2.4) together with the fact that date t has an equal chance $1/c$ of falling anywhere in a regime of length c. It can be verified that $\sum_{r=1}^{\infty} P(r) = 1$ as required. As in the independent case, we may extend from the finite-dimensional limit distributions implied by (3.3) to a limit stochastic process $\{F_t\}$.

Note the two important differences from the purely independent case, however. The limit is normalized by M^{-1}, not $M^{-1/2}$, and it does not depend on the idiosyncratic components at all, since these average out to zero. Because of these differing convergence rates, it appears difficult to build a model in which the limit depends on both 'macro' and 'micro' components. Also, be careful to note that the conclusion

$$\operatorname{plim} M^{-1} \sum_{i=1}^{M} 1_t^{(i)}(r) = P(r)$$

for any t, on which (3.3) depends, rules out a common component in the regimes processes $\tau_j^{(i)}$. These must be purely idiosyncratic. It is not hard to see that any tendency for regime switches to be co-ordinated in the population will lead to 'jumps' of the aggregate at the preferred dates, which would rule out time-invariance of the coefficients in (3.3).

We show directly that the aggregate model inherits the covariance structure of the micro processes, as follows.

Theorem 3.1 *Let Assumption 2 hold for the common components and Assumption 1 for the idiosyncratic components, which are also distributed independently of each other and of the common components. If $\gamma_h = E(F_0 F_h)$ then $\gamma_h > 0$ for all h and $\gamma_h \simeq h^{1-\alpha}L(h)$.*

4 Representation and Invariance Principle

Let $H = (3-\alpha)/2$ for $1 < \alpha < 2$, corresponding to Hurst's coefficient, so that the bounding cases $\alpha = 2$ and $\alpha = 1$ correspond to $H = \frac{1}{2}$ and $H = 1$ respectively. We next show that the aggregate process has the variance characteristics associated with long memory increments. Let

$$\sigma_T^2 = \sum_{g=1}^{T}\sum_{h=1}^{T}\gamma_{|g-h|} = E\left(\sum_{t=1}^{T} F_t^2\right)$$

and note from Theorem 2.1 that the sequence $\{\gamma_h\}$ is positive and monotone, and $\gamma_h \simeq h^{2H-2}L(h)$. It follows directly that, for any fixed g, $\sum_{h=1}^{T}\gamma_{|g-h|} = O(T^{2H-1}L(T))$. Hence

$$\sigma_T^2 = O(T^{2H}L(T)).$$

In view of the stationarity, we can assume the existence of a finite positive constant

$$\sigma^2 = \lim_{T\to\infty}(T^{2H}L(T))^{-1}\sigma_T^2.$$

The sequence $\{F_t\}$ is strictly stationary with finite variance γ_0, and purely nondeterministic, by construction. The Wold (1938) decomposition theorem (see e.g. Davidson (2000) Theorem 5.2.1) therefore implies the form

$$F_t = \sum_{j=0}^{\infty}\theta_j\eta_{t-j} \tag{4.1}$$

where the sequence $\{\eta_t\}$ is stationary and uncorrelated with variance $\sigma_\eta^2 = \gamma_0/\sum_{j=0}^{\infty}\theta_j^2$. However, $\{F_t\}$ is Gaussian, either by the CLT in the independent aggregation case or, in the common components model, because under Assumption 2, F_t in (3.3) is a linear function of Gaussian processes $E_t, K_t, K_{t-1}, K_{t-2}\ldots$ with fixed summable coefficients. According to the Wold construction, the residuals η_t are arbitrarily well approximated by finite linear combinations of the observed process. They are therefore themselves Gaussian, and, being uncorrelated, are independently and identically distributed. The conclusion may be stated formally as follows.

Theorem 4.1 *Under either Assumption 1 in the independent aggregation case, or Assumption 2 in the common components case, the limiting aggregate process $\{F_t, -\infty < t < \infty\}$ has representation (4.1) where $\sum_{j=0}^{\infty}\theta_j^2 < \infty$ and $\eta_t \sim NI(0, \gamma_0/\sum_{j=0}^{\infty}\theta_j^2)$.*

This shows that linearity need not be an intrinsic feature of the data generation process, in order for linear models to be useful for modelling purposes. As we show in Section 7, the ARFIMA(p,d,q) model could provide a good approximation in many cases, with $d = H - \frac{1}{2}$.

The next step is to establish the invariance principle. Write

$$Z_T^M(\xi) = \sigma_T^{-1}\sum_{t=1}^{[T\xi]} F_t^M,\ 0 \le \xi \le 1 \tag{4.2}$$

where $[x]$ denotes the largest integer not exceeding x.

Theorem 4.2 *$Z_T^M \xrightarrow{d} \sigma B_H$ as $M, T \to \infty$ (sequentially), where B_H denotes fractional Brownian motion of type 1*[3] *with parameter H.*

The expression '$M, T \to \infty$ (sequentially)' means that M must be taken to the limit for each $t = 1, \ldots, T$, with T fixed, and the limit of this procedure is taken with respect to T. It is clear that this is the case relevant to the present context, but note that the limit with respect to $T, M \to \infty$ (sequentially) may be different, as may any scheme of joint convergence by setting (say) $M = M(T)$ for some monotone increasing function. The next theorem in this section illustrates the importance of the distinction. See Phillips and Moon (1999) for a discussion of the relationships between sequential and joint convergence (weakly, or in probability) for double-indexed samples.

It may be the case (and this assertion is explored in the simulations reported in Section 7) that quite a low value of M is sufficient to yield an adequate linear approximation. However, the next result demonstrates that the the invariance properties obtained in Theorem 4.2 are not obtained without cross-sectional aggregation. In other words, the usual argument from time aggregation fails. We introduce the following extra assumptions.

Assumption 3 **(a)** *The sequence $\{(k_j, \tau_j),\ -\infty < j < \infty\}$ is i.i.d.*

(b) $\displaystyle\int_{\{\tau_j \le c\}} P(\sigma_k^{-1}|k_j| > c/\tau_j|\tau_j)dF(\tau_j) = o(c^{-\alpha}).$

Then, we have the following result.

Theorem 4.3 *Let $X_T(\xi) = (T^{1/\alpha}L(T))^{-1}\sum_{t=1}^{[T\xi]} X_t$, $0 \le \xi \le 1$, where X_t is defined in (2.1). If Assumptions 1 and 3 hold then $X_T \xrightarrow{d} \Lambda_\alpha$, where Λ_α is stable Levy motion with stability parameter α*

Since $1/\alpha < H$ in the range $1 < \alpha < 2$, note the implication of this result, that with $M = 1$ the process defined in (4.2) converges to zero, albeit slowly because $(3-\alpha)/2$ and $1/\alpha$ are quite close over most of the range $(1, 2)$. Of course, this fact points to the inappropriateness of normalising by the variance of the process, which is diverging as $T \to \infty$. Also, since the increments of the limit process have no variance, note how reversing the order of M and T in Theorem 4.2 cannot yield a Gaussian limit in this case.

A related analysis has been given independently by Mikosch et. al. (2002). These authors consider an independent aggregation model similar to Taqqu et al. (1997), in the context of modelling network traffic. They show that whether the limit in their model is Gaussian or Levy can be related to what, in our context, would be the relative rates of simultaneous increase of M and T. However, as just noted, such potentially interesting considerations are not really germane to the present analysis. Our two counting processes are strictly sequential, relating in the first case to the underlying data generation process, and in the second to the mode of its observation.

Assumption 3(a) is imposed just for simplicity. Results for dependent regimes are certainly available, but the additional complications with the proof go beyond the scope of the present paper, where the aim is simply to exhibit a counter-example to the Gaussian case. Assumption 3(b) is a natural extension of Assumption 1(d), and ensures that k_j does not itself contribute to the tail behaviour of the random variables $k_j\tau_j$ that feature in the proof, in such a way that α does not define the relevant power law. Again, this is just for simplicity. It will certainly be satisfied if the conditional probability declines exponentially with c, for example.

[3] See Robinson and Marinucci (1999), and also Davidson and de Jong (2000) for details.

5 The Nonstationary Case

In this section, we consider how the regimes model can be extended to the nonstationary case $0 < \alpha < 1$.[4] As noted, the expected regime duration is infinite in this case according to (2.3), and the calculations in Theorem 2.1 demonstrate that the process is not covariance stationary. However, consider the differences $\Delta X_t = \Delta m_t + \Delta\varepsilon_t$, where

$$\Delta m_t = \begin{cases} \Delta k_{J(t)}, & t = S_{J(t)-1} + 1 \\ 0, & \text{otherwise.} \end{cases}$$

where $\Delta k_{J(t)} = k_{J(t)} - k_{J(t)-1}$. In other words, the process is nonzero at date t only if t falls in regime $J(t)$, and $t-1$ in regime $J(t)-1$. However, there is now a further difficulty. Assume temporarily, for simplicity, that τ_j and k_j are independent. It then follows from (2.4) that

$$E(\Delta m_t^2) = E(\Delta k_{J(t)}^2 1_{\{t=S_{J(t)-1}+1\}}) = \frac{E(\Delta k_0^2)}{E(\tau_0)} = 0.$$

This is a somewhat paradoxical fact, since the probability of a regime switch occurring at date t is of course not zero. However, very long regimes arise with high enough probability that the set of switch dates has probability measure zero. This technical drawback prevents a covariance analysis comparable to Theorem 2.1 being performed in this case.

However, we can construct a model not subject to this difficulty. Consider the finite sequence m_t, $t = 1, ..., T$ where $m_0 = k_{J(0)}$. Assuming the usual stationary distribution for $\{k_j, \tau_j\}$, the expected duration of a realized regime, bounded at most by dates 0 and T, is

$$E_T(\tau_0) = \sum_{c=1}^{T} cP(\tau_0 = c) + T \sum_{c=T+1}^{\infty} P(\tau_0 = c) = O(T^{1-\alpha}) \tag{5.1}$$

and (still assuming k_j and τ_j independent of each other)

$$E(\Delta m_t^2) = \frac{E(\Delta k_0^2)}{E_T(\tau_0)} = O(T^{\alpha-1}).$$

Therefore consider the triangular array

$$X_{Tt} = m_{Tt} + \varepsilon_t, \;\; t = 1, \ldots, T, \;\; T \geq 1 \tag{5.2}$$

where $m_{Tt} = E_T(\tau_0)^{1/2} k_{J(t)}$. Note that the difference process is covariance stationary by construction, and the mean process now nondegenerate. We derive its properties under the following assumption.

Assumption 4 **(a)** *Assumptions 1(a), (c), (d) and (e) hold.*

(b) $P(\tau_0 = c) \simeq c^{-1-\alpha}L(c)$ *as* $c \to \infty$, $0 < \alpha < 1$, *where* $L(.)$ *is slowly varying at* ∞ *and* $\exists$ $\beta > 0$ *such that* $L(c)/\log^\beta c \to 0$.

(c) $E(k_0 k_s) \geq E(k_0 k_{s+1})$ *for* $s \geq 0$.

[4] The boundary case $\alpha = 1$ is also nonstationary, but requires special treatment and is not considered here. Note that is corresponds to the case $d = 0.5$ in the fractionally integrated model.

(d) *For $s \geq 0$,*

$$B[E(k_0k_s) - E(k_0k_{s+1})] \leq E(k_0k_s|\mathcal{T}) - E(k_0k_{s+1}|\mathcal{T}) \\ \leq B^{-1}[E(k_0k_s) - E(k_0k_{s+1})] \ a.s. \tag{5.3}$$

(e) $E(\varepsilon_0\varepsilon_h) \geq E(\varepsilon_0\varepsilon_{h+1})$ *for $h \geq 0$.*

Theorem 5.1 *Under Assumption 4 , $\gamma_h^\Delta = E(\Delta X_{Tt}\Delta X_{T,t+h})$ is independent of T and t, and $\gamma_h^\Delta < 0$ for $h \geq 1$, $\gamma_h^\Delta \simeq h^{-1-\alpha}L(h)$, and $\gamma_0^\Delta + 2\sum_{h=1}^s \gamma_h^\Delta = O(s^{-\alpha}L(s))$.*

Hence, consider the cross-sectional aggregation of independent micro-processes with this form. With T fixed, it is clear that if

$$\Delta F_{Tt}^M = M^{-1/2}\sum_{i=1}^M \Delta X_{Tt}^{(i)}$$

then

$$(\Delta F_{Tt_1}^M, \ldots, \Delta F_{Tt_K}^M) \xrightarrow{d} (\Delta F_{Tt_1}, \ldots, \Delta F_{Tt_K})$$

where the limit is Gaussian with covariances $\gamma_{|t_j-t_k|}^\Delta$, having the special sign and summability properties specified in Theorem 5.1. Although $\Delta m_{Tt} \neq 0$ only with probability of $O(T^{\alpha-1})$, taking M to the limit with T fixed ensures that each data point has a Gaussian weak limit with fixed variance, so that the sequence $\{\Delta F_{T1}, \ldots, \Delta F_{TT}\}$ is stationary and Gaussian. If T is now allowed to increase without limit, provided the convergences in M and T are strictly sequential, the same conclusion applies, so that previous arguments allow the extension to $\{\Delta F_t\}_{-\infty}^{\infty}$, stationary and Gaussian with autocovariance sequence $\{\gamma_h, h \geq 0\}$. While ΔF_t might be thought of as the weak limit of the sequence $\{\Delta F_{Tt}, T \geq t\}$, since its distribution is invariant with respect to T, this is a case of convergence only in the trivial sense.

The common components model suffers from the analogous difficulty that $E(1_t^{(i)}(r)) = 0$ from (3.4). A slightly different modification is required here. We do not need an explicit array structure, but assume the processes start at time $t = 0$, with $1_1^{(i)}(1) = 1$. In other words, all the regimes are initialised with K_0, at date $t = 1$. Then consider the process

$$X_t^{(i)} = \sum_{r=1}^t 1_t^{(i)}(r)\frac{K_{t-r} + k_{J(t,i)}^{(i)}}{P_t(1)} + E_t + \varepsilon_t^{(i)} \tag{5.4}$$

where

$$P_t(r) = E(1_t^{(i)}(r)) = \frac{\sum_{c=r}^\infty P(\tau_0 = c)}{E_t(\tau_0)} = O(t^{\alpha-1}).$$

and $E_t(\tau_0)$ is defined as in (5.1), but with t replacing T. Now, letting $\Delta F_t^M = M^{-1}\sum_{i=1}^M \Delta X_t^{(i)}$, it follows from previous arguments that, under the usual assumption of independent $\tau_j^{(i)}$,

$$\Delta F_t^M \xrightarrow{pr} \Delta F_t = K_{t-1} + \sum_{r=2}^t \left(\frac{P_t(r)}{P_t(1)} - \frac{P_{t-1}(r-1)}{P_{t-1}(1)}\right)K_{t-r} + \Delta E_t$$

The obvious modification of Theorem 3.1 is now the following.

Theorem 5.2 *Let Assumption 2 hold for the common components and Assumption 4 for the idiosyncratic components, which are also distributed independently of each other and of the common components. Then $E(\Delta F_t\Delta F_{t+h}) = \gamma_h^\Delta + O(t^{-\alpha})$ for $h \geq 0$, where $\gamma_h^\Delta < 0$ for $h \geq 1$, $\gamma_h^\Delta \simeq h^{-1-\alpha}L(h)$ and $\gamma_0^\Delta + 2\sum_{h=1}^\infty \gamma_h^\Delta = 0$.*

In either model, we can now reprise the analysis of Section 4. The chief difference is that ΔF_t rather than F_t is treated as the increment process, since of course F_t turns out to be nonstationary. The Hurst coefficient is $H = (1-\alpha)/2$, lying beween 0 and $\frac{1}{2}$, so that this is the so-called 'anti-persistent' case. We show the following.

Theorem 5.3 $\sigma_T^2 = \sum_{g=1}^{T}\sum_{h=1}^{T}\gamma_{|g-h|}^{\Delta} = O(T^{2H}L(T))$.

Theorem 5.4 *Under Assumption 4 in the independent aggregation case, plus Assumption 2 in the common components case, the representation*

$$\Delta F_t = \sum_{j=0}^{\infty}\theta_j\eta_{t-j} \tag{5.5}$$

holds where $\sum_{j=0}^{\infty}\theta_j^2 < \infty$, $\eta_t \sim NI(0, \gamma_0^{\Delta}/\sum_{j=0}^{\infty}\theta_j^2)$ *and* $\sum_{j=0}^{\infty}\theta_j = 0$.

Theorem 5.5 *Letting* $Z_T^M(\xi) = \sigma_T^{-1}F_{[T\xi]}^M$ *for* $0 \le \xi \le 1$, $Z_T^M \xrightarrow{d} \sigma B_H$ *as* $M, T \to \infty$ *(sequentially), where* B_H *denotes fractional Brownian motion of Type 1 with parameter* H.

While the features of these models have been contrived to achieve the limit result, rather than to be realistic, their essential message is that for cross-sectional aggregation to preserve covariance structure, it is necessary for the noise to be smaller than the mean process by an order of magnitude. Apart from this consideration, the rescaling is really no more than a choice of units of measurement. Note that the date $t = 0$ does not have to be the first observation in a sample, but can be moved as far back in the pre-sample period as desired, so that the nonstationary features of the 'start-up' period need not be an issue. However, we do need to be reassured that the cross-sectional CLT can "work" in this case, and some intuition can be provided as follows. The probability that $\Delta m_{T_t} \neq 0$ in (5.2) is only $O(T^{\alpha-1})$, and while the units of measurement have been scaled up by an equivalent factor to keep the variance positive, we should not expect this array to be uniformly integrable with respect to T. However, if we choose $M = T^{2-\alpha}$, then in each time period, the aggregate process is the sum of at least T nonzero components on average, and there is no problem about taking T to the limit. In practice though, rather than linking the rates, as is possible, we simply make the convergence strictly sequential.

Finally, note that a result corresponding to Theorem 4.3 is not expected. We have noted the absence of uniform integrability of the rescaled difference process, which casts doubt on whether a weak limit can exist in this case. This is an issue going beyond the scope of the present paper, but it offers an interesting problem for future research.

6 Discussion

The results of this paper point to three main conclusions. First, a fairly general class of nonlinear processes can exhibit the covariance structure associated with long memory. Second, there exists a sub-class of nonlinear processes, characterised by cross-sectional aggregation, that are observationally equivalent to fractionally integrated processes. Specifically, their normalised partial sums converge to fBM, and they have a Wold linear representation, with independent Gaussian increments. Third, there exist counter-examples demonstrating the necessity of the aggregation to obtain the last result, in which the limit of the normalised partial sums is demonstrably different from fBM.

These considerations serve to emphasise the fact that the autocovariance structure is only part of the characterisation of a fractionally integrated process. The error duration (ED) model proposed by Parke provides a useful illustration. This has the form (in Parke's notation)

$$y_t = \sum_{s=-\infty}^{t} g_{s,t}\varepsilon_s \tag{6.1}$$

where $g_{s,t}$ is the indicator of the period running from s to time $t = s + n_s$. The random variable n_s is a stochastic duration obeying a power law similar to our τ_j, and ε_s is analogous to our k_j. Our noise term ε_t is set to 0 here. With these definitions, the ED model can be accommodated in our independent aggregation framework by allowing M to depend on t.

Consider the stationary ED model, such that $1 < \alpha < 2$ where α denotes the power law parameter, as above. The number of nonzero terms in the sum at date t, M_t say, must settle down to a stationary integer random sequence. Since a new component starts up every period,

$$E(M_t) = \sum_{c=1}^{\infty} cP(n_s = c) < \infty$$

corresponding to the Parke (1999) parameter λ. In many cases, the 'birth' of a nonzero term will be matched by the 'death' of another, and then the situation is observationally equivalent to a switch of regime in a single process. In case there is no match of a birth or death, this can be treated as a component either leaving or joining the aggregate, although it might also be rationalised, in our set-up, by having $k_j = 0$ with positive probability.

It is now possible to see the sense in which the stationary ED process can be treated as fractionally integrated. When α is close to 2, $E(M_t)$ is correspondingly small, while as α approaches 1, it tends to ∞. Suppose, to take a concrete example, that

$$P(n_s = c) = \frac{c^{-\alpha-1}}{\zeta(1+\alpha)}. \tag{6.2}$$

where $\zeta(\cdot)$ denotes the Riemann zeta function. In this case

$$E(M_t) = \frac{1}{\zeta(1+\alpha)} \sum_{c=1}^{\infty} c^{-\alpha} = \frac{\zeta(\alpha)}{\zeta(1+\alpha)}.$$

Illustrative values are $\zeta(1.5)/\zeta(2.5) = 1.947$, $\zeta(1.1)/\zeta(2.1) = 6.784$, and $\zeta(1.01)/\zeta(2.01) = 61.49$. In other words, only for the case where α is very close to the nonstationary case of unity (corresponding to the I(d) process with d close to 0.5) is the number of terms in the aggregate large. Clearly, the central limit theorem cannot be invoked to justify Gaussianity in this process. While it may be that the shocks ε_t themselves are Gaussian, y_t is the sum of a randomly varying number of independent terms, and therefore its marginal distribution would be mixed Gaussian in that case. From this point of view, we must be careful to distinguish between the stationary ED process and the fractionally integrated process. In particular, the partial sums of the former process do not converge to fBM, in general.

In the nonstationary case of (6.1) the number of terms in the sum increases with time, and the conventional argument from time aggregation appears to imply a Gaussian limit. In view of the covariance structure demonstrated in Parke (1999), this suggests possible convergence to fBM with $\frac{1}{2} < d < 1$. However, the proof of this conjecture would require a different approach to that adopted for Theorem 4.2. Note that even with Gaussian shocks, a linear representation of the form (5.5) does not hold for the difference process; this is

$$\Delta y_t = \varepsilon_t - \sum_{s=-\infty}^{t-1} \Delta g_{s,t}\varepsilon_s$$

where the number of nonzero terms for $s < t$ is a random variable with mean falling between zero and 1.

Unlike the ED model, the models constructed by Diebold and Inoue (2001) are explicitly 'false', in the sense that they define stochastic arrays in which the incidence of regime shifting is linked to sample size. In all of their cases, allowing the sample size to increase sufficiently would reveal that the processes are nonlinear random walks, having the covariance characteristics of an I(1) process. The message of these authors is that modellers face a hazard of mis-identification, because the incidence of structural change is adventitiously linked to the length of available series.

However, like Parke (1999), they focus wholly on the issue of the autocovariance structure. This, as we have shown, is only one defining characteristic of a fractionally integrated process, and we have highlighted the existence of a linear representation as another. The nature of the connection between these features can be clarified informally by 'discretizing' the fBM. Let $X(\xi)$, $0 \leq \xi \leq 1$, be fBM with Hurst parameter H and, for convenience, normalised such that $EX(1)^2 = 1$. Fix a finite integer $n \geq 1$ and consider the sequence

$$x_{nt} = (2n)^{2H}(X((t+1)/2n + \tfrac{1}{2}) - X(t/2n + \tfrac{1}{2})), \quad t = -n, \dots, (n-1). \tag{6.3}$$

Then note that $x_{nt} \sim N(0,1)$ by construction, and also, as would be expected,

Theorem 6.1 $E(x_{nt}x_{n,t+h}) \simeq h^{2H-2}$.

Taking n large, let an approximate Wold decomposition, truncated at n lags, be applied to the sequence $x_{n1}, \dots, x_{nn}$. In view of the Gaussianity, the shock process in this decomposition is independently distributed, and in view of the autocovariance structure, it can also be seen that the linear representation approximates to the fractional integration model. What this shows is that if a partial sum process converges to fBM, then under some degree of time aggregation (averaging successive blocks of observations of length $[T^\beta]$ for $0 < \beta < 1$, say) the time-aggregated sequence (with $[T^{1-\beta}] + 1$ terms) should possess an increasingly exact linear representation, as T increases. We suggest the value of this remark is to show that, even if we do not postulate that the process in question is exactly linear in the sense of Theorem 4.1, there always exists a natural approach to distinguishing processes having fBM as their weak limit from alternatives. This is by tests of linearity.

7 Testing Linearity

In this section we attempt to quantify the practical role of our result, as an approximation theorem, by means of a small simulation experiment. A test of linearity is applied to generated models in the 'aggregate-of-independent-regimes' class. If the approximation is good with even a small M then arguably the result is more tolerant of failure of the assumptions, but this behaviour will clearly depend on the value of α, among other factors. We don't consider the common components model here, but since it is the independence of switch dates that delivers the crucial aggegation property in each case, the experiment should have quite general implications

For practical purposes we have to limit consideration to a class of linear parametric models with the right covariance structure, and the ARFIMA(p, d, q) class is the natural choice for this purpose. It is true that the class of models in our null hypothesis is larger than the ARFIMA class, but there are reasonable grounds to think that the ARFIMA class can approximate any linear process with the requisite properties pretty well. This is not a Monte Carlo study, since there is no special interest in determining the distribution of the estimators. The approximating ARFIMA models are simply fitted to large samples, of 20,000 data points each, so that the parameter estimates can be regarded as close to their probability limits. To simplify the model selection process, the class of models considered is restricted to the ARFIMA$(p, d, 0)$, and p was chosen to optimise the value of the consistent Schwarz (1978) criterion. For the selected equation, test statistics for model adequacy were recorded.

Our chosen diagnostic test for nonlinearity is the McLeod-Li (1983) portmanteau test, corresponding to the Box-Pierce statistic computed for the squared residuals. Although there are several possible tests for linearity, this one is appropriate since correlation in the squares is a natural dummy alternative hypothesis. Forcing a linear representation onto a process exhibiting periodic jumps in the local mean is likely to induce conditional heteroscedasticity, in the form of runs of larger than average residuals in the neighbourhood of the jumps. There is therefore hope that this test should be relatively powerful against the alternatives of interest. The Brock-Dechert-Scheinkman test of residual independence (Brock et al. 1996) was also considered, but in preliminary experiments this proved to have rather low power by comparison.

The model simulated is the Bernoulli-switching model described in Section 2, with the conditional mean duration generated from (2.6). The processes μ_j, and $k_j \sim N(0,1)$ and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ are mutually and serially independent in all cases. Experiments were conducted for $\alpha = 1.5$ and $\alpha = 0.5$, representing the stationary and nonstationary cases respectively, and for $\sigma_\varepsilon^2 = 0.2$ and $\sigma_\varepsilon^2 = 0.5$. Within these four cases, a range of values of M were examined. The models were estimated by the Whittle quasi-maximum likelihood procedure, based on the periodogram of the series. The following statistics are reported in the tables: the estimate of d; the value of p, selected by the Schwarz criterion; the largest AR root; and the residual Box Pierce and McLeod-Li statistics, computed with 25 lags in each case.

The 'true' values of the fractional integration parameter d should be $1 - \alpha/2 = 0.25$ in Tables 1 and 2 and $-\alpha/2 = -0.25$ in Tables 3 and 4. It is evident that there is a fairly constant asymptotic bias away from zero, in all these estimates, that varies only slightly with M. There is no obvious explanation for this, but it is plausibly related to the technical misspecification of the model.[5] Also note the much larger number of autoregressive terms needed to achieve an adequate representation in Tables 3 and 4. These latter models are fitted to the differences of the (nonstationary) generated series, and all the AR coefficients are negative, without exception.

In a sample of 20,000, the diagnostic statistics should have their asymptotic distributions when the null of independence is true — $\chi^2(25 - p)$ for the Box-Pierce and $\chi^2(25)$ for the McLeod-Li — and in the latter case, should reject with probability of, effectively, unity if the linear approximation is inadequate. The 5% critical value for the $\chi^2(25)$ is 37, and on this basis, the evidence indicates that quite a low value of M is adequate for the linear approximation, in three out of the four cases considered. As has been pointed out in Section 5, in the nonstationary case the noise will tend to dominate the regimes processes in the aggregate unless it is an order of magnitude smaller in variance. The results of Table 4 might be accounted for by this fact, although the estimates of d are, interestingly, quite similar to those obtained in Table 3 with much less noise. The poor approximations evident in Table 3 are explained by the fact, also pointed out in Section 5, that M needs to be greater than $O(T)$ in magnitude to ensure a good approximation, while there is no such limitation in the stationary case.

8 Concluding Remarks

In this paper we have re-considered the problem of distinguishing the phenomenon of fractional integration from classes of nonlinear long memory process. We emphasize that the correlation structure of the process is not the only relevant information contained in the data, and draw attention to the distinction between processes that have a linear representation and those that

[5] A couple of the samples have been re-estimated by conditional least squares, with nearly identical results.

do not. Linear long memory processes (or their partial sums, in the stationary case) converge to fractional Brownian motion under quite general conditions, but we show that a nonlinear process may have a non-Gaussian limit. Processes generated by cross-sectional aggregation may be linearised by virtue of their Gaussianity. Simulation experiments show that quite a modest degree of aggregation may be sufficient for a good linear approximation.

One simplification of the analysis has been to allow only purely nondeterministic processes with zero-mean increments, and we conclude by commenting briefly on the relaxation of this assumption. Note first that in the common components model, there is no difficulty about letting the process K_t have a non-zero mean. Assumption 2 must hold for the mean deviations in this case. Without going into details, we simply observe that the partial sum process for the case $1 < \alpha < 2$ should converge to a fBM with deterministic drift. However, inducing a drift in the independent aggegation model is more problematic. A constant $\mu^{(i)}$ may added to the regimes process, so that $E(X_t^{(i)}) = \mu^{(i)}$, but these 'micro-means' need to be 'small' on average, to avoid the mean of the aggregate process diverging under the normalization appropriate to the CLT. That is, we should need

$$M^{-1/2}\sum_{i=1}^{M}\mu^{(i)} \to \mu, \quad |\mu| < \infty.$$

Subject to this somewhat artificial requirement, however, a deterministic drift in the aggregate partial-sum process is a possibility.

9 Proofs

9.1 Proof of Theorem 2.1

By Assumption 1(e),

$$\gamma_h = E(m_t m_{t+h}) + E(\varepsilon_t \varepsilon_{t+h}).$$

We show that the first term satisfies the stated power law. The theorem will then follow because the second term forms a summable sequence, also by Assumption 1(e).

First, write

$$m_t = \sum_{j=-\infty}^{\infty} k_j 1_{(S_{j-1},S_j]}(t). \tag{9.1}$$

Note that by Assumptions 1(c) and 1(d), and the law of iterated expectations,

$$\begin{aligned} E(m_t m_{t+h}) &= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} E[k_i k_j 1_{(S_{i-1},S_i]}(t) 1_{(S_{j-1},S_j]}(t+h)] \\ &= \sum_{i=J(t)}^{\infty} E[k_{J(t)} k_i 1_{(S_{i-1},S_i]}(t+h)] \\ &= \sum_{i=J(t)}^{\infty} E[1_{(S_{i-1},S_i]}(t+h) E(k_{J(t)} k_i | \mathcal{T})] \\ &\in [B, B^{-1}] E^*(m_t m_{t+h}) \end{aligned} \tag{9.2}$$

where $J(t)$ is defined following (2.3), the notation $x \in [B, B^{-1}]y$ denotes that $By \leq x \leq B^{-1}y$ and we also define

$$E^*(m_t m_{t+h}) = \sum_{i=J(t)}^{\infty} E(k_{J(t)} k_i) P(S_{i-1} < t+h \leq S_i) \geq 0. \tag{9.3}$$

The second equality of (9.2) uses the fact that $1_{(S_{J(t)-1},S_{J(t)}]}(t) = 1$ by construction.

For the leading term of (9.2) $i = J(t)$, we show that

$$\begin{aligned} P(t+h \le S_{J(t)}) &= \sum_{c=1}^{\infty} \max\{0,(1-h/c)\} \frac{c}{E(\tau_0)} P(\tau_0 = c) \\ &\simeq \sum_{c=h+1}^{\infty} (1-h/c) c^{-\alpha} L(c) \\ &\simeq h^{1-\alpha} L(h) \end{aligned} \tag{9.4}$$

where $E(\tau_0)$ is defined in (2.3). The equality here combines (2.4) with the fact that (by stationarity) the position of t in regime $J(t)$ is uniformly distributed with probabilities $1/\tau_{J(t)}$, and hence

$$P(t+h \le S_{J(t)} | S_{J(t)} - S_{J(t)-1} = c) = \max\{0,(1-h/c)\}.$$

Note that for every $\varepsilon > 0$, some $\beta > 0$, and h large enough,

$$h^{1-\alpha-\varepsilon} \simeq \sum_{c=h+1}^{\infty} c^{-\alpha-\varepsilon} < \sum_{c=h+1}^{\infty} c^{-\alpha} L(c) < \sum_{c=h+1}^{\infty} c^{-\alpha} \log^{\beta} c \simeq h^{1-\alpha} \log^{\beta} h. \tag{9.5}$$

It follows that there exists a slowly varying function $L(h)$, satisfying Assumption 1(b), such that the equivalence in (9.4) holds.

Next, consider the case $i = J(t) + 1$ in (9.2). Note that

$$P(S_{J(t)} < t+h \le S_{J(t)+1}) = P(t+h \le S_{J(t)+1}) - P(t+h \le S_{J(t)}) \tag{9.6}$$

where by analogy with (9.4),

$$P(t+h \le S_{J(t)+1}) = \sum_{c=h+1}^{\infty} \max\{0,(1-h/c)\} \frac{c}{2E(\tau_0)} P(\tau_0 + \tau_1 = c) \tag{9.7}$$

and

$$\begin{aligned} P(\tau_0+\tau_1 = c) &= \sum_{j=1}^{c-1} P(\tau_1 = c - \tau_0 | \tau_0 = j) P(\tau_0 = j) \\ &\simeq \sum_{j=1}^{c-1} (c-j)^{-1-\alpha} j^{-1-\alpha} L(c-j) L(j) \\ &\simeq c^{-1-\alpha} L(c) \end{aligned} \tag{9.8}$$

using standard summability arguments (see e.g. Davidson and de Jong (2000) Lemma A.1). Hence, substituting into (9.7) yields similarly to (9.4)

$$P(t+h \le S_{J(t)+1}) \simeq h^{1-\alpha} L(h).$$

In other words, the two terms on the right-hand side of (9.6) have the same order of magnitude, so that their difference has this order of magnitude at most. The same argument can be applied, recursively, for each $i = 2, 3, \ldots$. It follows that

$$E^*(m_t m_{t+h}) \simeq h^{1-\alpha} L(h).$$

and the same property extends to $E(m_t m_{t+h})$, by assumption.

9.2 Proof of Theorem 2.2

The required probability is given, from (2.5) and (2.6), by

$$P(\tau_j > c) = \alpha \int_1^\infty \left(\frac{\mu}{\mu+1}\right)^{c-1} \mu^{-1-\alpha} d\mu. \tag{9.9}$$

Note that for $\mu > c-1$,

$$L_c \leq \left(\frac{\mu}{\mu+1}\right)^{c-1} \leq U_c \tag{9.10}$$

where L_c and U_c can be made arbitrarily close to e^{-1} and 1, respectively, by taking c large enough. Also, simple calculus shows that

$$\begin{aligned}\max_{1\leq\mu<\infty} \left(\frac{\mu}{\mu+1}\right)^{c-1} \mu^{-1-\alpha} &= \left(1 - \frac{1+\alpha}{c-1}\right)^{c-1} \left(\frac{c-1}{1+\alpha} - 1\right)^{-1-\alpha} \\ &\approx e^{-1-\alpha}(1+\alpha)^{1+\alpha}c^{-1-\alpha}\end{aligned}$$

where the approximation improves as c increases. Therefore, defining

$$A_1(c) = \alpha \int_1^c \left(\frac{\mu}{\mu+1}\right)^{c-1} \mu^{-1-\alpha} d\mu$$

$$A_2(c) = \alpha \int_c^\infty \left(\frac{\mu}{\mu+1}\right)^{c-1} \mu^{-1-\alpha} d\mu$$

such that $P(\tau_j \geq c) = A_1 + A_2$, note that for c large enough,

$$A_1(c) \leq e^{-1-\alpha}(1+\alpha)^{1+\alpha}c^{-\alpha}$$

and also, using (9.10) with c large enough,

$$e^{-1}c^{-\alpha} \leq A_2(c) \leq c^{-\alpha}.$$

Hence,

$$e^{-1} \leq c^\alpha[A_1(c) + A_2(c)] \leq 1 + e^{-1-\alpha}(1+\alpha)^{1+\alpha}$$

uniformly in c. We can conclude that $c^\alpha[A_1(c) + A_2(c)] \to C$ for some constant C in the specified interval, and the theorem follows.

9.3 Proof of Theorem 3.1

$$\begin{aligned}\gamma_h &= E\Bigg(\sum_{r=1}^\infty P(r)(K_{t-r} - \mu) + E_t\Bigg)\Bigg(\sum_{r=1}^\infty P(r)(K_{t+h-r} - \mu) + E_{t+h}\Bigg) \\ &= E(E_0E_h) + E(K_0^2)\sum_{v=1}^\infty P(v)P(h+v) \\ &+ \sum_{u=1}^\infty E(K_0K_u)\Bigg(\sum_{v=\max(1,1+u-h)}^\infty P(v)P(h+v-u) + \sum_{v=1}^\infty P(v)P(h+v+u)\Bigg)\end{aligned} \tag{9.11}$$

These terms are all positive on the assumptions, and summability of $\{E(K_0K_u), u \geq 0\}$ means that

$$\gamma_h = O\Bigg(\sum_{v=h}^\infty P(v)\Bigg) = O(h^{1-\alpha}L(h)).$$

9.4 Proof of Theorem 4.2

From Theorem 5.3 we can deduce that for $\delta > 0$,

$$E(Z_T^M(\xi+\delta) - Z_T^M(\xi))^2 \to \delta^{2H} \tag{9.12}$$

as $M, T \to \infty$ (sequentially), which is the covariance structure of fractional Brownian motion. In view of the Gaussianity of the finite dimensional distributions already established under the limit with respect to M, it remains only to establish the tightness of the sequence of measures with respect to T. For example, Theorem 15.6 of Billingsley (1968), cites a sufficient condition of the form

$$E|F_T(t_1 - t)|^\gamma |F_T(t - t_2)|^\gamma \leq |t_2 - t_1|^{2H} \tag{9.13}$$

for all $t_1 \leq t \leq t_2$, all $0 \leq t_1 < t_2 \leq 1$, $\gamma \geq 0$ and $H > \frac{1}{2}$, plus right-continuity at $t = 1$ with probability 1. It easy to show that a process satisfying (9.12) also satisfies (9.13) with $\gamma = 1$.

9.5 Proof of Theorem 4.3

Assume without loss of generality that $S_0 = 0$. Since τ_j is the duration of regime j,

$$\begin{aligned}
[T^{1/\alpha}L(T)]^{-1}\sum_{t=1}^{[T\xi]} X_t &= [T^{1/\alpha}L(T)]^{-1}\Bigg(\sum_{j=1}^{J([T\xi])-1} k_j\tau_j \\
&\qquad + k_{J([T\xi])}([T\xi] - S_{J([T\xi])-1}) + \sum_{t=1}^{[T\xi]}\varepsilon_t\Bigg) \\
&= \sigma_k[J(T)^{1/\alpha}L(J(T))]^{-1}\sum_{j=1}^{J([T\xi])-1} U_j + o_p(1)
\end{aligned}$$

where $J(T)$ is defined following (2.3) and

$$U_j = E(\tau_j)^{-1/\alpha}\frac{k_j\tau_j}{\sigma_k}$$

noting that, since $T = \sum_{j=1}^{J(T)}\tau_j$ where τ_j is an i.i.d. and integrable random variable,

$$\frac{J(T)^{1/\alpha}L(T)}{T^{1/\alpha}L(J(T))} \xrightarrow{pr} E(\tau_j)^{-1/\alpha}.$$

Further note that U_j is an i.i.d., zero-mean random variable. By Assumption 3(b)

$$P(|k_j|\tau_j > c) = \int_{\{\tau_j > c\}} P(\sigma_k^{-1}|k_j| > c/\tau_j|\tau_j)dF(\tau_j) + o(c^{-\alpha}).$$

Also note that

$$C_1P(\tau_j > c) \leq \int_{\{\tau_j > c\}} P(\sigma_k^{-1}|k_j| > c/\tau_j|\tau_j)dF(\tau_j) \leq P(\tau_j > c)$$

where C_1 is an almost sure lower bound of $P(\sigma_k^{-1}|k_j| > 1|\tau_j)$, and $C_1 > 0$ since a random variable with unit variance must have positive probability mass above 1. Hence

$$P(|U_j| > c) \simeq P(\tau_j > c) \simeq c^{-\alpha}L(c).$$

Let F_U denote the c.d.f. of U_j. Since $E(k_j) = 0$ and $\tau_j > 0$, both tails of the distribution obey the power law such that

$$1 - F_U(c) \simeq c^{-\alpha} L(\alpha), \qquad F_U(-c) \simeq c^{-\alpha} L(\alpha).$$

Thus, we have

$$\frac{1 - F_U(\xi c)}{1 - F_U(c)} \to \xi^{-\alpha}, \qquad \frac{F_U(-\xi c)}{F_U(-c)} \to \xi^{-\alpha}.$$

According to (e.g.) Theorem 9.34 of Breiman (1968), this condition is necessary and sufficient for F_U to lie in the domain of attraction of a stable law with parameter α. In other words,

$$a_{J(T)}^{-1} \left(\sum_{j=1}^{J([T\xi])-1} U_j - b_{J(T)} \right) \xrightarrow{d} \Lambda_\alpha(\xi)$$

where

$$nP(U_j > a_n c) \to c^{-\alpha} \quad \text{as } n \to \infty \tag{9.14}$$

and $b_T \to 0$. Note that setting $a_n = n^{1/\alpha} L(n)$ for a suitably chosen slowly varying function L solves (9.14) (see Davis (1983), or e.g. Feller (1966) Sections 9.6 and 17.5). Finally, the theorem follows by application of (e.g.) Embrechts et al. (1997) Theorem 2.4.10.

9.6 Proof of Theorem 5.1

In the following let T be finite, but large enough to allow consideration of $h < T$ as large as desired. Defining the $\mathcal{T}$-measurable random variable $Q(t,u) = \sum_{s=0}^{u-1} \tau_{J(t)+s}$, note that

$$P(Q(t,u) = h) \simeq h^{-1-\alpha} L(h)$$

for any t and $u > 0$, following from arguments in the proof of Theorem 2.1. Next note that

$$\Delta m_{Tt} \Delta m_{T,t+h} = \begin{cases} E_T(\tau_0) \Delta k_{J(t)}^2 & t = S_{J(t)-1} + 1, \; h = 0 \\ E_T(\tau_0) \Delta k_{J(t)} \Delta k_{J(t)+u}, & t = S_{J(t)-1} + 1, \; h = Q(t,u) \\ 0, & \text{otherwise.} \end{cases}$$

Applying arguments similar to those in the proof of Theorem 2.1, it follows using Assumption 1(d) that

$$E(\Delta m_{Tt}^2) \in [B, B^{-1}] E(\Delta k_0^2) \tag{9.15}$$

and

$$\begin{aligned} E(\Delta m_{Tt} \Delta m_{T,t+h}) &\in [B, B^{-1}] \sum_{u=1}^{\infty} E(\Delta k_0 \Delta k_u) P(Q(0,u) = h) \\ &\simeq h^{-1-\alpha} L(h). \end{aligned} \tag{9.16}$$

from Assumption 4(b). It follows from the fact that $\{\tau_j\}$ has a stationary distribution that the sequence defined by (9.15) and (9.16) is independent of t and T.

Therefore, consider

$$\gamma_h^\Delta = E(\Delta m_{Tt} \Delta m_{Tt+h}) + E(\Delta \varepsilon_t \Delta \varepsilon_{t+h}) \tag{9.17}$$

where the cross-products vanish by Assumption 4(e). The assumption further ensuresthe second right-hand side term is of smaller order in h than the first one. Under stationarity and Assumption 4,

$$\left.\begin{array}{l} E(\Delta k_0 \Delta k_u) = 2[E(k_0 k_u) - E(k_0 k_{u-1})] \\ E(\Delta\varepsilon_0 \Delta\varepsilon_u) = 2[E(\varepsilon_0\varepsilon_u) - E(\varepsilon_0\varepsilon_{u-1})] \end{array}\right\} \left\{\begin{array}{ll} < 0, & u = 1 \\ \leq 0, & u \geq 2. \end{array}\right. \tag{9.18}$$

These results therefore show that $\gamma_h^\Delta < 0$ for $h > 0$, and also that $\gamma_h^\Delta \simeq h^{-1-\alpha}L(h)$. Next, note that for any covariance stationary random sequence x_t,

$$\begin{aligned} E(\Delta x_t^2) + 2E(\Delta x_t \Delta x_{t-1}) + \cdots + 2E(\Delta x_t \Delta x_{t-h}) &= E(x_t - x_{t-1})(x_t + x_{t-1} - 2x_{t-h-1}) \\ &= 2[E(x_t x_{t+h}) - E(x_t x_{t+h+1})]. \end{aligned} \tag{9.19}$$

Therefore, in view of (9.15), (9.16) and the assumptions, and since (9.19) also applies to ε_t, we conclude from (9.17) and (9.18) that

$$\gamma_0^\Delta + 2\sum_{h=1}^{s} \gamma_h^\Delta = O(s^{-\alpha}L(s)).$$

9.7 Proof of Theorem 5.2

Note first that we can write

$$\Delta F_t = K_{t-1} + \sum_{r=1}^{t-1} \xi_{tr} K_{t-r-1} + \Delta E_t$$

where

$$\begin{aligned} -\xi_{tr} &= \frac{P_{t-1}(r)}{P_{t-1}(1)} - \frac{P_t(r+1)}{P_t(1)} \\ &= \frac{P(\tau_0 = r)}{\sum_{c=1}^{t} P(\tau_0 = c)} - \frac{P(\tau_0 = t)\sum_{c=1}^{r-1} P(\tau_0 = c)}{\sum_{c=1}^{t-1} P(\tau_0 = c)\sum_{c=1}^{t} P(\tau_0 = c)} \\ &= P(\tau_0 = r) + O(t^{-\alpha}L(r)). \end{aligned}$$

Letting $\Delta F_t^* = K_{t-1} + \sum_{r=1}^{\infty} P(\tau_0 = r)K_{t-r-1} + \Delta E_t$, note that ΔF_t^* is stationary and $E(\Delta F_t - \Delta F_t^*)^2 = O(t^{-\alpha})$, so we do the calculations for ΔF_t^*. To fix ideas, consider initially the case of K_t serially uncorrelated. In this case, substituting into the formula in (9.11), replacing $P(1)$ by 1 and $P(v+1)$ by $-P(\tau_0 = v)$ for $1 \leq v < t$, and 0 otherwise, yields

$$\begin{aligned} \gamma_0^\Delta &= E(\Delta E_0^2) + E(K_0^2)\Big(1 + \sum_{v=1}^{\infty} P(\tau_0 = v)^2\Big) \\ \gamma_h^\Delta &= E(\Delta E_0 \Delta E_h) + E(K_0^2)\Big(-P(\tau_0 = h) + \sum_{v=1}^{\infty} P(\tau_0 = v)P(\tau_0 = h + v)\Big), \quad h \geq 1. \end{aligned}$$

Verify first that these terms are $O(h^{-1-\alpha})$, and second that they are negative for $h \geq 1$, which in respect of $E(\Delta E_0 \Delta E_h)$ follows from (9.18) and Assumption 2(c). Third, observe that

$$1 + \sum_{v=1}^{\infty} P(\tau_0 = v)^2 + 2\sum_{h=1}^{\infty}\Big(-P(\tau_0 = h) + \sum_{v=1}^{\infty} P(\tau_0 = v)P(\tau_0 = h + v)\Big).$$

$$= \left(\sum_{v=t+1}^{\infty} P(\tau_0 = v) \right)^2 = 0.$$

These results, together with (9.19) in respect of the terms in $E(\Delta E_0 \Delta E_h)$, imply, as required,

$$\gamma_0^{\Delta} + 2\sum_{h=1}^{\infty} \gamma_0^{\Delta} = 0.$$

To extend these results to the case of serially correlated K_t, consider (9.11) again, and note that for any sequence $P(v)$, and all $u > 0$,

$$\begin{aligned} &\sum_{v=\max(1,1+u)}^{\infty} P(v)P(v-u) + \sum_{v=1}^{\infty} P(v)P(v+u) \\ &+2\sum_{h=1}^{\infty}\left(\sum_{v=\max(1,1+u-h)}^{\infty} P(v)P(h+v-u) + \sum_{v=1}^{\infty} P(v)P(h+v+u) \right) \\ &= 2\left(\sum_{v=1}^{\infty} P(v)\right)^2. \end{aligned}$$

9.8 Proof of Theorem 5.3

Theorem 5.1 establishes that, for any fixed g, $\sum_{h=-\infty}^{\infty} \gamma_{|g-h|}^{\Delta} = 0$. Hence, $0 \geq \gamma_h^{\Delta} \simeq h^{2H-2}L(h)$ implies that

$$\sum_{h=-T}^{T} \gamma_{|g-h|}^{\Delta} = -\left(\sum_{h=-\infty}^{-T-1} + \sum_{h=T+1}^{\infty} \right) \gamma_{|g-h|}^{\Delta} \simeq \sum_{h=T+1}^{\infty} h^{2H-2}L(h) \simeq T^{2H-1}L(T)$$

where the final rate of convergence follows, under Assumption 4(b), by an argument analogous to (9.5).

9.9 Proof of Theorem 5.4

The linear structure with independent increments follows from Wold's Theorem and the Gaussianity, exactly as for the case $1 < \alpha < 2$, and it remains to establish the summation properties of the coefficients. Note that $F_T = \sum_{s=1}^{T} \Delta F_s = \sum_{t=-\infty}^{T} a_{Tt}\eta_t$ where

$$a_{Tt} = \begin{cases} \sum_{j=0}^{T-t} \theta_j, & t > 0 \\ \sum_{j=1-t}^{T-t} \theta_j, & t \leq 0. \end{cases} \tag{9.20}$$

Hence

$$E(F_T^2) = \sigma_\eta^2 \sum_{t=-\infty}^{T} a_{Tt}^2 = \sigma_\eta^2 \sum_{t=-\infty}^{0} \left(\sum_{j=1-t}^{T-t} \theta_j \right)^2 + \sigma_\eta^2 \sum_{t=1}^{T} \left(\sum_{j=0}^{T-t} \theta_j \right)^2. \tag{9.21}$$

However, we also know from Theorem 5.3 that

$$E(F_T^2) = O(T^{1-\alpha}L(T)).$$

Considering the second block of terms on the right-hand side of (9.21), it is clear we have a contradiction unless the sequence of squared sums is $o(1)$ as $T \to \infty$, for any fixed t.

9.10 Proof of Theorem 5.5

We can deduce (9.12) in this case from Theorem 5.3, and since the finite dimensional distributions are known it remains, as in the case $1 < \alpha < 2$, to establish the tightness. However, it can be seen that Billingsley's (1968) Theorem 15.6 will not serve in this case since $H < \frac{1}{2}$. However, Theorem 3.1 of Davidson and de Jong (2000) (henceforth DdJ) can be substituted (and in fact, provides an alternative proof for Theorem 4.2).

By Theorem 5.4, the process has a linear representation as $M \to \infty$ with T fixed. Specifically, adapting the notation of Lemma 3.1 of DdJ, let

$$a_{Tt}(\xi+\delta,\xi) = \begin{cases} -\sum_{j=[T(\xi+\delta)]+1-t}^{\infty} \theta_j, & t > [T\xi] \\ \sum_{j=[T\xi]+1-t}^{\infty} \theta_j - \sum_{j=[T(\xi+\delta)]+1-t}^{\infty} \theta_j, & t \le [T\xi] \end{cases}$$

such that a_{Tt} defined in (9.20) becomes $a_{Tt}(1,0)$. Then, holding T fixed we can write

$$Z_T^M(\xi+\delta) - Z_T^M(\xi) \xrightarrow{d} \sigma_T^{-1} \sum_{t=-\infty}^{[T(\xi+\delta)]} a_{Tt}(\xi+\delta,\xi)\eta_t \text{ as } M \to \infty.$$

In view of Theorem 5.3, we have shown that

$$\sigma_T^2 = \sigma_\eta^2 \sum_{t=-\infty}^{T} a_{Tt}(1,0)^2 = O(T^{2H}L(T))$$

and hence,

$$\sigma_T^{-2}\sigma_\eta^2 \sum_{t=-\infty}^{[T(\xi+\delta)]} a_{Tt}(\xi+\delta,\xi)^2 \to \delta^{2H} \text{ as } T \to \infty. \tag{9.22}$$

We have therefore established conditions sufficient for Theorem 3.1 of DdJ. This result uses the linearity of the fractionally integrated process to establish the uniform tightness, and the conditions are easily established because here the increment process $\{\eta_t\}$ is i.i.d., so that DdJ's Lemma 3.2 holds trivially. The properties required to be satisfied by the moving average coefficients are those leading to condition (B.36) of DdJ, which corresponds here to (9.22). This completes the proof.

9.11 Proof of Theorem 6.1

It follows from the properties of fBM (see e.g. Davidson and de Jong (2000) equations (2.8)-(2.9)) that for $0 \le \xi < 1$ and $0 < \delta < 1 - \xi$,

$$E(X(\xi)X(\xi+\delta)) = \tfrac{1}{2}\big[\xi^{2H} + (\xi+\delta)^{2H} - \delta^{2H}\big]. \tag{9.23}$$

Therefore, for $0 \le \eta \le 1 - \xi - \delta$,

$$\begin{aligned} &E(X(\xi+\delta) - X(\xi))(X(\xi+\eta+\delta) - X(\xi+\eta)) \\ &\qquad = \tfrac{1}{2}\big[(\eta+\delta)^{2H} - 2(\eta)^{2H} + (\eta-\delta)^{2H}\big] \end{aligned}$$

Putting $\delta = 1/2n$ and $\eta = h/2n$ for integer $h \ge 1$, and substituting from (6.3), we therefore have

$$\begin{aligned} E(x_{nt}x_{n,t+h}) &= (2n)^{2H} E(X((t+h+1)/2n) - X((t+h)/2n))(X((t+1)/2n) - X(t/2n)) \\ &= \tfrac{1}{2}\big[(1/2n + h/2n)^{2H} + (h/2n - 1/2n)^{2H} - 2(h/2n)^{2H}\big] \end{aligned}$$

$$
\begin{aligned}
&= \tfrac{1}{2}h^{2H}\left[(1+1/h)^{2H} + (1-1/h)^{2H} - 2\right] \\
&\approx (H - \tfrac{1}{2})h^{2H-2}
\end{aligned}
$$

where the approximation is obtained from Taylor's expansions to second order of the first two terms around 1, and improves as h increases.

References

Beran. J. (1994) Statistics for Long Memory Processes. New York: Chapman and Hall

Billingsley, Patrick (1968) Convergence of Probability Measures, John Wiley, New York.

Breiman, Leo (1968), Probability, Addison-Wesley, Reading, Mass.

Brock, W. A., W. D. Dechert, J. Scheinkman and B. LeBaron (1996) A test for independence based on the correlation dimension, Econometric Reviews 15(3), 197–235.

Byers, D., J. Davidson and D. Peel (1997) Modelling political popularity: an analysis of long range dependence in opinion poll series, Journal of the Royal Statistical Society Series A, 160 (3), 471–90.

Byers, D., J. Davidson and D. Peel (2000) The dynamics of aggregate political popularity: evidence from eight countries. Electoral Studies 19, 1, 49–62.

Davidson, J. (1994) Stochastic Limit Theory: An Introduction for Econometricians. Oxford: Oxford University Press.

Davidson, J. (2000) Econometric Theory, Oxford: Blackwell Publishers.

Davidson, J. (2002a) Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes, Journal of Econometrics 106 (2002) 243–269.

Davidson, J. (2002b) A model of fractional cointegration, and tests for cointegration using the bootstrap. Journal of Econometrics 110, 187–212.

Davidson, J. and R. M. de Jong (2000) The functional central limit theorem and weak convergence to stochastic integrals II: fractionally integrated processes. Econometric Theory 16, 5, 621–642.

Davis, R. A. (1983) Stable limits for partial sums of dependent random variables. Annals of Probability 11(2) 262-269.

Diebold, F. X. and A. Inoue (2001) Long memory and regime switching. Journal of Econometrics 105, 131 –159

Ding, Z. and C. W. J. Granger (1996) Modelling volatility persistence of speculative returns: a new approach. Journal of Econometrics 73, 185–215.

Embrechts, P., C. Kluppelberg and T. Mikosch (1997) Modelling Extremal Events, Springer Verlag

Feller, W. (1966) An Introduction to Probability Theory and its Applications Vol 2, Wiley, New York

Granger, C. W. J. (1980) Long memory relationships and the aggregation of dynamic models, Journal of Econometrics 14, 227–238

Granger, C.W.J. and R. Joyeux (1980) An introduction to long memory time series models and fractional differencing. Journal of Time Series Analysis 1, 1, 15-29.

Hosking, J. R. M. (1981) Fractional differencing. Biometrika 68,1, 165-76.

Liu, M. (2000) Modeling long memory in stock market volatility. Journal of Econometrics 99, 139-171.

Marinucci, D. and P. M. Robinson (1999) Alternative forms of fractional Brownian motion. Journal of Statistical Inference and Planning 80, 111–122.

McLeod, A. I. and Li, W. K. (1983) Diagnostic checking ARMA time series models using squared-residual autocorrelations, Journal of Time Series Analysis 4, pp. 269–273.

Mikosch, T., S. Resnick, H. Rootzén and A. Stegeman: (2002) Is network traffic approximated by stable Lévy motion or fractional Brownian motion? Ann. Appl. Probab. 12, 23-68

Parke, W. R., (1999) What is fractional integration? Review of Economics and Statistics 81, 632 –638.

Phillips, P. C. B. and H. R. Moon (1999) Linear regression limit theory for nonstationary panel data, Econometrica 67, 1057–1112.

Robinson, P. M. (1978) Statistical inference for a random coefficient autoregressive model, Scan. J. Statist. 5, 163–8.

Schwarz, G. (1978) Estimating the dimension of a model. Annals of Statistics 6, 461–4.

Taqqu, M. S., W. Willinger and R. Sherman (1997) Proof of a fundamental result in self-similar traffic modeling.Computer Communication Review 27, 5–23.

Wold, H (1938) A Study in the Analysis of Stationary Time Series. Uppsala: Almqvist and Wicksell

M	100	20	10	5	1
d	0.31	0.31	0.31	0.30	0.28
p	2	2	2	2	2
$\lambda_{\max}$	0.37	0.38	0.39	0.42	0.42
B-P(25)	28	15	18	16	11
M-L(25)	29	45	115	222	1027

Table 1: $\alpha = 1.5$, $\sigma_\epsilon^2 = 0.2$

M	100	20	10	5	1
d	0.30	0.29	0.28	0.28	0.21
p	2	2	2	2	3
$\lambda_{\max}$	0.31	0.32	0.35	0.37	0.49
B-P(25)	16	22	24	17	25
M-L(25)	20	36	38	64	994

Table 2: $\alpha = 1.5$, $\sigma_\epsilon^2 = 0.5$

M	1000	100	20	10	5
d	-0.33	-0.32	-0.34	-0.30	-0.33
p	4	5	4	4	4
$\lambda_{\max}$	0.50	0.53	0.45	0.43	0.44
B-P(25)	10	10	25	29	24
M-L(25)	137	428	1104	2240	2844

Table 3: $\alpha = 0.5$, $\sigma_\epsilon^2 = 0.2$

M	1000	100	20	10	5
d	-0.38	-0.39	-0.36	-0.41	-0.42
p	9	9	9	8	9
$\lambda_{\max}$	0.71	0.70	0.72	0.67	0.71
B-P(25)	39	10	59	16	51
M-L(25)	30	8	36	101	239

Table 4: $\alpha = 0.5$, $\sigma_\epsilon^2 = 0.5$