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

James Davidson[*] Philipp Sibbertsen[†]
Cardiff University Universität Dortmund

This version, September 2002

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. 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. 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. 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.

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. Arguably, such behaviour can always be accounted for, in a finite sample, by postulating a root close to but below unity. However, the reliance on asymptotic theory that lies at the heart of nonstationary analysis makes the 'small-sample' explanation of apparent persistence inherently unsatisfactory.

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. As detailed in Davidson (2002b), cointegration theory can be adapted straightforwardly to this set-up. 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', and so at most a finite number of autoregressive lags. Unless a unit root is involved, all these models exhibit exponentially short memory. On the other hand, 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 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) and Parke (1999) propose similar models, involving the aggregation of persistent shocks whose durations follow a power law distribution. Taqqu et al., in the context of modelling ethernet traffic, 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 fractional Brownian motion (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 fractional Brownian motion. 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 fractional 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 develops the properties of the aggregate process, showing that it has a linear representation and deriving an invariance principle. It is also shown that different limits arise without aggregation. Section 4 relates our findings to the cited literature on these models, and Section 5 reports some simulations of tests of linearity in an ARFIMA framework. Section 6 contains concluding remarks, and Section 7 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,[1] and

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

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.

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, and also features such as deterministic components, 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 $a_n \sim C b_n$, 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) $P(\tau_j = c) \simeq c^{-1-\alpha} L(c)$ *as* $c \to \infty$, $0 < \alpha < 2$, *where* $L(.)$ *is slowly varying at* ∞ *and* $\exists$ $\beta > 0$ *such that* $L(c)/\log^{\beta} c \to 0$. [2]

(c) $E(k_j) = 0$, $E(k_j^2) = \sigma_k^2 < \infty$, $E(k_j k_{j+s}) \ge 0$ *for* $s \ge 0$, *and* $\sum_{s=0}^{\infty} E(k_j k_{j+s}) < \infty$. *If* $0 < \alpha < 1$, *then in addition,* $E(k_j k_{j+s}) \ge E(k_j k_{j+s+1})$ *for* $s \ge 0$.

(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* $s \ge 0$,

$$BE(k_j k_{j+s}) \le E(k_j k_{j+s} | \mathcal{T}) \le B^{-1} E(k_j k_{j+s}) \ a.s. \tag{2.2}$$

---

[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.

[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 $L^2 = L$.

If $0 < \alpha < 1$ then in addition, for $s \geq 0$,

$$B[E(k_j k_{j+s}) - E(k_j k_{j+s+1})] \leq E(k_j k_{j+s}|\mathcal{T}) - E(k_j k_{j+s+1}|\mathcal{T})$$
$$\leq B^{-1}[E(k_j k_{j+s}) - E(k_j k_{j+s+1})] \ a.s. \qquad (2.3)$$

(e) $\{\varepsilon_t\}_{-\infty}^{\infty}$ *is strictly stationary with* $E(\varepsilon_t) = 0$ *and* $E(\varepsilon_t^2) = \sigma_\varepsilon^2$, $\sum_{s=0}^{\infty} E(\varepsilon_t \varepsilon_{t+h}) < \infty$, *and* $E(m_t \varepsilon_{t+h}) = E(\varepsilon_t m_{t+h}) = 0$ *for all* $h \geq 0$. *If* $0 < \alpha < 1$, *then in addition,* $E(\varepsilon_j \varepsilon_{j+s}) \geq E(\varepsilon_j \varepsilon_{j+s+1})$ *for* $s \geq 0$.

Thus, note that (b) implies the key power law property

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

Note the important distinction between the cases $1 < \alpha < 2$ and $0 < \alpha \leq 1$. In the first case the expected duration of a regime is finite, given according to Assumption 1(b) by

$$E(\tau) = \sum_{c=1}^{\infty} cP(\tau = c) < \infty. \qquad (2.4)$$

However, this sum diverges in the second case and the process is nonstationary.

Assumption 1(c) controls the dependence of successive regimes in a fairly natural manner. All the restrictions hold if the regimes are 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. Note that given Assumption 1(c), (2.3) implies (2.2), but the reverse implication does not hold.

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, the process is covariance stationary (and hence strictly stationary) and long memory, in the case $1 < \alpha < 2$, and difference stationary (the differences strictly and covariance stationary) in the case $0 < \alpha < 1$.[3]

Theorem 2.1 *Let Assumption 1 hold.*

(i) *If* $1 < \alpha < 2$, *then letting* $\gamma_h = E(X_t X_{t+h})$, $\gamma_h > 0$ *for all* h *and* $\gamma_h \simeq h^{1-\alpha} L(h)$.

(ii) *If* $0 < \alpha < 1$, *then letting* $\gamma_h = E(\Delta X_t \Delta X_{t+h})$, $\gamma_h < 0$ *for* $h \geq 1$, $\gamma_h \simeq h^{-1-\alpha} L(h)$, *and* $\gamma_0 + 2 \sum_{h=1}^{\infty} \gamma_h = 0$.

---

[3]The case $\alpha = 1$ requires a different treatment from that developed here. Essentially, the sequence $\{\gamma_h\}$ is slowly varying in case (i) of the theorem and the partial sums slowly varying in case (ii). This case will not be treated explicitly.

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

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

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

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

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 The Aggregation Model: Representation and Invariance Principle

Our interest in these processes is their relationship with the phenomena of long memory and fractional integration. Cross-sectional aggregation is a crucial feature of the analysis, and our approach follows essentially that of Taqqu et. al. (1997). Suppose that stochastic regime processes govern the behaviour of agents in the economy at the micro level, and that what is observed is the aggregate of their activities. 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

$$E(F_t^M F_{t+h}^M) = \gamma_h \qquad 1 < \alpha < 2$$

5

$$E(\Delta F_t^M \Delta F_{t+h}^M) = \gamma_h \qquad 0 < \alpha < 1$$

follows from the independence, where γ_h is defined by Theorem 2.1, parts (i) and (ii) respectively. Let $\{F_t\}_{-\infty}^{\infty}$ denote the limiting random process as $M \to \infty$, defined by the relation

$$(F_{t_1}^M, \ldots, F_{t_K}^M) \xrightarrow{d} (F_{t_1}, \ldots, F_{t_K}) \qquad 1 < \alpha < 2 \tag{3.1}$$

$$(\Delta F_{t_1}^M, \ldots, \Delta F_{t_K}^M) \xrightarrow{d} (\Delta F_{t_1}, \ldots, \Delta F_{t_K}) \qquad 0 < \alpha < 1 \tag{3.2}$$

where $t_1, \ldots, t_K$ is any finite collection of time coordinates. Under the assumptions, note that the limits in (3.1) and (3.2) are 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 $\{\gamma_h > 0, h = 0, 1, 2, \ldots\}$, 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. We assume the distributions to be identical solely to avoid undue complications.

Let $H = (3 - \alpha)/2$ for $1 < \alpha < 2$ and $H = (1 - \alpha)/2$ for $0 < \alpha < 1$, corresponding to Hurst's coefficient. We next show that the aggregate processes, whether partial sums or 'partial sums of differences', have the variance characteristics associated with long memory increments. Define the partial sum variance

$$\sigma_T^2 = \sum_{g=1}^{T} \sum_{h=1}^{T} \gamma_{|g-h|} = \left\{ \begin{array}{ll} E\left(\sum_{t=1}^{T} F_t\right)^2, & 1 < \alpha < 2 \\ E(F_T^2), & 0 < \alpha < 1 \end{array} \right\}.$$

Theorem 3.1 $\sigma_T^2 = O(T^{2H} L(T))$.

The bounding cases of the stationary process, $\alpha = 2$ and $\alpha = 1$, correspond to $H = \frac{1}{2}$ and $H = 1$ respectively, whereas the bounding cases of the difference-stationary process, $\alpha = 0$ and $\alpha = 1$, correspond to $H = \frac{1}{2}$ and $H = 0$ respectively. In view of the stationarity, we can assume the existence of a finite positive constant

$$\sigma^2 = \lim_{T \to \infty} (T^{2H} L(T))^{-1} \sigma_T^2 \tag{3.3}$$

Our next result extends the often-quoted rationale for the ARMA class of processes, based on Wold's theorem, to the sphere of long memory models.

Theorem 3.2 *The limiting aggregate process $\{F_t, -\infty < t < \infty\}$ has a linear moving average representation*

$$F_t = \sum_{j=0}^{\infty} \theta_j \eta_{t-j} \qquad 1 < \alpha < 2 \tag{3.4}$$

$$\Delta F_t = \sum_{j=0}^{\infty} \theta_j \eta_{t-j} \qquad 0 < \alpha < 1 \tag{3.5}$$

where $\sum_{j=0}^{\infty} \theta_j^2 < \infty$, $\eta_t \sim NI(0, \sigma_\eta^2)$, $\sigma_\eta^2 = \gamma_0 / \sum_{j=0}^{\infty} \theta_j^2$, and $\sum_{j=0}^{\infty} \theta_j = 0$ in case (3.5).

In other words, the (by construction) strictly nonlinear process F_t, generated by aggregating independent stochastic regime processes, nonetheless has a strictly linear representation such that the shock sequence is i.i.d. 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

6

we show in Section 5, 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. To unify notation, write

$$Z_T^M(\xi) = \left\{ \begin{array}{ll} \sigma_T^{-1} \sum_{t=1}^{[T\xi]} F_t^M & 1 < \alpha < 2, \ H = (3 - \alpha)/2, \\ \sigma_T^{-1} F_{[T\xi]}^M & 0 < \alpha < 1, \ H = (1 - \alpha)/2, \end{array} \right\}, \ 0 \le \xi \le 1 \qquad (3.6)$$

where $[x]$ denotes the largest integer not exceeding x. The following result is obtained by adapting Theorem 3.1 of Davidson and de Jong (2000).

Theorem 3.3 $Z_T^M \xrightarrow{d} \sigma B_H$ as $M, T \to \infty$ (sequentially), where B_H denotes fractional Brownian motion[4] with parameter H.

The expression '$M, T \to \infty$ (sequentially)' means that M must be taken to the limit for each $t = 1, \dots, T$, 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.

For the case $1 < \alpha < 2$, Theorem 3.3 may also be obtained by a direct appeal to Billingsley (1968) Theorem 15.6 to establish the tightness, given that Gaussianity and the covariance structure are already given. However, this approach does not work for the case $0 < \alpha < 1$. Note that the case $\alpha = 1$ corresponds to the boundary-nonstationary case, analogous to $d = \frac{1}{2}$ in the fractional integration model and, as remarked previously, requires an alternative approach.

It may be the case (and this assertion is explored in the simulations reported in Section 5) that quite a low value of M is sufficient to yield an adequate linear approximation. However, the next result shows that, without some degree of cross-sectional aggregation, the invariance properties demonstrated are not obtained. In other words, the usual argument from time aggregation fails. We focus attention on the case $1 < \alpha < 2$, and introduce the following extra assumptions.

Assumption 2 (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 3.4 *For* $1 < \alpha < 2$, *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 2 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 (3.6) 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

---

[4] This is fractional Brownian motion of type I, as defined by Robinson and Marinucci (1999). See also Davidson and de Jong (2000) for details.

limit process have no variance, note how reversing the order of M and T in Theorem 3.3 cannot yield a Gaussian limit in this case.

Assumption 2(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 2(b) is a natural extension of Assumption 1(d), and ensures that k_j does not itself contribute to the tail behaviour of U_j 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.

A result corresponding to Theorem 3.4 is not expected in the nonstationary case, however. The time aggregation involved in forming the partial sums is an essential feature of this result, and the diverging variance does not appear to be a feature of the nonstationary regimes process. Rather, it appears that this process will exhibit regimes even in the limit. That is to say, as T increases, a regime must sooner or later arise that accounts for a segment of the realization of order T. We may conjecture that a nonstationary limit process does exist in this case, and may even have Gaussian features, but it would apparently not be a.s. continuous.

4 Discussion

The results of the last section demonstrate three main conclusions. First, that a fairly general class of nonlinear processes can exhibit the covariance structure associated with long memory. Second, that there exists a sub-class of nonlinear processes, characterised by 'sideways aggregation', that are observationally equivalent to fractionally integrated processes. Specifically, their normalised partial sums converge to fractional Brownian motion (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)

$$y_t = \sum_{s=-\infty}^{t} g_{s,t} \varepsilon_s \qquad (4.1)$$

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 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,

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

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

$$P(n_s = c) = \frac{c^{-\alpha-1}}{\zeta(1+\alpha)}. \tag{4.2}$$

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 (4.1) 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 3.3. Note that even with Gaussian shocks, a linear representation of the form (3.5) 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 'discretising' the fBM. Let $X(\xi)$, $0 \le \xi \le 1$, be fBM with Hurst parameter H and, for convenience, normalised such that $EX(1)^2 = 1$. Fix a finite integer $n \ge 1$ and consider the sequence

$$x_{nt} = (2n)^{2H}(X((t+1)/2n + \tfrac{1}{2}) - X(t/2n + \tfrac{1}{2})), \quad t = -n, \ldots, (n-1). \tag{4.3}$$

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

Theorem 4.1 $E(x_{nt}x_{n,t+h}) \simeq h^{2H-2}$.

Taking n large, let an approximate Wold decomposition, truncated at n lags, be applied to the sequence $x_{n1}, \ldots, x_{nn}$. 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 3.2, there always exists a natural approach to distinguishing processes having fBM as their weak limit from alternatives. This is by tests of linearity.

5 Testing Linearity

In this section we consider tests of linearity applied to models in the aggregated-regimes class. For a practical implementation, we have to limit consideration to a class of linear models having the right covariance structure in a finite parameterisation. The ARFIMA(p, d, q) class is the natural choice for this purpose. It is true that the class of models we would wish to include in the null hypothesis is larger than the ARFIMA class, but there are grounds for thinking that the ARFIMA class can approximate linear processes with the requisite properties pretty well. Recall that these processes have the structure $X_t = \sum_{j=0}^\infty b_j U_{t-j}$ where U_t represents a stable ARMA(p, q) process, and $b_j = \Gamma(d + j)/(\Gamma(d)\Gamma(j + 1)) \simeq j^{d-1}$. Since $U_t = \sum_{k=0}^\infty \phi_k \eta_{t-j-k}$ where ϕ_k decays geometrically and $\eta_t \sim iid(0, \sigma_\eta^2)$, we have the representation $X_t = \sum_{j=0}^\infty \theta_j \eta_{t-j}$ where

$$\theta_j = \sum_{k=0}^j \phi_k b_{j-k}.$$

The ϕ_k can be of either sign and can decay monotonically, sinusoidally and discontinuously, so a wide range of decay patterns of the θ_j coefficients can be approximated while preserving the property $\theta_j \simeq j^{d-1}$.

We report some simulations in which ARFIMA models are fitted to aggregated-regime processes. This is not a Monte Carlo study, since there is no special interest in determining the distribution of the estimators. The approximating 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.

The chosen diagnostic test for nonlinearity is the McLeod-Li (1983) portmanteau test, corresponding to the Box-Pierce statistic computed for the squared residuals. Correlation in the squares is in any case a natural dummy alternative hypothesis in a diagnostic test for nonlinearity, but these models in particular are likely to resemble GARCH processes. 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 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

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

$\sigma_\varepsilon^2 = 0.5$. 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. The poor approximation evident in Table 3, even with $M = 1000$, can be explained by the fact that with $\alpha = 0.5$, the occurrence of regimes running for hundreds and even thousands of periods is significant. While rare, these account for a significant proportion of the data points; see the remarks in the last paragraph of Section 4. For the averaging to succeed when the noise level is low, it is necessary that 'most' of the time periods coincide with a break, in one or more components of the aggregate. Regime changes are much more frequent in the case $\alpha = 1.5$. However, Table 4 shows that when the noise component is relatively large, the linear approximation can still be quite good, even in the nonstationary case.

6 Conclusion

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 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.

7 Proofs

7.1 Proof of Theorem 2.1(i)

By Assumption 1(e),
$$\gamma_h = E(m_t m_{t+h}) + E(\varepsilon_t \varepsilon_{t+h}).$$

---

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

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

$$m_t = \sum_{j=-\infty}^{\infty} k_j 1_{(S_{j-1}, S_j]}(t). \tag{7.1}$$

Letting $J(t) = \min\{j : t \leq S_j\}$, note that by Assumptions 1(c) and 1(d), and the law of iterated expectations,

$$
\begin{aligned}
E(m_t m_{t+h}) &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} E[k_i k_j 1_{(S_{i-1}, S_i]}(t) 1_{(S_{j-1}, S_j]}(t+h)] \\
&= \sum_{i=J(t)}^{\infty} E[k_{J(t)} k_i 1_{(S_{i-1}, S_i]}(t+h)] \\
&= \sum_{i=J(t)}^{\infty} E[1_{(S_{i-1}, S_i]}(t+h) E(k_{J(t)} k_i | \mathcal{T})] \\
&\in [B, B^{-1}] E^*(m_t m_{t+h})
\end{aligned} \tag{7.2}
$$

where the notation $x \in [B, B^{-1}]y$ denotes that $By \leq x \leq B^{-1}y$ and we also define

$$E^*(m_t m_{t+h}) = \sum_{i=J(t)}^{\infty} E(k_{J(t)} k_i) P(S_{i-1} < t+h \leq S_i) \geq 0. \tag{7.3}$$

The second equality of (7.2) uses the fact that $1_{(S_{J(t)-1}, S_{J(t)}]}(t) = 1$ by construction.

For the leading term of (7.2) $i = J(t)$, we show that

$$
\begin{aligned}
P(t+h \leq S_{J(t)}) &= \sum_{c=1}^{\infty} \max\{0, (1-h/c)\} \frac{c}{E(\tau)} P(\tau = c) \\
&\simeq \sum_{c=h+1}^{\infty} (1-h/c) c^{-\alpha} L(c) \\
&\simeq h^{1-\alpha} L(h)
\end{aligned} \tag{7.4}
$$

where $E(\tau)$ is defined in (2.4). The equality here uses the following two facts. First, by stationarity, the position of t in regime $J(t)$ is uniformly distributed with probabilities $1/\tau_{J(t)}$, and hence

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

Second,

$$P(S_{J(t)} - S_{J(t)-1} = c) = \frac{cP(\tau = c)}{E(\tau)} \tag{7.5}$$

where the left-hand side probability is to be thought of (given the definition of $J(t)$) as the probability of a randomly selected observation X_t falling in a regime of duration c. Note that for every $\varepsilon > 0$, some $\beta > 0$, and h large enough,

$$h^{1-\alpha-\varepsilon} \simeq \sum_{c=h+1}^{\infty} c^{-\alpha-\varepsilon} < \sum_{c=h+1}^{\infty} c^{-\alpha} L(c) < \sum_{c=h+1}^{\infty} c^{-\alpha} \log^\beta c \simeq h^{\alpha-1} \log^\beta h \tag{7.6}$$

13

where the equivalence of convergence rates can be shown by an integral approximation. It follows that there exists a slowly varying function $L(h)$, satisfying Assumption 1(b), such that the equivalence in (7.4) holds.

Next, consider the case $i = J(t) + 1$ in (7.2). Note that

$$P(S_{J(t)} < t + h \leq S_{J(t)+1}) = P(t + h \leq S_{J(t)+1}) - P(t + h \leq S_{J(t)}) \tag{7.7}$$

where by analogy with (7.4),

$$P(t + h \leq S_{J(t)+1}) = \sum_{c=h+1}^{\infty} \max\{0, (1 - h/c)\} \frac{c}{2E(\tau)} P(\tau_1 + \tau_2 = c) \tag{7.8}$$

and

$$
\begin{aligned}
P(\tau_1 + \tau_2 = c) &= \sum_{j=1}^{c-1} P(\tau_2 = c - \tau_1 | \tau_1 = j) P(\tau_1 = j) \\
&\simeq \sum_{j=1}^{c-1} (c - j)^{-1-\alpha} j^{-1-\alpha} L(c - j) L(j) \\
&\simeq c^{-1-\alpha} L(c) \tag{7.9}
\end{aligned}
$$

using standard summability arguments (see e.g. Davidson and de Jong (2000) Lemma A.1). Hence, substituting into (7.8) yields similarly to (7.4)

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

In other words, the two terms on the right-hand side of (7.7) 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. ∎

7.2 Proof of Theorem 2.1(ii)

From (7.1),

$$
\begin{aligned}
\Delta m_t &= \sum_{j=-\infty}^{\infty} k_j [1_{(S_{j-1}, S_j]}(t) - 1_{(S_{j-1}, S_j]}(t - 1)] \\
&= \begin{cases} \Delta k_{J(t)}, & t = S_{J(t)-1} + 1 \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
$$

where $\Delta k_{J(t)} = k_{J(t)} - k_{J(t)-1}$. In other words, the process is nonzero only when date t falls in regime $J(t)$ and $t - 1$ in regime $J(t) - 1$. Hence, defining the $\mathcal{T}$-measurable random variable $Q(t, i) = \sum_{s=0}^{i-1} \tau_{J(t)+s}$,

$$
\Delta m_t \Delta m_{t+h} = \begin{cases} \Delta k_{J(t)}^2, & h = 0 \\ \Delta k_{J(t)} \Delta k_{J(t)+i}, & h = Q(t, i) \\ 0, & \text{otherwise.} \end{cases}
$$

14

Therefore note that

$$E(\Delta m_t^2) = E(\Delta k_{J(t)}^2).$$

For the cases $h > 0$, note first that under stationarity and Assumption 1(c),

$$E(\Delta k_j \Delta k_{j+i}) = 2[E(k_j k_{j+i}) - E(k_j k_{j+i-1})] \begin{cases} < 0, & i = 1 \\ \leq 0, & i \geq 2. \end{cases} \quad (7.10)$$

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

$$E(\Delta m_t \Delta m_{t+h}) = \sum_{i=1}^{\infty} E[1_{\{Q(t,i)=h\}} E(\Delta k_{J(t)} \Delta k_{J(t)+i} | \mathcal{T})]$$

$$\in [B^{-1}, B] E^*(\Delta m_t \Delta m_{t+h}) \quad (7.11)$$

where

$$E^*(\Delta m_t \Delta m_{t+h}) = \sum_{i=1}^{\infty} E(\Delta k_{J(t)} \Delta k_{J(t)+i}) P(Q(t,i) = h) < 0. \quad (7.12)$$

Therefore, consider

$$\gamma_h = E(\Delta m_t \Delta m_{t+h}) + E(\Delta \varepsilon_t \Delta \varepsilon_{t+h}) \quad (7.13)$$

where the cross-products vanish by Assumption 1(e). The assumption further ensures, by analogy with (7.10), that the second right-hand side term is both negative and of smaller order than the first one. These results therefore show that $\gamma_h < 0$ for $h > 0$, and also that $\gamma_h \simeq h^{-1-\alpha} L(h)$ from Assumption 1(b).

Next, note that for any covariance stationary random sequence x_t,

$$E(\Delta x_t^2) + 2E(\Delta x_t \Delta x_{t-1}) + \cdots + 2E(\Delta x_t \Delta x_{t-h}) = E(x_t - x_{t-1})(x_t + x_{t-1} - 2x_{t-h-1})$$

$$= 2[E(x_t x_{t+h}) - E(x_t x_{t+h+1})]. \quad (7.14)$$

Therefore, in view of (7.11), (7.12) and the assumptions,

$$E(\Delta m_t^2) + 2\sum_{h=1}^{\infty} E(\Delta m_t \Delta m_{t+h}) = E(\Delta k_{J(t)}^2) + 2\sum_{i=1}^{\infty} \sum_{h=1}^{\infty} E[1_{\{Q(t,i)=h\}} E(\Delta k_{J(t)} \Delta k_{J(t)+i} | \mathcal{T})]$$

$$= E(\Delta k_{J(t)}^2) + 2\sum_{i=1}^{\infty} E(\Delta k_{J(t)} \Delta k_{J(t)+i})$$

$$= 0. \quad (7.15)$$

Note that the second equality in (7.15) holds by linearity of the expectation, so that

$$\sum_{h=1}^{\infty} E[1_{\{Q(t,i)=h\}} E(\Delta k_{J(t)} \Delta k_{J(t)+i} | \mathcal{T})] = E\left[\sum_{h=1}^{\infty} 1_{\{Q(t,i)=h\}} E(\Delta k_{J(t)} \Delta k_{J(t)+i} | \mathcal{T})\right]$$

$$= E[E(\Delta k_{J(t)} \Delta k_{J(t)+i} | \mathcal{T})]$$

$$= E(\Delta k_{J(t)} \Delta k_{J(t)+i}).$$

Since (7.14) also applies to ε_t, we conclude from (7.13) that

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

which completes the proof. ∎

7.3 Proof of Theorem 2.2

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

$$P(\tau_j > c) = \alpha \int_1^\infty \left(\frac{\mu}{\mu+1}\right)^{c-1} \mu^{-1-\alpha} d\mu. \tag{7.16}$$

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

$$L_c \leq \left(\frac{\mu}{\mu+1}\right)^{c-1} \leq U_c \tag{7.17}$$

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 (7.17) 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. ∎

7.4 Proof. of Theorem 3.1

In the case $1 < \alpha < 2$, note from Theorem 2.1(i) that the sequence $\{\gamma_h\}$ is positive and monotone, and $\gamma_h \simeq h^{2H-2}L(h)$. It follows directly that $\sum_{h=1}^T \gamma_{|g-s|} = O(T^{2H-1}L(T))$.

In the case $0 < \alpha < 1$, Theorem 2.1(ii) establishes that $\sum_{h=1}^\infty \gamma_{|g-s|} = 0$. Hence, $0 \geq \gamma_h \simeq h^{2H-2}L(h)$ implies that

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

where the final rate of convergence follows, under Assumption 1(b), by an argument analogous to (7.6). ∎

16

7.5 Proof of Theorem 3.2

The sequences $\{F_t\}$ $(1 < \alpha < 2)$, and $\{\Delta F_t\}$ $(0 < \alpha < 1)$, respectively, are strictly stationary with finite variance γ_0, and purely nondeterministic, by construction. The Wold decomposition theorem (see e.g. Davidson (2000) Theorem 5.2.1) therefore implies the forms (3.4) and (3.5) respectively, where the sequence $\{\eta_t\}$ is stationary and uncorrelated with variance

$$\sigma_\eta^2 = \frac{\gamma_0}{\sum_{j=0}^\infty \theta_j^2}.$$

Further, $\{F_t\}$ $(1 < \alpha < 2)$ and $\{\Delta F_t\}$ $(0 < \alpha < 1)$, respectively, are Gaussian by the Lindeberg Levy central limit theorem, as noted in the text. 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. This completes the proof for the case $1 < \alpha < 2$.

Considering the case $0 < \alpha < 1$, note that

$$F_T = \sum_{s=1}^T \Delta F_s = \sum_{t=-\infty}^T a_{Tt} \eta_t$$

where

$$a_{Tt} = \begin{cases} \sum_{j=0}^{T-t} \theta_j, & t > 0 \\ \sum_{j=1-t}^{T-t} \theta_j, & t \le 0. \end{cases} \tag{7.18}$$

Hence

$$E(F_T^2) = \sigma_\eta^2 \sum_{t=-\infty}^T a_{Tt}^2 = \sigma_\eta^2 \sum_{t=-\infty}^0 \left(\sum_{j=1-t}^{T-t} \theta_j\right)^2 + \sigma_\eta^2 \sum_{t=1}^T \left(\sum_{j=0}^{T-t} \theta_j\right)^2. \tag{7.19}$$

However, we also know from Theorem 3.1 that

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

Considering the second block of terms on the right-hand side of (7.19), it is clear we have a contradiction unless the sequence of squared sums is $o(1)$ as $T \to \infty$, for any fixed t. ∎

7.6 Proof of Theorem 3.3

From Theorem 3.1 we can deduce that for $\delta > 0$,

$$E(Z_T^M(\xi + \delta) - Z_T^M(\xi))^2 \to \delta^{2H} \tag{7.20}$$

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.

By Theorem 3.2, the process has a linear representation as $M \to \infty$ with T fixed. Specifically, adapting the notation of Lemma 3.1 of Davidson and de Jong (2000) (henceforth DdJ), generalise (7.18) by defining

$$a_{Tt}(\xi + \delta, \xi) = \begin{cases} \begin{cases} \sum_{j=0}^{[T(\xi+\delta)]-t} \theta_j, & t > [T\xi] \\ \sum_{j=[T\xi]+1-t}^{[T(\xi+\delta)]-t} \theta_j, & t \le [T\xi] \end{cases}, & 1 < \alpha < 2 \\[3em] \begin{cases} -\sum_{j=[T(\xi+\delta)]+1-t}^\infty \theta_j, & t > [T\xi] \\ \sum_{j=[T\xi]+1-t}^\infty \theta_j - \sum_{j=[T(\xi+\delta)]+1-t}^\infty \theta_j, & t \le [T\xi] \end{cases}, & 0 < \alpha < 1 \end{cases}$$

17

such that a_{Tt} previously defined 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 3.1, 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,

$$\sigma_T^{-2}\sigma_\eta^2 \sum_{t=-\infty}^{[T(\xi+\delta)]} a_{Tt}(\xi + \delta, \xi)^2 \to \delta^{2H} \text{ as } T \to \infty. \tag{7.21}$$

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 (7.21). This completes the proof. ∎

7.7   Proof of Theorem 3.4

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

$$[T^{1/\alpha}L(T)]^{-1}\sum_{t=1}^{[T\xi]} X_t = [T^{1/\alpha}L(T)]^{-1}\left(\sum_{j=1}^{J([T\xi])-1} k_j\tau_j\right.$$

$$\left. + k_{J([T\xi])}([T\xi] - S_{J([T\xi])-1}) + \sum_{t=1}^{[T\xi]} \varepsilon_t\right)$$

$$= \sigma_k[J(T)^{1/\alpha}L(J(T))]^{-1}\sum_{j=1}^{J([T\xi])-1} U_j + o_p(1)$$

where $J(T)$ is defined following (7.1), 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 2(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

$$C_1 P(\tau_j > c) \leq \int_{\{\tau_j > c\}} P(\sigma_k^{-1}|k_j| > c/\tau_j|\tau_j)dF(\tau_j) \leq P(\tau_j > c)$$

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

$$nP(U_j > a_n c) \to c^{-\alpha} \quad \text{as } n \to \infty \tag{7.22}$$

and $b_T \to 0$. Note that setting $a_n = n^{1/\alpha}L(n)$ for a suitably chosen slowly varying function L solves (7.22) (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. ∎

7.8 Proof of Theorem 4.1

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

$$E(X(\xi)X(\xi + \delta)) = \tfrac{1}{2}\big[\xi^{2H} + (\xi + \delta)^{2H} - \delta^{2H}\big]. \tag{7.23}$$

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

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

Putting $\delta = 1/2n$ and $\eta = h/2n$ for integer $h \geq 1$, and substituting from (4.3), we therefore have

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

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

19

References

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

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

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. (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. Forthcoming in *Journal of Econometrics* Annals: Proceedings of the Cardiff Conference on Long Memory and Nonlinear Time Series, 2000, and at http://www.cf.ac.uk/carbs/davidsonje/bootfrac2.pdf.

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

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.

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

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

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.

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.