Journal of Econometrics 105 (2001) 131–159

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Long memory and regime switching

Francis X. Diebold[a,*], Atsushi Inoue[b]

[a]Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297, USA

[b]Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC 27695-8109, USA

Abstract

The theoretical and empirical econometric literatures on long memory and regime switching have evolved largely independently, as the phenomena appear distinct. We argue, in contrast, that they are intimately related, and we substantiate our claim in several environments, including a simple mixture model, Engle and Smith's (Rev. Econom. Statist. 81 (1999) 553–574) stochastic permanent break model, and Hamilton's (Econometrica 57 (1989) 357–384) Markov-switching model. In particular, we show analytically that stochastic regime switching is easily confused with long memory, even asymptotically, so long as only a "small" amount of regime switching occurs, in a sense that we make precise. A Monte Carlo analysis supports the relevance of the theory and produces additional insights.

JEL classification: C12; C13; C22; C50

Keywords: Regime switching; Fractional integration; Mixture model; Stochastic permanent break (STOPBREAK) model; Markov-switching model; Structural change

1. Introduction

Motivated by early empirical work in macroeconomics (e.g., Diebold and Rudebusch, 1989) and later empirical work in finance (e.g., Ding et al., 1993),

* Corresponding author. Tel.: +1-215-898-1507; fax: +1-215-573-4217.
E-mail address: fdiebold@sas.upenn.edu (F.X. Diebold).

0304-4076/01/$ - see front matter
PII: S0304-4076(01)00073-2

recent years have witnessed a renaissance in both the theoretical and empirical econometrics of long memory and fractional integration. Recent contributions include, among many others, Ding et al. (1993), Robinson (1995), Comte and Renault (1998), and Andersen et al. (2001a, b).

The fractional unit root boom of the 1990s was preceded by the integer unit root boom of the 1980s. In that literature, the classic work of Perron (1989) made clear the ease with which stationary deviations from broken trend can be misinterpreted as $I(1)$ with drift. More generally, it is now widely appreciated that structural change and unit roots are easily confused, as emphasized for example by Stock (1994), who summarizes the huge subsequent literatures on unit roots and on structural change, and the interrelationships between the two.

The recent more general long-memory literature, in contrast, pays little attention to the possibility of confusing long memory and structural change. It is striking, for example, that the otherwise masterful surveys by Robinson (1994a), Beran (1994), and Baillie (1996) do not so much as mention the issue. The possibility of confusing long memory and structural change has of course arisen occasionally, in literatures ranging from mathematical statistics (Bhattacharya et al., 1983; Künsch, 1986; Teverovsky and Taqqu, 1997) to applied hydrology (Klemeš, 1974), but the warnings have had little impact. [1] We can only speculate as to the reasons, but they are probably linked to the facts that, on the one hand, theoretical work such as Bhattacharya et al. (1983) is highly abstract and appears lacking in motivation and intuition, while on the other hand, simulation examples such as Klemeš (1974), although well motivated and intuitive, offer neither theoretical justification nor Monte Carlo evidence.

In this paper we work with simple and intuitive econometric models, and we provide both theoretical justification and Monte Carlo evidence to support the claim that long memory and structural change are easily confused. In Section 2, we set the stage by considering alternative definitions of long memory and the relationships among them. In addition, we review the mechanisms for generating long memory that have been stressed previously in the literature, which are very different from those that we develop and therefore provide interesting contrast. In Section 3, we work with several simple models of structural change, or more precisely, stochastic regime switching, and we show when and why they produce realizations that appear to have long memory. In Section 4 we present a Monte Carlo analysis, which verifies the predictions of the theory and produces additional insights. We conclude in Section 5.

[1] Econometrics literature touching on the possibility of confusing long memory and structural change includes Hidalgo and Robinson (1996) and Lobato and Savin (1997).

2. Long memory: Definitions and origins

Here we briefly consider alternative definitions of long memory and the relationships among them, as well as "explanations" of the origin of long memory that have been stressed in the literature. Reviewing various definitions of long memory sets the stage for our subsequent analysis, which is heavily influenced by one of the definitions. Reviewing standard explanations of long memory, which are very different from ours, again sets the stage for our subsequent approach and provides useful contrast. Throughout this section and this paper, in keeping with the motivation above, our interest in long memory centers not on unit-root processes, but rather on mean-reverting fractionally integrated processes ($I(d)$, $0 < d < 1$), and when used without qualification "long memory" should be taken to mean $I(d)$, $0 < d < 1$.

2.1. Definitions

Traditionally, long memory has been defined in the time domain in terms of decay rates of long-lag autocorrelations, or in the frequency domain in terms of rates of explosion of low-frequency spectra. A long-lag autocorrelation definition of long memory for a covariance stationary process x is

$$\gamma_x(\tau) = c\tau^{2d-1} \quad \text{as } \tau \rightarrow \infty$$

and a low-frequency spectral definition of long memory is

$$f_x(\omega) = g\omega^{-2d} \quad \text{as } \omega \rightarrow 0^+.$$

An even more general low-frequency spectral definition of long memory is simply

$$f_x(\omega) = \infty \quad \text{as } \omega \rightarrow 0^+,$$

as in Heyde and Yang (1997). The long-lag autocorrelation and low-frequency spectral definitions of long memory are well known to be equivalent under conditions given, for example, in Beran (1994).

A third definition of long memory involves the rate of growth of variances of partial sums

$$\text{var}(S_T) = O(T^{2d+1}),$$

where $S_T = \sum_{t=1}^{T} x_t$. There is a tight connection between this variance-of-partial-sum definition of long memory and the spectral and autocorrelation definitions of long memory. In particular, because the spectral density at frequency zero is the limit of $(1/T)S_T$, a covariance stationary process has long memory in the generalized spectral sense of Heyde and Yang if and only if it has long memory for some $d > 0$ in the variance-of-partial-sum sense (see also Barndorff-Nielsen and Cox, 1989, p. 13). Our subsequent analysis is heavily influenced by the variance-of-partial-sum definition of long memory.

2.2. Origins

There is a natural desire to understand the appearance of long memory. Most econometric attention has focused on the role of aggregation.[2] Here we briefly review the two central aggregation-based routes to long memory, in order to contrast them to our subsequent perspective, which is very different.

First, following Granger (1980), consider the aggregation of $i=1,\ldots,N$ cross-sectional AR(1) units

$$x_{it} = \alpha_i x_{i,t-1} + \varepsilon_{it},$$

where ε_{it} is the white noise, $\varepsilon_{it} \perp \varepsilon_{jt}$, and $\alpha_i \perp \varepsilon_{jt}$ for all i,j,t. As $N \to \infty$, the spectrum of the aggregate $x_t = \sum_{i=1}^{N} x_{it}$ can be approximated as

$$f_x(\omega) \approx \frac{N}{2\pi} \mathrm{E}(\mathrm{var}(\varepsilon_{it})) \int \frac{1}{|1-\alpha \mathrm{e}^{\mathrm{i}\omega}|^2} \, \mathrm{d}F(\alpha),$$

where F is the c.d.f. governing the α's. If F is a beta distribution, i.e., if

$$\mathrm{d}F(\alpha) = \frac{2}{B(p,q)} \alpha^{2p-1}(1-\alpha^2)^{q-1} \, \mathrm{d}\alpha, \quad 0 \leqslant \alpha \leqslant 1,$$

then the τth autocovariance of x_t is

$$\gamma_x(\tau) = \frac{2}{B(p,q)} \int_0^1 \alpha^{2p+\tau-1}(1-\alpha^2)^{q-2} \, \mathrm{d}\alpha = A\tau^{1-q}.$$

Thus $x_t \sim I(1-q/2)$.

Granger's (1980) elegant bridge from cross-sectional aggregation to long-memory dynamics has since been generalized by a number of authors. For example, Chambers (1998) considers temporal aggregation in addition to cross-sectional aggregation, in both discrete and continuous time, and Lippi and Zaffaroni (1999) replace Granger's assumed beta distribution with weaker semiparametric assumptions.

An alternative route to long memory, which also involves aggregation, has been studied by Mandelbrot and his coauthors (e.g., Cioczek-Georges and Mandelbrot, 1995) and Taqqu and his coauthors (e.g., Taqqu et al., 1997). It has found wide application in the modeling of aggregate traffic on computer networks, although the basic idea is widely applicable. Define the stationary continuous-time binary series $W(t)$, $t \geqslant 0$ so that $W(t)=1$ during "on" periods and $W(t)=0$ during "off" periods. The lengths of the on and off periods

[2] Although aggregation is by far the most heavily studied route to long memory, it is not the only one to have appeared in the literature. In particular, Chen et al. (1999) have recently constructed a class of nonlinear Markov processes that can display long memory.

are iid at all leads and lags, and on and off periods alternate. Consider M such series, $W^{(m)}(t)$, $t \geqslant 0$, $m = 1, \ldots, M$, and define their aggregate in the interval $[0, Tt]$ by

$$W_M^*(Tt) = \int_0^{Tt} \left(\sum_{m=1}^{M} W^{(m)}(u) \right) \mathrm{d}u.$$

Let $F_1(x)$ and $F_2(x)$ denote the cumulative density functions of durations during on and off periods, and assume that they have power-law tails,

$$1 - F_1(x) \sim c_1 x^{-\alpha_1} L_1(x), \quad 1 < \alpha_1 < 2,$$

$$1 - F_2(x) \sim c_2 x^{-\alpha_2} L_2(x), \quad 1 < \alpha_2 < 2,$$

which imply infinite variance. Now let $M \to \infty$ and then let $T \to \infty$. Then $W_M^*(Tt)$, appropriately standardized, converges to a fractional Brownian motion.[3]

Parke (1999) considers a closely related discrete-time error duration model, $y_t = \sum_{s=-\infty}^{t} g_{s,t}\varepsilon_s$, where $\varepsilon_t \sim \mathrm{iid}(0, \sigma^2)$, $g_{s,t} = 1(t \leqslant s + n_s)$, and n_s is a stochastic duration. Effectively Parke, as with Mandelbrot and Taqqu, is concerned with an output equal to the sum of all current and past shocks that are still "alive", and in Parke's model, again as in the Mandelbrot–Taqqu model, long memory arises when the lifetime duration (n_s in Parke's setup) has infinite variance. The Mandelbrot–Taqqu–Parke approach beautifully illustrates the intimate connection between long memory and heavy tails.[4]

Both the Granger and the Mandelbrot–Taqqu–Parke approaches generate rich long-memory dynamics from aggregation of much simpler processes. Yet the aggregation approach is not fully satisfying. On the theoretical side, it works from assumptions that are both rigid and difficult to verify, and on the empirical side, we know of no evidence that long memory tends to be relatively more prevalent in aggregated as opposed to disaggregated series. With this in mind, we now explore a very different and complementary route to long memory: structural change.

3. Long memory and structural change

Structural change is likely widespread in economic relationships; see Stock and Watson (1996) for a recent and persuasive empirical analysis. There are of course huge econometric literatures on testing for structural change, and

[3] For general background on fractional Brownian motion and infinite variance, see Samorodnitsky and Taqqu (1994). For background on the link between infinite variance and power law tails, see Embrechts et al. (1997).

[4] Liu (1995) also establishes a link between long memory and infinite variance.

on estimating models of structural change and stochastic regime switching. In what follows, we focus on simple models of stochastic regime switching.

3.1. A mixture model

We will show that a standard mixture model with appropriately time-varying mixture weight ("break probability") will display $\mathrm{var}(S_T)$ behavior that matches that of an $I(d)$ process. Specifically, let

$$v_t = \begin{cases} 0 & \text{w.p. } 1-p, \\ w_t & \text{w.p. } p, \end{cases}$$

where $w_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_w^2)$. Note that $\mathrm{var}(\sum_{t=1}^T v_t) = pT\sigma_w^2 = O(T)$. If, however, instead of requiring constant p we allow it to change appropriately with sample size, then we can immediately obtain a partial sum variance that grows consistent with long memory.

Proposition 1. If $p = O(T^{2d-2})$, $0 < d < 1$, *then* $\mathrm{var}(S_T) = O(T^{2(d-1)+1})$.

Proof. $\mathrm{var}(\sum_{t=1}^T v_t) = O(T^{2d-2})T\sigma_w^2 = O(T^{2d-1}) = O(T^{2(d-1)+1})$. □

It is straightforward to move to a richer model for the mean of a series:

$$\mu_t = \mu_{t-1} + v_t,$$

$$v_t = \begin{cases} 0 & \text{w.p. } 1-p, \\ w_t & \text{w.p. } p, \end{cases}$$

where $w_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_w^2)$. Note that $\mathrm{var}(\sum_{t=1}^T v_t) = pT\sigma_w^2 = O(T)$. As before, let $p = O(T^{2d-2})$, $0 < d < 1$, so that the variance of partial sums of v_t grows at the rate corresponding to $I(d-1)$ behavior, which implies that the variance of partial sums of $\mu_t = \sum_{t=1}^T v_t$ grows at the rate corresponding to $I(d)$ behavior.

It is also straightforward to move to an even richer "mean-plus-noise model" in state space form:

$$y_t = \mu_t + \varepsilon_t,$$

$$\mu_t = \mu_{t-1} + v_t,$$

$$v_t = \begin{cases} 0 & \text{w.p. } 1-p, \\ w_t & \text{w.p. } p, \end{cases}$$

where $w_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_w^2)$ and $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\varepsilon^2)$, which will display the same behavior of variances of partial sums when $p = O(T^{2d-2})$, $0 < d < 1$.

Many additional generalizations could of course be entertained. The Balke and Fomby (1989) model of infrequent permanent shocks, for example, is a straightforward extension of the simple mixture model described above. Whatever the model, the key idea is to let p decrease with the sample size, so that regardless of the sample size, realizations tend to have just a few breaks.

3.2. The stochastic permanent break model

Engle and Smith (1999) propose the "stochastic permanent break" (STOPBREAK) model

$$y_t = \mu_t + \varepsilon_t,$$

$$\mu_t = \mu_{t-1} + q_{t-1}\varepsilon_{t-1},$$

where $q_t = q(|\varepsilon_t|)$ is nondecreasing in $|\varepsilon_t|$ and bounded by zero and one, so that bigger innovations have more permanent effects, and $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\varepsilon^2)$. They use $q_t = \varepsilon_t^2/(\gamma + \varepsilon_t^2)$ for $\gamma > 0$.

Quite interestingly for our purposes, Engle and Smith show that their model is an approximation to the mean-plus-noise model:

$$y_t = \mu_t + \varepsilon_t,$$

$$\mu_t = \mu_{t-1} + v_t,$$

where

$$v_t = \begin{cases} 0 & \text{w.p. } 1 - p, \\ w_t & \text{w.p. } p, \end{cases}$$

$\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\varepsilon^2)$ and $w_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_w^2)$. They note, moreover, that although other approximations to the mean-plus-noise model are available, the STOPBREAK model is designed to bridge the gap between transience and permanence of shocks and therefore provides a better approximation, for example, than an exponential smoother, which is the best linear approximation.

We now introduce a simple modification. We allow γ to change with T and write

$$y_t = \mu_t + \varepsilon_t,$$

$$\mu_t = \mu_{t-1} + \frac{\varepsilon_{t-1}^2}{\gamma_T + \varepsilon_{t-1}^2}\varepsilon_{t-1}.$$

Variances of partial sums of this process will match those of a fractionally integrated process under certain conditions, the key element of which involves the nature of time variation in γ.

Proposition 2. If (a) $\mathrm{E}(\varepsilon_t^6) < \infty$ *and* (b) $\gamma_T \to \infty$ *as* $T \to \infty$ *and* $\gamma_T = O(T^\delta)$ *for some* $\delta > 0$, *then the variances of partial sums of y grow at a rate corresponding to* $I(1-\delta)$ *behavior.*

Proof. Note that

$$\begin{aligned}
\mathrm{var}\left(\sum_{t=1}^{T} \Delta y_t\right) &= \mathrm{var}\left(\varepsilon_T - \varepsilon_0 + \sum_{t=1}^{T} \frac{\varepsilon_{t-1}^3}{\gamma_T + \varepsilon_{t-1}^2}\right) \\
&= \mathrm{var}(\varepsilon_T - \varepsilon_0) + \mathrm{var}\left(\sum_{t=1}^{T} \frac{\varepsilon_{t-1}^3}{\gamma_T + \varepsilon_{t-1}^2}\right) - 2\mathrm{E}\left(\frac{\varepsilon_0^4}{\gamma_T + \varepsilon_0^2}\right) \\
&= 2\sigma^2 - 2\mathrm{E}\left(\frac{\varepsilon_0^4}{\gamma_T + \varepsilon_0^2}\right) \\
&\quad + T\left\{\mathrm{E}\left(\frac{\varepsilon_{t-1}^6}{(\gamma_T + \varepsilon_{t-1}^2)^2}\right) - \left[\mathrm{E}\left(\frac{\varepsilon_{t-1}^3}{\gamma_T + \varepsilon_{t-1}^2}\right)\right]^2\right\} \\
&= O(T/\gamma_T^2) \\
&= O(T^{1-2\delta}) \\
&= O(T^{2(-\delta)+1}),
\end{aligned}$$

where the second equality follows from the maintained assumption that $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\varepsilon^2)$, the fourth equality follows from Assumption (a), and the fifth from Assumption (b). Thus variances of partial sums of Δy grow at a rate corresponding to $I(-\delta)$ behavior, so variances of partial sums of y grow at a rate corresponding to $I(1-\delta)$ behavior. $\square$

It is interesting to note that the standard STOPBREAK model corresponds to $\gamma_T = \gamma$, which corresponds to $\delta = 0$. Hence the standard STOPBREAK model is $I(1)$.

3.3. The Markov-switching model

All of the models considered thus far are effectively mixture models. The mean-plus-noise model and its relatives are explicit mixture models, and the STOPBREAK model is an approximation to the mean-plus-noise model. We now consider a richer dynamic mixture model, the Markov-switching model of Hamilton (1989).

Let $\{s_t\}_{t=1}^T$ be the (latent) sample path of two-state first-order autoregressive process, taking the value 0 or 1, with transition probability matrix given by

$$M = \begin{pmatrix} p_{00} & 1-p_{00} \\ 1-p_{11} & p_{11} \end{pmatrix}.$$

The ijth element of M gives the probability of moving from state i (at time $t-1$) to state j (at time t). Note that there are only two free parameters, the staying probabilities, p_{00} and p_{11}. Let $\{y_t\}_{t=1}^T$ be the sample path of an observed time series that depends on $\{s_t\}_{t=1}^T$ such that the density of y_t conditional upon s_t is

$$f(y_t|s_t;\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(y_t-\mu_{s_t})^2}{2\sigma^2}\right).$$

Thus, y_t is Gaussian white noise with a potentially switching mean, and we write

$$y_t = \mu_{s_t} + \varepsilon_t,$$

where $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0,\sigma^2)$ and s_t and ε_τ are independent for all t and τ.[5]

Variances of partial sums of the Markov-switching process will match those of a fractionally integrated process under certain conditions, which as the reader who has progressed to this point will surely suspect, involve the nature of time variation in p_{00} and p_{11}.

Proposition 3. Assume that (a) $\mu_0 \neq \mu_1$ *and that* (b) $p_{00} = 1 - c_0 T^{-\delta_0}$ *and* $p_{11} = 1 - c_1 T^{-\delta_1}$, *with* $\delta_0, \delta_1 > 0$ *and* $0 < c_0, c_1 < 1$. *Then the variances of partial sums of y grow at a rate corresponding to* $I((1/2)\max(\min(\delta_0,\delta_1) - |\delta_0 - \delta_1|, 0))$ *behavior.*

Proof. Let $\xi_t = (I(s_t=0)\ I(s_t=1))'$, $\mu = (\mu_0, \mu_1)$, and $\Gamma_j = \mathrm{E}(\xi_t \xi_{t-j}')$. Then $y_t = \mu'\xi_t + \varepsilon_t$, so

$$\begin{aligned} \mathrm{var}\left(\sum_{t=1}^T y_t\right) &= \mathrm{var}\left(\sum_{t=1}^T \mu'\xi_t\right) + T\sigma^2 \\ &= \mu'\left[\Gamma_0 T + \sum_{j=1}^T (T-j)(\Gamma_j + \Gamma_j')\right]\mu + T\sigma^2. \end{aligned}$$

[5] In the present example, only the mean switches across states. We could, of course, examine richer models with several parameters switching across states, but the simple model used here illustrates the basic idea.

For every T, the Markov chain is ergodic. Hence the unconditional variance–covariance matrix of ξ_t is

$$\Gamma_0 = \frac{(1-p_{00})(1-p_{11})}{(2-p_{00}-p_{11})^2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = O(1).$$

Now let

$$\lambda = p_{00} + p_{11} - 1 = 1 - c_0 T^{-\delta_0} - c_1 T^{-\delta_1}.$$

Then the jth autocovariance matrix of ξ_t is [6]

$$\Gamma_j = M^j \Gamma_0 = \frac{(1-p_{00})(1-p_{11})}{(2-p_{00}-p_{11})^2}\begin{bmatrix} \lambda^j & -\lambda^j \\ -\lambda^j & \lambda^j \end{bmatrix},$$

$$M^j = \begin{bmatrix} \dfrac{(1-p_{11})+\lambda^j(1-p_{00})}{2-p_{00}-p_{11}} & \dfrac{(1-p_{11})-\lambda^j(1-p_{11})}{2-p_{00}-p_{11}} \\ \dfrac{(1-p_{00})-\lambda^j(1-p_{00})}{2-p_{00}-p_{11}} & \dfrac{(1-p_{00})+\lambda^j(1-p_{11})}{2-p_{00}-p_{11}} \end{bmatrix}.$$

Thus

$$\begin{aligned} \frac{1}{T}\operatorname{var}\left(\sum_{t=1}^{T} y_t\right) &= O(1) + O\left(\frac{(1-p_{00})(1-p_{11})}{(1-\lambda)(2-p_{00}-p_{11})^2}\right) \\ &= O(1) + O(T^{\min(\delta_0,\delta_1)-|\delta_0-\delta_1|}), \end{aligned}$$

which in turn implies that

$$\operatorname{var}\left(\sum_{t=1}^{T} y_t\right) = O(T^{\max(\min(\delta_0,\delta_1)-|\delta_0-\delta_1|+1,1)}),$$

which completes the proof. □

It is interesting to note that the transition probabilities do not depend on T in the standard Markov-switching model, which corresponds to $\delta_0 = \delta_1 = 0$. Thus the standard Markov-switching model is $I(0)$, unlike the mean-plus-noise or STOPBREAK models, which are $I(1)$.[7]

Although we have not worked out the details, we conjecture that results similar to those reported here could be obtained in straightforward fashion for the threshold autoregressive (TAR) model, the smooth transition TAR model, and for reflecting barrier models of various sorts, by allowing the thresholds or reflecting barriers to change appropriately with sample size.

[6] See Hamilton (1994, p. 683) for the formula giving M^j in terms of p_{00}, p_{11}, and λ.

[7] In spite of the fact that it is truly $I(0)$, the standard Markov-switching model can nevertheless generate high persistence at short lags, as noted by Timmermann (2000) and verified in our subsequent Monte Carlo.

Similarly, we conjecture that Balke and Fomby (1997) threshold cointegration may be confused with fractional cointegration, for a suitably adapted series of thresholds, and that the Diebold and Rudebusch (1996) dynamic factor model with Markov-switching factor may be confused with fractional cointegration, for suitably adapted transition probabilities.

4. A Monte Carlo exploration

Our analysis thus far suggests that, under certain plausible conditions amounting to nonzero but "small" amounts of structural change, long memory and structural change may be confused. Motivated by the theory, we now perform a series of related Monte Carlo experiments. We simulate 10,000 realizations from various models of stochastic regime switching, and we characterize the finite-sample inference to which a researcher armed with a standard estimator of the long-memory parameter would be led.

We use the log-periodogram regression estimator proposed by Geweke and Porter-Hudak (GPH, 1983) and refined by Robinson (1994b, 1995). In particular, let $I(\omega_j)$ denote the sample periodogram at the jth Fourier frequency, $\omega_j = 2\pi j/T$, $j = 1, 2, \ldots, [T/2]$. The estimator of the parameter of fractional integration, d, is then based on the least-squares regression

$$\log[I(\omega_j)] = \beta_0 + \beta_1 \log(\omega_j) + u_j,$$

where $j = 1, 2, \ldots, m$, and $\hat{d} = -1/2\hat{\beta}_1$.[8] The least-squares estimator of β_1, and hence $\hat{d}$, is asymptotically normal and the corresponding theoretical standard error, $\pi(24m)^{-1/2}$, depends only on the number of periodogram ordinates used.

Of course, the actual value of the estimate of $\hat{d}$ also depends upon the particular choice of m. While the formula for the theoretical standard error suggests choosing a large value of m in order to obtain a small standard error, doing so may induce a bias in the estimator, because the relationship underlying the GPH regression in general holds only for frequencies close to zero. It turns out that consistency requires that m grows with sample size, but at a slower rate. Use of $m = \sqrt{T}$ has emerged as a popular rule of thumb, which we adopt.

[8] The calculations in Hurvich and Beltrao (1994) suggest that the estimator proposed by Robinson (1994b, 1995), which leaves out the very lowest frequencies in the regression in the GPH regression, has larger MSE than the original Geweke and Porter-Hudak (1983) estimator defined over all of the first m Fourier frequencies. For that reason, we include periodogram ordinates at all of the first m Fourier frequencies.

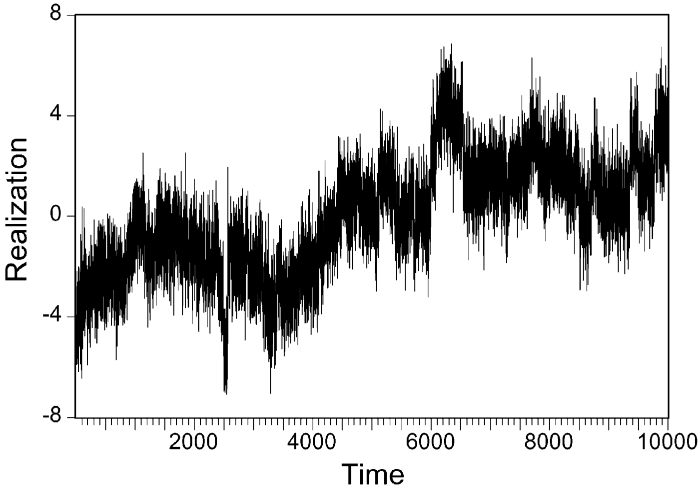

Fig. 1. Realization of length 10,000 of the mean-plus-noise process described in the text.

4.1. Mean-plus-noise model

We first consider the finite-sample behavior of the mean-plus-noise model. We parameterize the model as

$$y_t = \mu_t + \varepsilon_t,$$

$$\mu_t = \mu_{t-1} + v_t,$$

$$v_t = \begin{cases} 0 & \text{w.p. } 1 - p, \\ w_t & \text{w.p. } p, \end{cases}$$

where $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$ and $w_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$, $t = 1, 2, \ldots, T$. To build intuition before proceeding to the Monte Carlo, we first show in Fig. 1 a specific realization of the mean-plus-noise model with $p = 0.01$ and $T = 10,000$. It is clear that there are only a few breaks, with lots of noise superimposed. In Fig. 2 we plot the average log periodogram against log frequency for the same process, using $\sqrt{10,000}$ periodogram ordinates, where the averaging is done across 10,000 replications. The low-frequency log-periodogram looks approximately linear.

Now we proceed to the Monte Carlo analysis. We vary p and T, examining all pairs of $p \in \{0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1\}$ and $T \in \{100, 200, 300, 400, 500, 1000, \ldots, 5000\}$. In Table 1 we report the empirical sizes of nominal 5% tests of $d = 0$. They are increasing in T and p, which makes sense for two reasons. First, for fixed $p > 0$, the null is in fact false, so

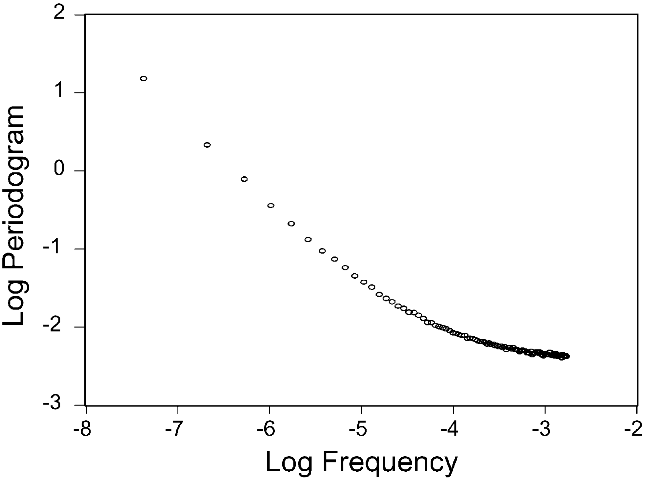

Fig. 2. The average of each of the first 100 log-periodogram ordinates across 10,000 simulated realizations of the mean-plus-noise process, each of length 10,000.

Table 1
Mean-plus-noise model: empirical sizes of nominal 5% tests of $d=0$[a]

	p						
T	0.0001	0.0005	0.001	0.005	0.01	0.05	0.1
100	0.165	0.169	0.180	0.268	0.355	0.733	0.866
200	0.141	0.167	0.202	0.412	0.590	0.941	0.979
300	0.141	0.186	0.239	0.554	0.749	0.986	0.994
400	0.128	0.195	0.276	0.670	0.867	0.996	0.998
500	0.127	0.210	0.313	0.755	0.922	0.999	1.000
1000	0.146	0.337	0.519	0.953	0.996	1.000	1.000
1500	0.177	0.452	0.674	0.993	1.000	1.000	1.000
2000	0.207	0.555	0.778	0.999	1.000	1.000	1.000
2500	0.234	0.634	0.850	1.000	1.000	1.000	1.000
3000	0.265	0.703	0.898	1.000	1.000	1.000	1.000
3500	0.293	0.762	0.931	1.000	1.000	1.000	1.000
4000	0.318	0.807	0.957	1.000	1.000	1.000	1.000
4500	0.355	0.840	0.972	1.000	1.000	1.000	1.000
5000	0.380	0.873	0.983	1.000	1.000	1.000	1.000

[a] T denotes sample size and p denotes the mixture probability. We report the fraction of 10,000 trials in which inference based on the Geweke–Porter–Hudak procedure leads to rejection of the hypothesis that $d=0$, using a nominal 5% test based on $\sqrt{T}$ periodogram ordinates.

Table 2
Mean-plus-noise model: empirical sizes of nominal 5% tests of $d=1$[a]

	p						
T	0.0001	0.0005	0.001	0.005	0.01	0.05	0.1
100	0.373	0.376	0.376	0.353	0.343	0.290	0.277
200	0.481	0.475	0.470	0.424	0.383	0.277	0.258
300	0.516	0.505	0.494	0.419	0.355	0.246	0.250
400	0.523	0.509	0.488	0.390	0.315	0.239	0.238
500	0.568	0.548	0.526	0.399	0.318	0.230	0.231
1000	0.639	0.596	0.542	0.351	0.262	0.223	0.211
1500	0.681	0.617	0.550	0.303	0.235	0.201	0.198
2000	0.705	0.616	0.542	0.277	0.216	0.200	0.197
2500	0.721	0.621	0.526	0.255	0.211	0.195	0.194
3000	0.735	0.616	0.520	0.247	0.200	0.193	0.185
3500	0.751	0.624	0.502	0.232	0.208	0.182	0.185
4000	0.761	0.619	0.497	0.226	0.195	0.181	0.179
4500	0.767	0.615	0.495	0.223	0.191	0.181	0.183
5000	0.763	0.593	0.463	0.209	0.199	0.179	0.184

[a] T denotes sample size and p denotes the mixture probability. We report the fraction of 10,000 trials in which inference based on the Geweke–Porter–Hudak procedure leads to rejection of the hypothesis that $d=1$, using a nominal 5% test based on $\sqrt{T}$ periodogram ordinates.

power increases in T by consistency of the test.[9] Second, for fixed T, we have more power to detect $I(1)$ behavior as p grows, because we have a greater number of nonzero innovations, whose effects we can observe.

The thesis of this paper is that structural change may be easily confused with fractional integration, so it is important to be sure that we are not rejecting the $d=0$ hypothesis simply because of a unit root. Hence we also test the $d=1$ hypothesis. The results appear in Table 2, which reports empirical sizes of nominal 5% tests of $d=1$, executed by testing $d=0$ on differenced data using the GPH procedure. The $d=1$ rejection frequencies decrease with T, because the null is in fact true. They also decrease sharply with p, because the effective sample size grows quickly as p grows.

In Fig. 3 we plot kernel estimates of the density of $\hat{d}$ for $T\in\{400,1000,2500,5000\}$ and $p\in\{0.0001,0.0005,0.001,0.005,0.01,0.05,0.1\}$.[10] Fig. 3 illuminates the way in which the GPH rejection frequencies increase with p and T. The density estimates shift gradually to the right as p and T increase. For small p, the estimated densities are bimodal in some cases. Evidently

[9] When $p=0$, the process is white noise and hence $I(0)$. For all $p>0$, the change in the mean process is iid, and hence the mean process is $I(1)$, albeit with highly non-Gaussian increments. When $p=1$, the mean process is a Gaussian random walk.

[10] Here and in all subsequent density estimation, we select the bandwidth by Silverman's rule, and we use an Epanechnikov kernel.

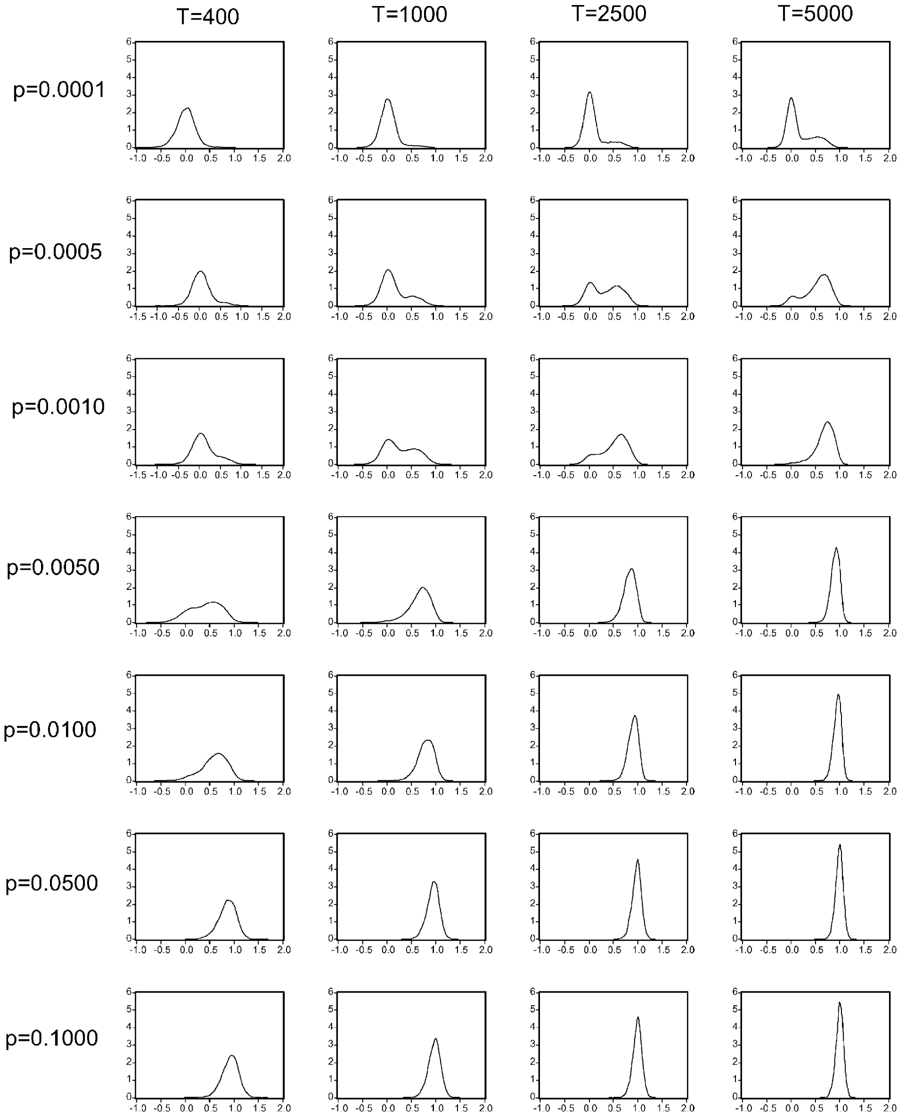

Fig. 3. Kernel density estimates of the distribution of the Geweke–Porter–Hudak log-periodogram regression estimates of the fractional integration parameter d, based on $\sqrt{T}$ periodogram ordinates. T denotes sample size and p denotes the mixture probability.

the bimodality results from a mixture of two densities: one is the density of $\hat{d}$ when no structural change occurs, and the other is the density of $\hat{d}$ when there is at least one break.

Table 3
Mean-plus-noise model: mean estimate of the fractional integration parameter, d[a]

	p						
T	0.0001	0.0005	0.001	0.005	0.01	0.05	0.1
100	0.008	0.015	0.030	0.120	0.210	0.580	0.727
200	0.005	0.029	0.056	0.232	0.392	0.752	0.851
300	0.012	0.048	0.090	0.338	0.503	0.827	0.894
400	0.013	0.061	0.117	0.412	0.591	0.864	0.920
500	0.016	0.075	0.141	0.476	0.648	0.892	0.937
1000	0.036	0.155	0.275	0.668	0.795	0.943	0.970
1500	0.052	0.223	0.377	0.754	0.850	0.964	0.981
2000	0.072	0.290	0.455	0.802	0.884	0.970	0.985
2500	0.087	0.334	0.512	0.831	0.904	0.977	0.989
3000	0.103	0.387	0.567	0.857	0.917	0.981	0.990
3500	0.121	0.430	0.608	0.872	0.928	0.985	0.992
4000	0.133	0.467	0.640	0.887	0.936	0.986	0.992
4500	0.156	0.498	0.667	0.897	0.943	0.988	0.996
5000	0.168	0.529	0.694	0.907	0.949	0.988	0.995

[a] T denotes sample size and p denotes the mixture probability. We report the average (across 10,000 trials) of the Geweke–Porter–Hudak log-periodogram regression estimates of the fractional integration parameter d, based on $\sqrt{T}$ periodogram ordinates.

Finally, in Table 3, we show the mean value of $\hat{d}$ for the same wide range of T and p values as in Tables 1 and 2. This facilitates a crude check of the asymptotics, by checking whether taking $p = O(T^{2d-2})$, as T increases, produces the appearance of fractional integration with parameter d. The results are encouraging, as illustrated, for example, by the underlined entries in the table, obtained by taking $p^* = 2.6T^{2d-2} = 2.6T^{2\times 0.5-2}$ and then underlining the mean value of $\hat{d}$ whose p is closest to p^*. As the theory predicts, all of the resulting mean values of $\hat{d}$ are close to 0.5.

4.2. Stochastic permanent break model

Next, we consider the finite-sample behavior of the STOPBREAK model:

$$y_t = \mu_t + \varepsilon_t,$$

$$\mu_t = \mu_{t-1} + \frac{\varepsilon_{t-1}^2}{\gamma + \varepsilon_{t-1}^2}\varepsilon_{t-1},$$

with $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$. In Fig. 4 we show a specific realization of the STOPBREAK process with $\gamma = 500$ and $T = 10{,}000$. Because the evolution of the STOPBREAK process is smooth, as it is only an *approximation* to a mixture model, we do not observe sharp breaks in the realization. In Fig. 5 we

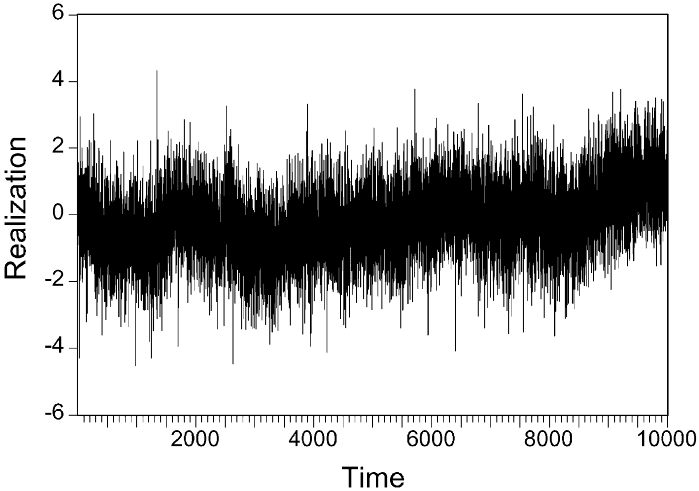

Fig. 4. Realization of length 10,000 of the STOPBREAK process described in the text.

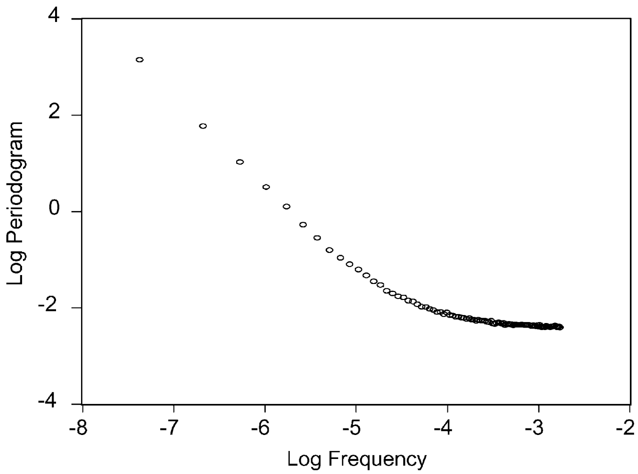

Fig. 5. The average of each of the first 100 log-periodogram ordinates across 10,000 simulated realizations of the STOPBREAK process, each of length 10,000.

plot the average log periodogram against log frequency, using $\sqrt{10,000}$ periodogram ordinates and 10,000 replicated realizations of the process. The low-frequency periodogram looks linear.

In the Monte Carlo experiment, we examine all pairs of $\gamma \in \{10^{-5}, 10^{-4}, \ldots, 10^{3}, 10^{4}\}$ and $T \in \{100, 200, 300, 400, 500, 1000, 1500, \ldots, 5000\}$. In Table 4

Table 4
Stochastic permanent break model: empirical sizes of nominal 5% tests of $d = 0$[a]

	γ									
T	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1	10	10^2	10^3	10^4
100	0.976	0.978	0.978	0.978	0.977	0.971	0.844	0.195	0.167	0.166
200	0.995	0.995	0.995	0.995	0.996	0.995	0.972	0.292	0.134	0.134
300	0.999	0.999	0.999	0.999	0.998	0.999	0.993	0.444	0.123	0.121
400	0.999	0.999	0.999	1.000	1.000	0.999	0.998	0.587	0.120	0.119
500	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.697	0.115	0.114
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.951	0.118	0.101
1500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.158	0.091
2000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.213	0.090
2500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.266	0.082
3000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.341	0.084
3500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.415	0.078
4000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.486	0.084
4500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.554	0.081
5000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.615	0.080

[a] T denotes sample size and γ denotes the STOPBREAK parameter. We report the fraction of 10,000 trials in which inference based on the Geweke–Porter–Hudak procedure leads to rejection of the hypothesis that $d = 0$, using a nominal 5% test based on $\sqrt{T}$ periodogram ordinates.

we report the empirical sizes of nominal 5% tests of $d = 0$. The $d = 0$ rejection frequencies are increasing in T and decreasing in γ, which makes intuitive sense. First consider fixed γ and varying T. The STOPBREAK process is $I(1)$ for all $\gamma < \infty$, so the null of $d = 0$ is in fact false, and power increases in T by consistency of the test. Now consider fixed T and varying γ. For all $\gamma < \infty$, the change in the mean process is iid, and hence the mean process is $I(1)$, albeit with non-Gaussian increments. But we have less power to detect $I(1)$ behavior as γ grows, because we have a smaller effective sample size.[11] In fact, as γ approaches ∞, the process approaches $I(0)$ white noise.

As before, we also test the $d = 1$ hypothesis by employing GPH on differenced data. In Table 5 we report the empirical sizes of nominal 5% tests of $d = 1$. The rejection frequencies tend to be decreasing in T and increasing in γ, which makes sense for the reasons sketched above. In particular, because the STOPBREAK process is $I(1)$, the $d = 1$ rejection frequencies should naturally drop toward nominal size as T grows. Alternatively, it becomes progressively easier to reject $d = 1$ as γ increases, for any fixed T, because the STOPBREAK process gets closer to $I(0)$ as γ increases.

[11] The last two columns of the table, however, reveal a nonmonotonicity in T: empirical size first drops and then rises with T.

Table 5
Stochastic permanent break model: empirical sizes of nominal 5% tests of $d = 1$[a]

	γ									
T	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1	10	10^2	10^3	10^4
100	0.176	0.174	0.175	0.176	0.175	0.176	0.379	0.729	0.742	0.742
200	0.133	0.133	0.133	0.133	0.134	0.141	0.318	0.833	0.842	0.843
300	0.121	0.121	0.122	0.123	0.121	0.124	0.250	0.838	0.848	0.849
400	0.122	0.122	0.121	0.119	0.123	0.119	0.206	0.841	0.859	0.859
500	0.116	0.116	0.116	0.117	0.116	0.114	0.196	0.861	0.872	0.873
1000	0.098	0.098	0.098	0.098	0.100	0.099	0.133	0.892	0.913	0.912
1500	0.091	0.091	0.092	0.092	0.096	0.095	0.110	0.904	0.922	0.923
2000	0.088	0.088	0.088	0.088	0.088	0.086	0.101	0.906	0.936	0.936
2500	0.084	0.084	0.084	0.085	0.080	0.079	0.087	0.910	0.945	0.946
3000	0.082	0.082	0.083	0.083	0.081	0.082	0.084	0.911	0.947	0.947
3500	0.082	0.082	0.083	0.081	0.079	0.081	0.082	0.910	0.953	0.954
4000	0.083	0.083	0.083	0.083	0.081	0.081	0.083	0.904	0.960	0.959
4500	0.079	0.079	0.079	0.079	0.077	0.078	0.083	0.904	0.961	0.961
5000	0.075	0.075	0.075	0.075	0.076	0.076	0.079	0.891	0.963	0.963

[a] T denotes sample size and γ denotes the STOPBREAK parameter. We report the fraction of 10,000 trials in which inference based on the Geweke–Porter–Hudak procedure leads to rejection of the hypothesis that $d = 1$, using a nominal 5% test based on $\sqrt{T}$ periodogram ordinates.

In Fig. 6 we show kernel estimates of the density of $\hat{d}$ for $T \in \{400, 1000, 2500, 5000\}$ and $\gamma \in \{10^{-5}, 10^{-4}, \ldots, 10^3, 10^4\}$. As the sample size grows, the estimated density shifts to the right and the median of $\hat{d}$ approaches unity for $\gamma < 10{,}000$. This is expected because the STOPBREAK process is $I(1)$. However, as γ increases, the effective sample size required to detect this nonstationarity also increases. As a result, when γ is large, the median of $\hat{d}$ is below unity even for a sample of size 5000.

4.3. Markov switching

Lastly, we analyze the finite-sample properties of the Markov-switching model. The model is

$$y_t = \mu_{s_t} + \varepsilon_t,$$

where $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, and s_t and ε_τ are independent for all t and τ. We take $\mu_0 = 0$ and $\mu_1 = 1$. In Fig. 7 we plot a specific realization with $p_{00} = p_{11} = 0.9995$ and $T = 10{,}000$. It appears that the regime has changed several times in this particular realization. In Fig. 8 we plot the corresponding average log periodogram against log frequency, using $\sqrt{10{,}000} = 100$ periodogram

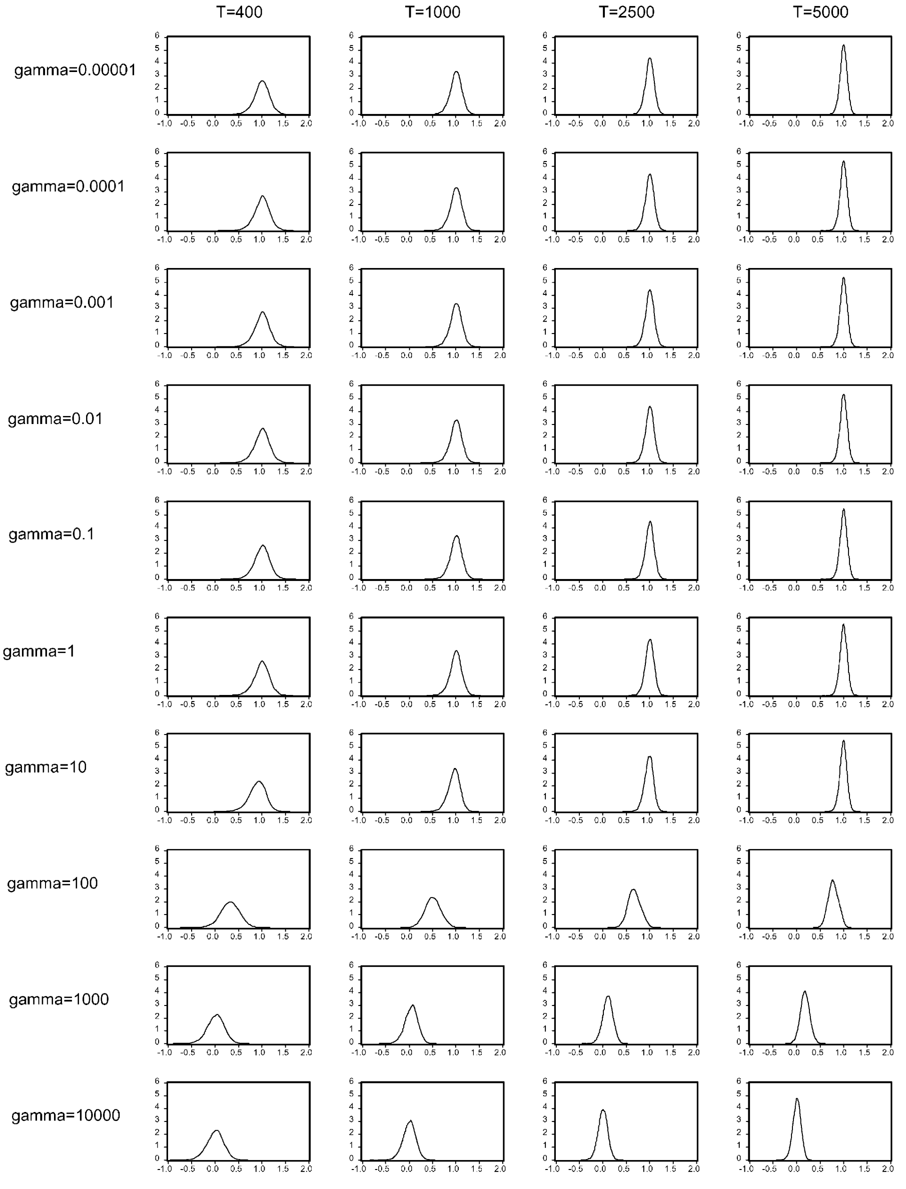

Fig. 6. Kernel density estimates of the distribution of the Geweke–Porter–Hudak log-periodogram regression estimates of the fractional integration parameter d, based on $\sqrt{T}$ periodogram ordinates. T denotes sample size and γ denotes the STOPBREAK parameter.

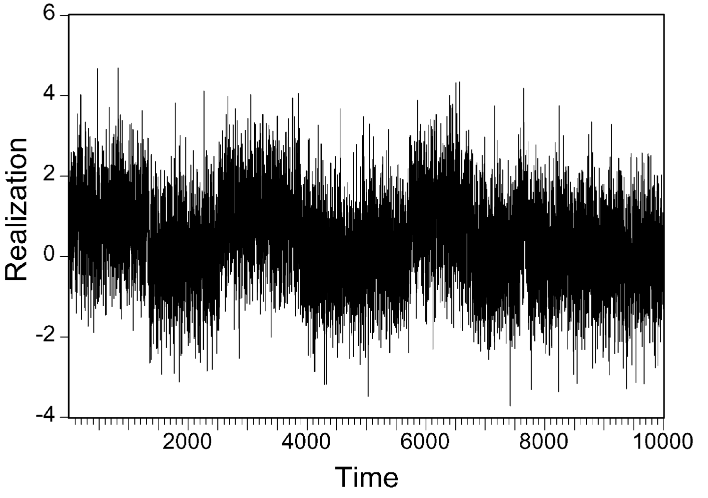

Fig. 7. Realization of length 10,000 of the Markov-switching process described in the text.

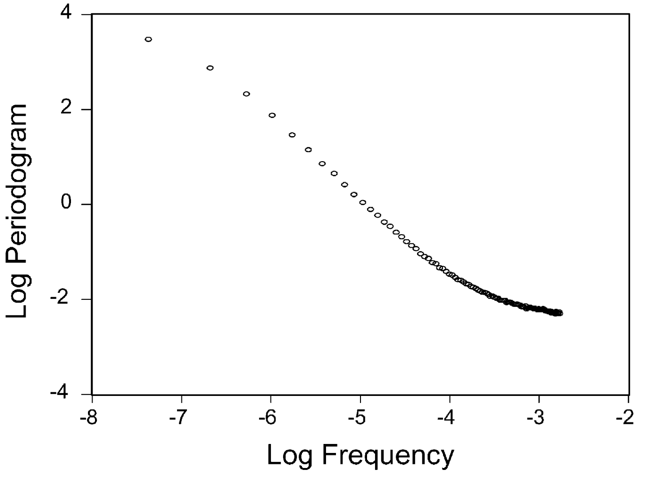

Fig. 8. The average of each of the first 100 log-periodogram ordinates across 10,000 simulated realizations of the Markov-switching process, each of length 10,000.

ordinates and 10,000 replicated realizations of the process; low-frequency linearity of the log periodogram appears to be a good approximation.

In the Monte Carlo analysis, we explore $p_{00} \in \{0.95, 0.99, 0.999\}$, $p_{11} \in \{0.95, 0.99, 0.999\}$, and $T \in \{100, 200, 300, 400, 500, 1000, 1500, \ldots, 5000\}$. In

Table 6
Markov-switching model: empirical sizes of nominal 5% tests of $d = 0$[a]

T	p_{00} / p_{11}	0.95 / 0.95	0.95 / 0.99	0.95 / 0.999	0.99 / 0.95	0.99 / 0.99	0.99 / 0.999	0.999 / 0.95	0.999 / 0.99	0.999 / 0.999
100		0.417	0.332	0.186	0.329	0.375	0.208	0.186	0.213	0.196
200		0.476	0.425	0.180	0.420	0.618	0.257	0.179	0.262	0.232
300		0.478	0.482	0.187	0.482	0.761	0.313	0.189	0.313	0.290
400		0.487	0.514	0.185	0.522	0.858	0.350	0.190	0.353	0.344
500		0.460	0.529	0.188	0.541	0.907	0.392	0.191	0.393	0.398
1000		0.383	0.559	0.214	0.561	0.981	0.549	0.212	0.554	0.628
1500		0.317	0.552	0.210	0.547	0.991	0.643	0.216	0.644	0.758
2000		0.266	0.523	0.213	0.522	0.995	0.716	0.215	0.716	0.849
2500		0.235	0.498	0.213	0.506	0.996	0.772	0.217	0.775	0.903
3000		0.205	0.472	0.211	0.462	0.997	0.813	0.208	0.810	0.941
3500		0.188	0.444	0.211	0.458	0.997	0.847	0.207	0.849	0.963
4000		0.174	0.415	0.203	0.432	0.997	0.869	0.206	0.865	0.975
4500		0.155	0.405	0.200	0.399	0.998	0.887	0.201	0.888	0.984
5000		0.143	0.367	0.195	0.375	0.997	0.902	0.190	0.903	0.990

[a] T denotes sample size, and $1 - p_{00}$ and $1 - p_{11}$ denote the Markov transition probabilities. We report the fraction of 10,000 trials in which inference based on the Geweke–Porter–Hudak procedure leads to rejection of the hypothesis that $d = 0$, using a nominal 5% test based on $\sqrt{T}$ periodogram ordinates.

Table 6 we show the empirical sizes of nominal 5% tests of $d = 0$. When both p_{00} and p_{11} are well away from unity, such as when $p_{00} = p_{11} = 0.95$, the rejection frequencies eventually decrease as the sample size increases, which makes sense because the process is $I(0)$. In contrast, when both p_{00} and p_{11} are large, such as when $p_{00} = p_{11} = 0.999$, the rejection frequency is increasing in T. This would appear inconsistent with the fact that the Markov-switching model is $I(0)$ for any fixed p_{00} and p_{11}, but it is not, as the dependence of rejection frequency on T is not monotonic. If we included $T > 5000$ in the design, we would eventually see the rejection frequency decrease.

In Table 7 we tabulate the empirical sizes of nominal 5% tests of $d = 1$. Although the persistence of the Markov-switching model is increasing in p_{00} and p_{11}, it turns out that it is nevertheless very easy to reject $d = 1$ in this particular experimental design.

In Fig. 9 we plot kernel estimates of the density of $\hat{d}$ for $p_{00}, p_{11} \in \{0.95, 0.99, 0.999\}$ and $T \in \{400, 1000, 2500, 5000\}$. When both p_{00} and p_{11} are away from unity, the estimated density tends to shift to the left and the median of $\hat{d}$ converges to zero as the sample size grows. When both p_{00} and p_{11} are near unity, the estimated density tends to shift to the right. These observations are consistent with our theory as discussed before. When p_{00} and p_{11} are close to unity and when T is relatively small, the regime does

Table 7
Markov-switching model: empirical sizes of nominal 5% tests of $d = 1$[a]

T	p_{00} / p_{11}	0.95 / 0.95	0.95 / 0.99	0.95 / 0.999	0.99 / 0.95	0.99 / 0.99	0.99 / 0.999	0.999 / 0.95	0.999 / 0.99	0.999 / 0.999
100		0.633	0.676	0.736	0.675	0.674	0.732	0.735	0.730	0.735
200		0.784	0.783	0.831	0.787	0.761	0.822	0.833	0.824	0.832
300		0.840	0.816	0.853	0.823	0.776	0.840	0.858	0.842	0.849
400		0.876	0.843	0.866	0.840	0.778	0.844	0.862	0.843	0.852
500		0.912	0.875	0.886	0.877	0.814	0.870	0.885	0.867	0.875
1000		0.969	0.939	0.918	0.938	0.869	0.901	0.915	0.901	0.903
1500		0.983	0.959	0.927	0.959	0.902	0.909	0.927	0.907	0.903
2000		0.994	0.978	0.948	0.980	0.935	0.931	0.944	0.931	0.927
2500		0.996	0.986	0.957	0.986	0.954	0.941	0.954	0.939	0.930
3000		0.998	0.988	0.960	0.991	0.969	0.948	0.959	0.945	0.934
3500		0.998	0.993	0.966	0.994	0.981	0.956	0.966	0.958	0.947
4000		0.999	0.994	0.971	0.995	0.984	0.962	0.970	0.961	0.948
4500		0.999	0.996	0.975	0.997	0.990	0.966	0.974	0.966	0.951
5000		1.000	0.998	0.975	0.997	0.992	0.968	0.977	0.969	0.951

[a] T denotes sample size, and $1 - p_{00}$ and $1 - p_{11}$ denote the Markov transition probabilities. We report the fraction of 10,000 trials in which inference based on the Geweke–Porter–Hudak procedure leads to rejection of the hypothesis that $d = 1$, using a nominal 5% test based on $\sqrt{T}$ periodogram ordinates.

not change with positive probability and, as a result, the estimated densities appear bimodal.

In closing this sub-section, we contrast our results for the Markov-switching model with those of Rydén et al. (1998), who find that the Markov-switching model does a poor job of mimicking long memory, which would seem to conflict with both our theoretical and Monte Carlo results. However, our theory requires that all diagonal elements of the transition probability matrix be near unity. In contrast, nine of the ten Markov-switching models estimated by Rydén et al. have at least one diagonal element well away from unity. Only their estimated model H satisfies our condition, and its dynamics are in fact highly persistent. Hence the results are entirely consistent.

5. Summary and concluding remarks

We have argued that structural change in general, and stochastic regime switching in particular, are intimately related to long memory and easily confused with it, so long as only a small amount of regime switching occurs in an observed sample path. We provided theoretical analysis of several environments, including a simple mixture model, Engle and Smith's (1999) stochastic permanent break model, and Hamilton's (1989) Markov-switching

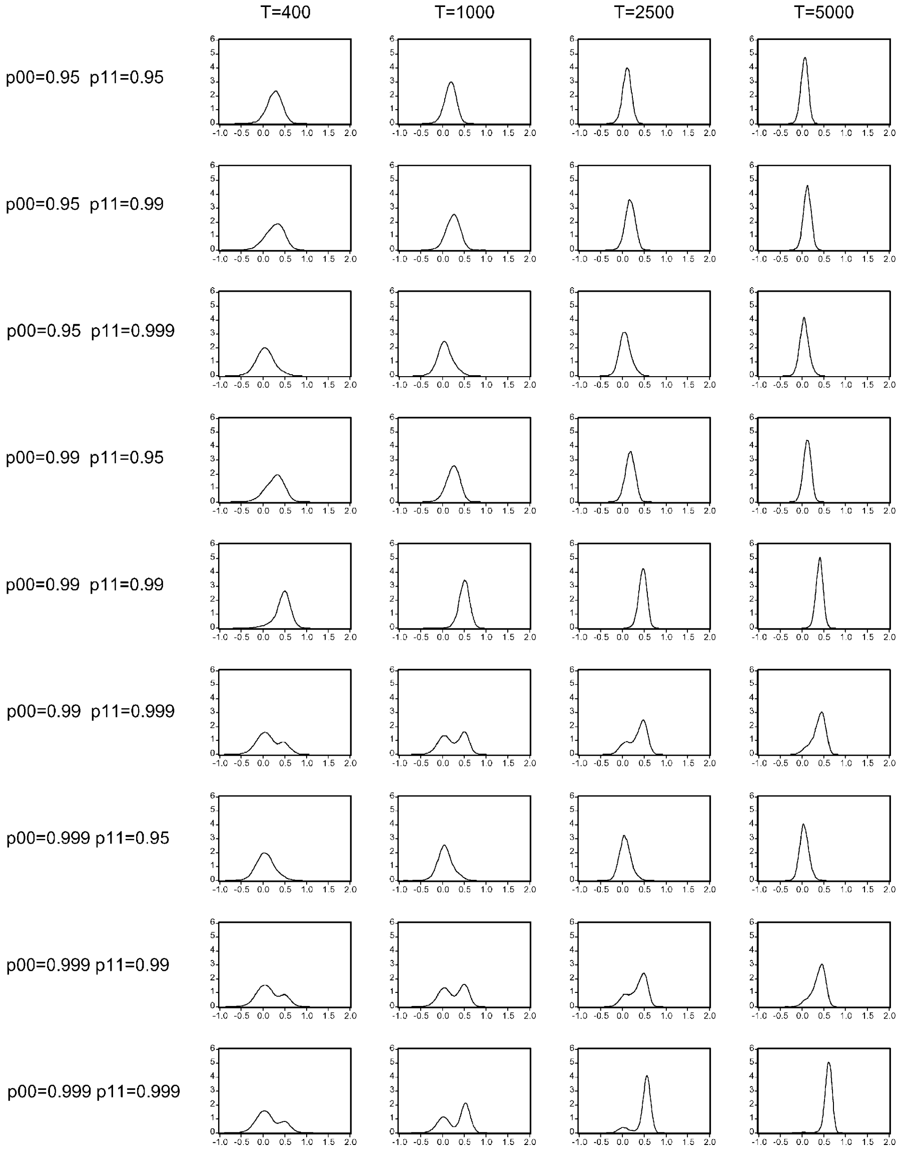

Fig. 9. Kernel density estimates of the distribution of the Geweke–Porter–Hudak log-periodogram regression estimates of the fractional integration parameter d, based on $\sqrt{T}$ periodogram ordinates. T denotes sample size, and $1 - p_{00}$ and $1 - p_{11}$ denote the Markov transition probabilities.

model. Simulations support the relevance of the theory in finite samples and make clear that the confusion is not merely a theoretical curiosity, but rather a distinct possibility in routine empirical economic and financial applications.

In what remains of this paper, we furnish additional perspective on our results, in two ways. First, we expand on the nature of the confusion between regime switching and the appearance of long memory studied in our paper. Second, we situate our work within the context of several related recent contributions.

5.1. On the theoretical perspective

Our device of letting certain parameters such as mixture probabilities vary with T is simply a thought experiment that proves useful for thinking about the appearance of long memory. We view our theory as effectively providing a "local to no breaks" perspective, in parallel to the use of "local-to-unity" asymptotics in autoregressions, a thought experiment that proves useful for characterizing the distribution of the dominant root. But just as in the context of a local-to-unity autoregressive root, which does not require that one literally believe that the root satisfies $\rho = 1 - c/T$ as T grows, we do not require that one literally believe that the mixture probability satisfies $p = cT^{2d-2}$ as T grows.

In practice, and in the Monte Carlo analysis that we performed, we are not interested in, and we do not explore, models with truly time-varying parameters (such as time-varying mixture probabilities). Similarly, we are not interested in expanding samples with size approaching infinity. Instead, our interest centers on fixed-parameter models with fixed finite T, the dynamics of which are in fact either $I(0)$ or $I(1)$. The theory suggests that confusion with fractional integration will result when only a small amount of breakage occurs, and therefore that the larger is T, the smaller must be the break probability, if confusion is to arise.

Our approach differs in an important sense from that of Granger–Mandelbrot–Taqqu–Parke, who develop models that are truly fractionally integrated and could for example be used to simulate realizations of such processes. In contrast, we use a certain asymptotic perspective to guide our thinking about whether and when finite-sample paths of truly $I(0)$ or $I(1)$ processes might nevertheless *appear* fractionally integrated, not as a device for producing sample paths of truly fractionally integrated processes. In our framework, the appearance of long memory requires models with rare breaks with lingering effects, whether truly $I(0)$ or $I(1)$. The fact that they occur rarely makes it hard to divine their nature and, under certain conditions, makes them appear fractionally integrated.

In all of the asymptotic thought experiments entertained in this paper, breaks are of the same stochastic size when they occur, but their probability of occurrence drops with T. We have not emphasized the dual case in which breaks occur with constant probability, but their stochastic size shrinks with T, because it is less likely to generate confusion between regime switching and fractional integration. Consider, for example, the following simple model for the mean of a series, in which a "break" in the mean occurs each period:

$$\mu_t = \mu_{t-1} + \sqrt{c} w_t,$$

where $w_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$, $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0,\sigma_\varepsilon^2)$, and $c > 0$. If c shrinks with T at an appropriate rate, then the variances of partial sums of μ_t will obviously grow at rates consistent with fractional integration. Yet it is unlikely that a realization of such a process would be mistaken as fractionally integrated; breaks occur each period, so it is easy to learn about the associated dynamics, and even the simplest data analytic methods would reveal that $\Delta\mu_t$ is simply independent white noise (with fixed variance $c(T)$, for any fixed T).

5.2. Related work

Here we sketch the relationship of our work to several recent and closely related contributions. First, Granger and Teräsvirta (1999) consider the following simple nonlinear process:

$$y_t = \mathrm{sign}(y_{t-1}) + \varepsilon_t,$$

where $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0,\sigma^2)$. This process behaves like a regime-switching process and, theoretically, the autocorrelations should decline exponentially. They show, however, that as the tail probability of ε_t decreases (presumably by decreasing the value of σ^2) so that there are fewer regime switches for any fixed sample size, long memory seems to appear, and the implied d estimates begin to grow. The Granger–Teräsvirta results, however, are based on single realizations (not Monte Carlo analysis), and no theoretical explanation is provided.

Second, in contemporaneous, independent and complementary work, Granger and Hyung (1999) develop a theory closely related to ours. They consider a mean-plus-noise model, and they show that its autocorrelations decay very slowly if $p = O(1/T)$. Their result is a special case of ours, with $d = 0.5$. Importantly, moreover, we show that $p = O(1/T)$ is not necessary to obtain the appearance of long memory, and we provide a link between the convergence rate of p and the apparent long-memory parameter d. We also provide related results for STOPBREAK models and Markov-switching models, as well as an extensive Monte Carlo analysis of finite-sample effects. On the other hand, Granger and Hyung consider some interesting topics which we have not considered in the present paper, such as common breaks in multivariate time series.

Finally, we note that our results are in line with those of Mikosch and Stărică (1999), who find structural change in asset return dynamics and argue that it could be responsible for evidence of long memory. We believe, however, that the temptation to jump to conclusions of "structural change producing spurious inferences of long memory" should be resisted, as such conclusions are potentially naive. Even if the "truth" is structural change, long memory may be a convenient shorthand description, which may remain very useful for tasks such as prediction.[12] Moreover, at least in the sorts of circumstances studied in this paper, "structural change" and "long memory" are effectively different labels for the same phenomenon, in which case attempts to label one as "true" and the other as "spurious" may be of dubious value.

Acknowledgements

Much of this paper was written while Diebold visited the Stern School of Business, New York University, whose hospitality is gratefully acknowledged. For helpful comments we thank the editor and two referees, participants in the June 1999 London School of Economics Financial Markets Group Conference on the Limits to Forecasting in Financial Economics, the October 1999 Cowles Foundation Conference on Econometrics, and the January 2000 North American Winter Meeting of the Econometric Society, as well as Torben Andersen, Tim Bollerslev, Rohit Deo, Steve Durlauf, Clive Granger, Bruce Hansen, Cliff Hurvich, Adrian Pagan, Fallaw Sowell, Jim Stock, and Mark Watson. None of those thanked, of course, are responsible in any way for the outcome. The National Science Foundation provided research support.

References

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001a. The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96, 42–55.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001b. The distribution of realized stock return volatility. Journal of Financial Economics, in press.

Baillie, R.T., 1996. Long-memory processes and fractional integration in econometrics. Journal of Econometrics 73, 5–59.

Balke, N.S., Fomby, T.B., 1989. Shifting trends, segmented trends, and infrequent permanent shocks. Journal of Monetary Economics 28, 61–85.

Balke, N.S., Fomby, T.B., 1997. Threshold cointegration. International Economic Review 38, 627–645.

[12] In a development that supports this conjecture, Clements and Krolzig (1998) show that fixed-coefficient autoregressions often outperform Markov-switching models for forecasting in finite samples, even when the true data-generating process is Markov switching.

Barndorff-Nielsen, O.E., Cox, D.R., 1989. Asymptotic Techniques for Use in Statistics. Chapman and Hall, New York.

Beran, J., 1994. Statistics for Long-Memory Processes. Chapman and Hall, New York.

Bhattacharya, R.N., Gupta, V.K., Waymire, E., 1983. The Hurst effect under trends. Journal of Applied Probability 20, 649–662.

Chambers, M., 1998. Long memory and aggregation in macroeconomic time series. International Economic Review 39, 1053–1072.

Chen, X., Hansen, L.P., Carrasco, M., 1999. Nonlinearity and temporal dependence, Manuscript, Department of Economics, University of Chicago.

Cioczek-Georges, R., Mandelbrot, B.B., 1995. A class of micropulses and antipersistent fractional Brownian motion. Stochastic Processes and Their Applications 60, 1–18.

Clements, M.P., Krolzig, H.-M., 1998. A comparison of the forecast performance of Markov-switching and threshold autoregressive models of U.S. GNP. Econometrics Journal 1, 47–75.

Comte, F., Renault, E., 1998. Long memory in continuous-time stochastic volatility models. Mathematical Finance 8, 291–323.

Diebold, F.X., Rudebusch, G.D., 1989. Long memory and persistence in aggregate output. Journal of Monetary Economics 24, 189–209.

Diebold, F.X., Rudebusch, G.D., 1996. Measuring business cycles: a modern perspective. Review of Economics and Statistics 78, 67–77.

Ding, Z., Engle, R.F., Granger, C.W.J., 1993. A long-memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83–106.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events. Springer, New York.

Engle, R.F., Smith, A.D., 1999. Stochastic permanent breaks. Review of Economics and Statistics 81, 553–574.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long-memory time series models. Journal of Time Series Analysis 4, 221–238.

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

Granger, C.W.J., Hyung, N., 1999. Occasional structural breaks and long memory. Discussion Paper 99-14, University of California, San Diego.

Granger, C.W.J., Teräsvirta, T., 1999. A simple nonlinear time series model with misleading linear properties. Economics Letters 62, 161–165.

Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton.

Heyde, C.C., Yang, Y., 1997. On defining long range dependence. Journal of Applied Probability 34, 939–944.

Hidalgo, J., Robinson, P.M., 1996. Testing for structural change in a long-memory environment. Journal of Econometrics 70, 159–174.

Hurvich, C.M., Beltrao, K.I., 1994. Automatic semiparametric estimation of the memory parameter of a long-memory time series. Journal of Time Series Analysis 15, 285–302.

Klemeš, V., 1974. The hurst phenomenon: a puzzle? Water Resources Research 10, 675–688.

Künsch, H.R., 1986. Discrimination between monotonic trends and long-range dependence. Journal of Applied Probability 23, 1025–1030.

Lippi, M., Zaffaroni, P., 1999. Contemporaneous aggregation of linear dynamic models in large economies. Manuscript, Research Department, Bank of Italy.

Liu, M., 1995. Modeling long memory in stock market volatility. Manuscript, Department of Economics, Duke University.

Lobato, I.N., Savin, N.E., 1997. Real and spurious long-memory properties of stock-market data. Journal of Business and Economic Statistics 16, 261–283.

Mikosch, T., Stărică, C., 1999. Change of structure in financial time series, long range dependence and the GARCH Model. Manuscript, Department of Statistics, University of Pennsylvania.
Parke, W.R., 1999. What is fractional integration? Review of Economics and Statistics 81, 632–638.
Perron, P., 1989. The great crash, the oil price shock and the unit root hypothesis. Econometrica 57, 1361–1401.
Robinson, P.M., 1994a. Time series with strong dependence. In: Sims, C.A. (Ed.), Advances in Econometrics: Sixth World Congress, Vol. 1. Cambridge University Press, Cambridge, UK.
Robinson, P.M., 1994b. Semiparametric analysis of long-memory time series. Annals of Statistics 22, 515–539.
Robinson, P.M., 1995. Log-periodogram regression of time series with long-range dependence. Annals of Statistics 23, 1048–1072.
Rydén, T., Teräsvirta, T., Åsbrink, S., 1998. Stylized facts of daily return series and the hidden markov model. Journal of Applied Econometrics 13, 217–244.
Samorodnitsky, G., Taqqu, M.S., 1994. Stable Non-Gaussian Random Processes. Chapman and Hall, London.
Stock, J.H., 1994. Unit roots and trend breaks, In: Engle, R.F., McFadden, D. (Eds.), Handbook of Econometrics, Vol. IV. North-Holland, Amsterdam.
Stock, J.H., Watson, M.W., 1996. Evidence on structural instability in macroeconomic time series relations. Journal of Business and Economic Statistics 14, 11–30.
Taqqu, M.S., Willinger, W., Sherman, R., 1997. Proof of a fundamental result in self-similar traffic modeling. Computer Communication Review 27, 5–23.
Teverovsky, V., Taqqu, M.S., 1997. Testing for long-range dependence in the presence of shifting means or a slowly-declining trend, using a variance-type estimator. Journal of Time Series Analysis 18, 279–304.
Timmermann, A., 2000. Moments of markov switching models. Journal of Econometrics 96, 75–111.