Long Memory and Structural Change

Francis X. Diebold
Stern School, NYU
University of Pennsylvania
and NBER

Atsushi Inoue
North Carolina State University

May 1999
This Draft/Print: November 8, 1999

Correspondence to:
F.X. Diebold
Department of Finance
Stern School of Business
New York University
44 West 4th Street, K-MEC, Suite 9-190
New York, NY 10012-1126
fdiebold@stern.nyu.edu

Abstract: The huge theoretical and empirical econometric literatures on long memory and on structural change have evolved largely independently, as the phenomena appear distinct. We argue, in contrast, that they are intimately related. In particular, we show analytically that stochastic regime switching is easily confused with long memory, 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 asymptotic theory in finite samples and produces additional insights.

Acknowledgments: This paper grew from conversations with Bruce Hansen and Jim Stock at the New England Complex Systems Institute's International Conference on Complex Systems, September 1997. For helpful comments we thank participants in the June 1999 LSE Financial Markets Group Conference on the Limits to Forecasting in Financial Economics and the October 1999 Cowles Foundation Conference on Econometrics, as well as Torben Andersen, Tim Bollerslev, Rohit Deo, Steve Durlauf, Clive Granger, Cliff Hurvich, Adrian Pagan, Fallaw Sowell, and Mark Watson. None of those thanked, of course, are responsible in any way for the outcome. The National Science Foundation provided research support.

1. Introduction

Motivated by early empirical work in macroeconomics (e.g., Diebold and Rudebusch, 1989, Sowell, 1992) and later empirical work in finance (e.g., Ding, Engle, and Granger, 1993; Andersen and Bollerslev, 1997), the last decade has witnessed a renaissance in the econometrics of long memory and fractional integration, as surveyed for example by Baillie (1996).

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 literature on unit roots, structural change, and interrelationships between the two.

The recent more general long-memory literature, in contrast, pays comparatively little attention to confusing long memory and structural change. It is telling, for example, that the otherwise masterful surveys by Robinson (1994a), Beran (1994) and Baillie (1996) don't so much as mention the issue. The possibility of confusing long memory and structural change has of course arisen occasionally, in a number of literatures including applied hydrology (Klemeš, 1974), econometrics (Hidalgo and Robinson, 1996, Lobato and Savin, 1997), and mathematical statistics (Bhattacharya, Gupta and Waymire, 1983, Künsch, 1986, Teverovsky and Taqqu, 1997), but those warnings have had little impact. We can only speculate as to the reasons, but they are probably linked to the facts that (1) simulation examples such as Klemeš (1974) are interesting, but they offer neither theoretical justification nor Monte Carlo evidence, and (2) theoretical work such as Bhattacharya, Gupta and Waymire (1983) often seems highly abstract and lacking in intuition.

In this paper we provide both rigorous theory and Monte Carlo evidence to support the claim that long memory and structural change are easily confused, all in the context of simple and intuitive econometric models. In Section 2, we set the stage by considering alternative definitions of long memory and the relationships among them, and we motivate the definition of long memory that we shall adopt. In addition, we review the mechanisms for generating long memory that have been stressed previously in the literature, which differ rather sharply 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 how and when they produce realizations that appear to have long memory. In Section 4 we present an extensive finite-sample Monte Carlo analysis, which verifies the predictions of the theory and produces additional insights. We conclude in Section 5.

2. Long Memory

Here we consider alternative definitions of long memory and the relationships among them, we elaborate upon the definition that we adopt, and we review the mechanisms for generating long memory that have been stressed in the literature.

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 is

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

and a low-frequency spectral definition of long memory is

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

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

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

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 final definition of long memory involves the rate of growth of variances of partial sums,

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

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 definition of long memory (and hence also the autocorrelation definition of long memory). In particular, because the spectral density at frequency zero is the limit of $\frac{1}{T} S_T$, a 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. Hence the variance-of-partial-sum definition of long memory is quite general, and we shall make heavy use of it in our theoretical analysis in Section 3, labeling a series as I(d) if $\mathrm{var}(S_T) = O(T^{2d+1})$.[1]

Origins

There is a natural desire to understand the nature of various mechanisms that could generate long memory. Most econometric attention has focused on the role of aggregation. Here we briefly review two such aggregation-based perspectives on long memory, in order to contrast them to our subsequent perspective, which is quite different.

First, following Granger (1980), consider the aggregation of i = 1, ..., N cross-sectional

[1] Sowell (1990) also makes heavy use of this insight in a different context.

units

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

where ε_{it} is white noise, $\varepsilon_{it} \perp \varepsilon_{jt}$, and $\alpha_i \perp \varepsilon_{jt}$ for all i, j, t. As $N \rightarrow \infty$, the spectrum of the aggregate

$x_t = \sum_{i=1}^{N} x_{it}$ can be approximated as

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

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

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

then the τ-th autocovariance of x_t is

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

Thus $x_t \sim I\left(1 - \frac{q}{2}\right)$.

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

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, Willinger and Sherman, 1997). It has found wide application in the modeling of aggregate traffic on computer networks, although the basic idea is much more widely applicable. Define the stationary continuous-time binary series $W(t), t \geq 0$ so that $W(t) = 1$ during "on" periods and $W(t) = 0$ during "off" periods. The lengths of the on and off periods are iid at all leads and lags, and on and off periods alternate. Consider M sources, $W^{(m)}(t)$, $t \geq 0$, m = 1, ..., M, and define the aggregate packet count in the interval [0, Tt] by

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

Let $F_1(x)$ denote the c.d.f. of durations of on periods, and let $F_2(x)$ denote that for off periods, and assume that

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

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

Note that the power-law tails imply infinite variance. Now first let $M \to \infty$ and then let $T \to \infty$. Then it can be shown that $W_M^*(Tt)$, appropriately standardized, converges to a fractional Brownian motion.

Parke (1999) considers a closely related discrete-time error duration model, $y_t = \sum_{s=-\infty}^{t} g_{s,t}\varepsilon_s$,

where $\varepsilon_t \sim \text{iid}(0,\sigma^2)$, $g_{s,t} = 1(t \le s + n_s)$, and n_s is a stochastic duration. Long memory arises when n_s has infinite variance. The Mandelbrot-Taqqu-Parke approach beautifully illustrates the intimate connection between long memory and heavy tails, echoing earlier work summarized in Samorodnitsky and Taqqu (1993).[2]

3. Long Memory and Stochastic Regime Switching: Asymptotic Analysis

We now explore a very different route to long memory – structural change. Structural change is likely widespread in economics, as forcefully argued by Stock and Watson (1996). There are of course huge econometric literatures on testing for structural change, and on estimating models of structural change or stochastic parameter variation.

A Mixture Model with Constant Break Size, and Break Probability Dropping with T

We will show that a mixture model with a particular form of mixture weight linked to sample size will appear to have I(d) behavior. Let

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

where $w_t \overset{iid}{\sim} N(0,\sigma_w^2)$. Note that $\text{var}\left(\sum_{t=1}^{T} v_t\right) = pT\sigma_w^2 = O(T)$. If instead of forcing p to be

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

constant we allow it to change appropriately with sample size, then we can immediately obtain a long memory result. In particular, we have

Proposition 1: If $p = O(T^{2d-2})$, $0<d<1$, then $v_t = I(d-1)$.

Proof:

$$\operatorname{var}\left(\sum_{t=1}^{T} v_t\right) = O(T^{2d-2})T\sigma_w^2 = O(T^{2d-1}) = O(T^{2(d-1)+1}).$$

Hence $v_t = I(d-1)$, by the variance-of-partial sum definition of long memory. □

It is a simple matter 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 $w_t \overset{iid}{\sim} N(0,\sigma_w^2)$. Note that $\operatorname{var}\left(\sum_{t=1}^{T} v_t\right) = pT\sigma_w^2 = O(T)$. As before, let $p = O(T^{2d-2})$, $0<d<1$, so that $v_t = I(d-1)$, which implies that $\mu_t = \sum_{t=1}^{T} v_t = I(d)$.

It is also a simple matter 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 $w_t \overset{iid}{\sim} N(0,\sigma_w^2)$ and $\varepsilon_t \overset{iid}{\sim} N(0,\sigma_\varepsilon^2)$, which will display the same long memory property when $p = O(T^{2d-2})$, $0<d<1$.

Many additional generalizations could of course be entertained. The Balke-Fomby (1989) model of infrequent permanent shocks, for example, is a straightforward generalization of the simple mixture models 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.

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 $\varepsilon_t \overset{iid}{\sim} N(0,\sigma_\varepsilon^2)$. 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

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

$\varepsilon_t \overset{iid}{\sim} N(0,\sigma_\varepsilon^2)$, and $w_t \overset{iid}{\sim} N(0,\sigma_w^2)$. They note, moreover, that although other approximations 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 allow γ to change with T and write:

$$y_t = \mu_t + \varepsilon_t$$

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

This process will appear fractionally integrated under certain conditions, the key element of which involves the nature of time variation in γ.

Proposition 2: If (a) $E(\varepsilon_t^6)<\infty$ and (b) $\gamma_T\rightarrow\infty$ as $T\rightarrow\infty$ and $\gamma_T=O(T^{\delta})$ for some $\delta>0$, then

$y=I(1-\delta)$.

Proof: Note that

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

where the second equality follows from the maintained assumption that $\varepsilon_t \overset{iid}{\sim} N(0,\sigma_\varepsilon^2)$, the fourth equality follows from Assumption (a), and the fifth from Assumption (b). Thus, $\Delta y_t=I(-\delta)$, so

$y_t=I(1-\delta)$. □

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).

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 the Markov-switching model of Hamilton (1987), which is a rich dynamic mixture model.

Let $\{s_t\}_{t=1}^{T}$ be the (latent) sample path of two-state first-order autoregressive process, taking just the two values 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 ij-th 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 $\varepsilon_t \overset{iid}{\sim} N(0,\sigma^2)$ and s_t and ε_τ are independent for all t and τ.[3]

[3] In the present example, only the mean switches across states. We could, of course, examine richer models with Markov switching dynamics, but the simple model used here

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 $y = I(\min(\delta_0, \delta_1))$.

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

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

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 *j*th autocovariance matrix of ξ_t is[4]

illustrates the basic idea.

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

$$\Gamma_j = M^j \Gamma_0$$

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

$$= \frac{1}{c_0 T^{-\delta_0} + c_1 T^{-\delta_1}} \begin{bmatrix} 1-p_{11} & 1-p_{11} \\ 1-p_{00} & 1-p_{00} \end{bmatrix} + \frac{1}{c_0 T^{-\delta_0} + c_1 T^{-\delta_1}} \begin{bmatrix} 1-p_{00} & -(1-p_{11}) \\ -(1-p_{00}) & 1-p_{11} \end{bmatrix} \lambda^j$$

$$= O(T^{\min(\delta_0,\delta_1)}) + O(T^{\min(\delta_0,\delta_1)\lambda^j}).$$

Thus

$$\frac{1}{T}\mathrm{var}(\textstyle\sum_{t=1}^{T} y_t) = O(1) + O\left(\frac{T^{\min(\delta_0,\delta_1)}}{1-\lambda}\right) = O\left(T^{2\min(\delta_0,\delta_1)}\right),$$

which in turn implies that

$$\mathrm{var}(\textstyle\sum_{t=1}^{T} y_t) = O\left(T^{2\min(\delta_0,\delta_1)+1}\right).$$

Hence $y = I(\min(\delta_0,\delta_1))$. □

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).[5]

[5] 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 pointed out by Timmermann (1999), and

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

Additional Discussion

Before proceeding, it is important to appreciate what we do and what we don't do, and how we use the theory sketched above. Our device of letting certain parameters such as mixture probabilities vary with T is simply a thought experiment, which we hope proves useful for thinking about the appearance of long memory. The situation parallels the use of "local-to-unity" asymptotics in autoregressions, a thought experiment that proves useful for characterizing the distribution of the dominant root. We view our theory as effectively providing a "local to no breaks" perspective. 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 $\rho = 1 - \frac{c}{T}$ 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 our subsequent Monte Carlo analysis, 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

---

as verified in our subsequent Monte Carlo.

infinity. Instead, our interest centers on fixed-parameter models in fixed finite sample sizes, the dynamics of which are in fact either I(0) or I(1). The theory suggests that confusion with I(d) will result when only a small amount of breakage occurs, and it suggests that the larger the sample size, the smaller must be the break probability, in order to maintain the necessary small amount of breakage.

In short, we use the theory to guide our thinking about whether and when finite-sample paths of truly I(0) or I(1) processes might nevertheless appear I(d), $0<d<1$, *not* as a device for producing sample paths of truly fractionally integrated processes.

4. Long Memory and Stochastic Regime Switching: Finite-Sample Analysis

The theoretical results in Section 3 suggest that, under certain plausible conditions amounting to a nonzero but "small" amount of structural change, long memory and structural change are easily confused. Motivated by our 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 estimation 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 j-th Fourier frequency, $\omega_j = 2\pi j/T$, $j = 1, 2, ..., [T/2]$. The log-periodogram estimator of d is then based on the least squares regression,

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

where $j = 1, 2, ..., m$, and $\hat{d} = -1/2\hat{\beta}_1$.[6] The least squares estimator of β_1, and hence $\hat{d}$, is asymptotically normal and the corresponding theoretical standard error, $\pi\cdot(24\cdot m)^{-\frac{1}{2}}$, depends only on the number of periodogram ordinates used.

Of course, the actual value of the estimate of 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 grow with sample size, but at a slower rate. Use of $m=\sqrt{T}$ has emerged as a popular rule of thumb, which we adopt.

Mean Plus Noise Model

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

$$y_t = \mu_t + \varepsilon_t$$

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

[6] 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.

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

where $\varepsilon_t \overset{iid}{\sim} N(0,1)$, and $w_t \overset{iid}{\sim} N(0,1)$, t = 1, 2, ..., T. To build intuition before proceeding to the Monte Carlo, we first show in Figure 1 a specific realization of the mean plus noise process with p=0.01 and T=10000. It is clear that there are only a few breaks, with lots of noise superimposed. In Figure 2 we plot the log periodogram against log frequency for the same realization, using $\sqrt{10000}$ periodogram ordinates. We superimpose the linear regression line; the implied d estimate is 0.785.

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 $T \in \{100, 200, 300, 400, 500, 1000, ..., 5000\}$. 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 power increases in T by consistency of the test.[7] Second, for fixed T, we have more power to detect I(1) behavior as p grows, because we have a greater number of non-zero innovations, whose effects we can observe.

The thesis of this paper is that structural change is 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,

[7] 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.

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 Figure 3 we plot kernel estimates of the density of $\hat{d}$ for $T \in \{400,1000,2500,5000\}$ and $p \in \{0.0001,0.0005,0.001,0.005,0.01,0.05,0.1\}$.[8] Figure 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 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.

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 $\varepsilon_t \overset{iid}{\sim} N(0,1)$. In Figure 4 we show a specific realization of the STOPBREAK process with $\gamma = 500$ and T=10000. Because the evolution of the STOPBREAK process is smooth, as it is

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

only an *approximation* to a mixture model, we do not observe sharp breaks in the realization. In Figure 5 we plot the realization's log periodogram against log frequency, with linear regression line superimposed, using $\sqrt{10000}$ periodogram ordinates. The implied d estimate is 0.67.

In the Monte Carlo experiment, we examine all pairs of $\gamma \in \{10^{-5}, 10^{-4}, ..., 10^{3}, 10^{4}\}$ and $T \in \{100,200,300,400,500,1000,1500,...,5000\}$. In Table 3 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 gamma grows, because we have a smaller effective sample size.[9] 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 4 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.

In Figure 6 we show kernel estimates of the density of $\hat{d}$ for $T \in \{400,1000,2500,5000\}$

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

and $\gamma \in \{10^{-5}, 10^{-4}, ..., 10^3, 10^4\}$. As the sample size grows, the estimated density shifts to the right and the median of $\hat{d}$ approaches unity for $\gamma < 10000$. 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.

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 $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$, and s_t and ε_τ are independent for all t and τ. We take $\mu_0 = 0$ and $\mu_1 = 1$.

In Figure 7 we plot a specific realization with $p_{00} = p_{11} = 0.9995$ and T=10000, and in Figure 8 we plot the corresponding log periodogram against log frequency, with the linear regression line superimposed, using $\sqrt{10000} = 100$ periodogram ordinates. It appears that the regime has changed several times in this particular realization, and the implied d estimate is 0.616, which is consistent with our theory in Section 3.

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 $T \in \{100, 200, 300, 400, 500, 1000, 1500, ..., 5000\}$. In Table 5 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 both 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 6 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 Figure 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 not change with positive probability and, as a result, the estimated densities appear bimodal.

Finally, we contrast our results for the Markov switching model with those of Rydén, Teräsvirta and Åsbrink (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, Teräsvirta and Åsbrink have at least one diagonal element well away from unity. Only their estimated model H satisfies our condition, and its autocorrelation function does in fact decay very slowly. Hence the results are entirely consistent.

5. Concluding Remarks

We have argued that structural change in general, and stochastic regime switching in particular, are intimately related and easily confused, so long as only a small amount of regime switching occurs. Simulations support the relevance of the theory in finite samples and make clear that the confusion is not merely a theoretical curiosity, but rather is likely to be relevant in routine empirical economic and financial applications.

We close by sketching the relationship of our work to two close cousins by Granger and his coauthors. First, Granger and Teräsvirta (1999) consider the following simple nonlinear process,

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

where $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$. 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 and independent work, Granger and Hyung (1999) develop a theory closely related to ours. They consider a mean plus noise process and its Markov switching version, and they show that, if $p=O(1/T)$, the autocorrelations of the mean plus noise process decay very slowly. 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 long memory, and we provide a link between the convergence rate of p and the 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. Hence the two papers are complements rather than substitutes.

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.[10] 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" are of dubious value.

[10] 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.

References

Andersen, T.G. and Bollerslev, T. (1997), "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns," *Journal of Finance*, 52, 975-1005.

Baillie, R.T. (1996), "Long-Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics*, 73, 5-59.

Balke, N.S. and Fomby, T.B. (1989), "Shifting Trends, Segmented Trends, and Infrequent Permanent Shocks," *Journal of Monetary Economics*, 28, 61-85.

Balke, N.S. and Fomby, T.B. (1997), "Threshold Cointegration," *International Economic Review*, 38, 627-645.

Bhattacharya, R.N., Gupta, V.K. and Waymire, E. (1983), "The Hurst Effect Under Trends," *Journal of Applied Probability*, 20, 649-662.

Beran, J. (1994), *Statistics for long-memory processes*. New York: Chapman and Hall.

Chambers, M. (1998), "Long Memory and Aggregation in Macroeconomic Time Series," *International Economic Review*, 39, 1053-1072.

Cioczek-Georges, R. and Mandelbrot, B.B. (1995), "A Class of Micropulses and Antipersistent Fractional Brownian Motion," *Stochastic Processes and Their Applications*, 60, 1-18.

Clements, M.P. and 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.

Diebold, F.X. and Rudebusch, G.D. (1989), "Long Memory and Persistence in Aggregate Output," *Journal of Monetary Economics*, 24, 189-209.

Diebold, F.X. and Rudebusch, G.D. (1996), "Measuring Business Cycles: A Modern Perspective," *Review of Economics and Statistics*, 78, 67-77.

Ding, Z., Engle, R.F. and Granger, C.W.J. (1993), "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83-106.

Engle, R.F. and Smith, A.D. (1999), "Stochastic Permanent Breaks," *Review of Economics and Statistics*, 81, forthcoming.

Geweke, J. and 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. and Hyung, N. (1999), "Occasional Structural Breaks and Long Memory," Discussion Paper 99-14, University of California, San Diego.

Granger, C.W.J., and 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: Princeton University Press.

Heyde, C.C. and Yang, Y. (1997), "On Defining Long Range Dependence," *Journal of Applied Probability*, 34, 939-944.

Hidalgo, J. and Robinson, P.M. (1996), "Testing for Structural Change in a Long Memory Environment," *Journal of Econometrics*, 70, 159-174.

Hurvich, C.M. and 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. and 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. and 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. and 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, forthcoming.

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 C.A. Sims (ed.), *Advances in Econometrics: Sixth World Congress Vol.1*. Cambridge, UK: Cambridge University Press.

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. and Åsbrink, S. (1998), "Stylized Facts of Daily Return Series and the Hidden Markov Model," *Journal of Applied Econometrics*, 13, 217-244.

Samorodnitsky, G. and Taqqu, M.S. (1993), *Stable Non-Gaussian Random Processes*. London: Chapman and Hall.

Sowell, F. (1990), "The Fractional Unit Root Distribution," *Econometrica*, 50, 495-505.

Sowell, F.B. (1992), "Modeling Long-Run Behavior with the Fractional ARIMA Model," *Journal of Monetary Economics*, 29, 277-302.

Stock, J.H. (1994), "Unit Roots and Trend Breaks," in R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics, Volume IV.* Amsterdam: North-Holland.

Stock, J.H. and 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. and Sherman, R. (1997), "Proof of a Fundamental Result in Self-Similar Traffic Modeling," *Computer Communication Review*, 27, 5-23.

Teverovsky, V. and 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. (1999), "Moments of Markov Switching Models," *Journal of Econometrics*,

forthcoming.

Table 1
Mean Plus Noise Model
Empirical Sizes of Nominal 5% Tests of d=0

	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

Notes to table: T denotes sample size, and p denotes the mixture probability. We report the fraction of 10000 trials in which inference based on the Geweke-Porter-Hudak procedure led 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

	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

Notes to table: T denotes sample size, and p denotes the mixture probability. We report the fraction of 10000 trials in which inference based on the Geweke-Porter-Hudak procedure led to rejection of the hypothesis that d=1, using a nominal 5% test based on $\sqrt{T}$ periodogram ordinates.

Table 3
Stochastic Permanent Break Model
Empirical Sizes of Nominal 5% Tests of d=0

	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

Notes to table: T denotes sample size, and γ denotes the STOPBREAK parameter. We report the fraction of 10000 trials in which inference based on the Geweke-Porter-Hudak procedure led to rejection of the hypothesis that d=0, using a nominal 5% test based on $\sqrt{T}$ periodogram ordinates.

Table 4
Stochastic Permanent Break Model
Empirical sizes of nominal 5% tests of d=1

	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

Notes to table: T denotes sample size, and γ denotes the STOPBREAK parameter. We report the fraction of 10000 trials in which inference based on the Geweke-Porter-Hudak procedure led to rejection of the hypothesis that d=1, using a nominal 5% test based on $\sqrt{T}$ periodogram ordinates.

Table 5
Markov Switching Model
Empirical Sizes of Nominal 5% Tests of d=0

	p_{00}	**0.95**	**0.95**	**0.95**	**0.99**	**0.99**	**0.99**	**0.999**	**0.999**	**0.999**
T	p_{11}	**0.95**	**0.99**	**0.999**	**0.95**	**0.99**	**0.999**	**0.95**	**0.99**	**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

Notes to table: T denotes sample size, and p_{00} and p_{11} denote the Markov staying probabilities. We report the fraction of 10000 trials in which inference based on the Geweke-Porter-Hudak procedure led to rejection of the hypothesis that d=0, using a nominal 5% test based on $\sqrt{T}$ periodogram ordinates.

Table 6
Markov Switching Model
Empirical Sizes of Nominal 5% Tests of d=1

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

Notes to table: T denotes sample size, and p_{00} and p_{11} denote the Markov staying probabilities. We report the fraction of 10000 trials in which inference based on the Geweke-Porter-Hudak procedure led to rejection of the hypothesis that d=1, using a nominal 5% test based on $\sqrt{T}$ periodogram ordinates.

Figure 1
Mean-Plus-Noise Realization

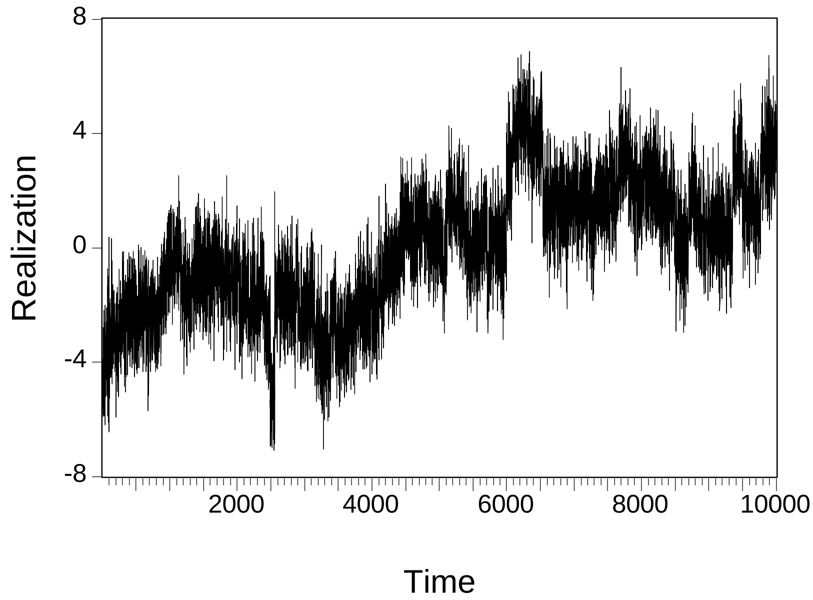

Figure 2
Mean-Plus-Noise Model
Low-Frequency Log Periodogram

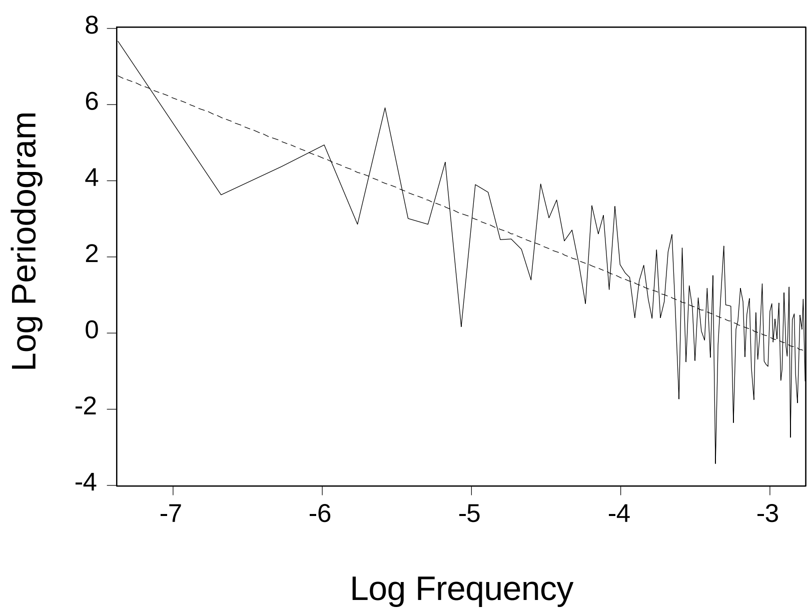

Figure 3
Mean-Plus-Noise Model
Distribution of Long-Memory Parameter Estimate

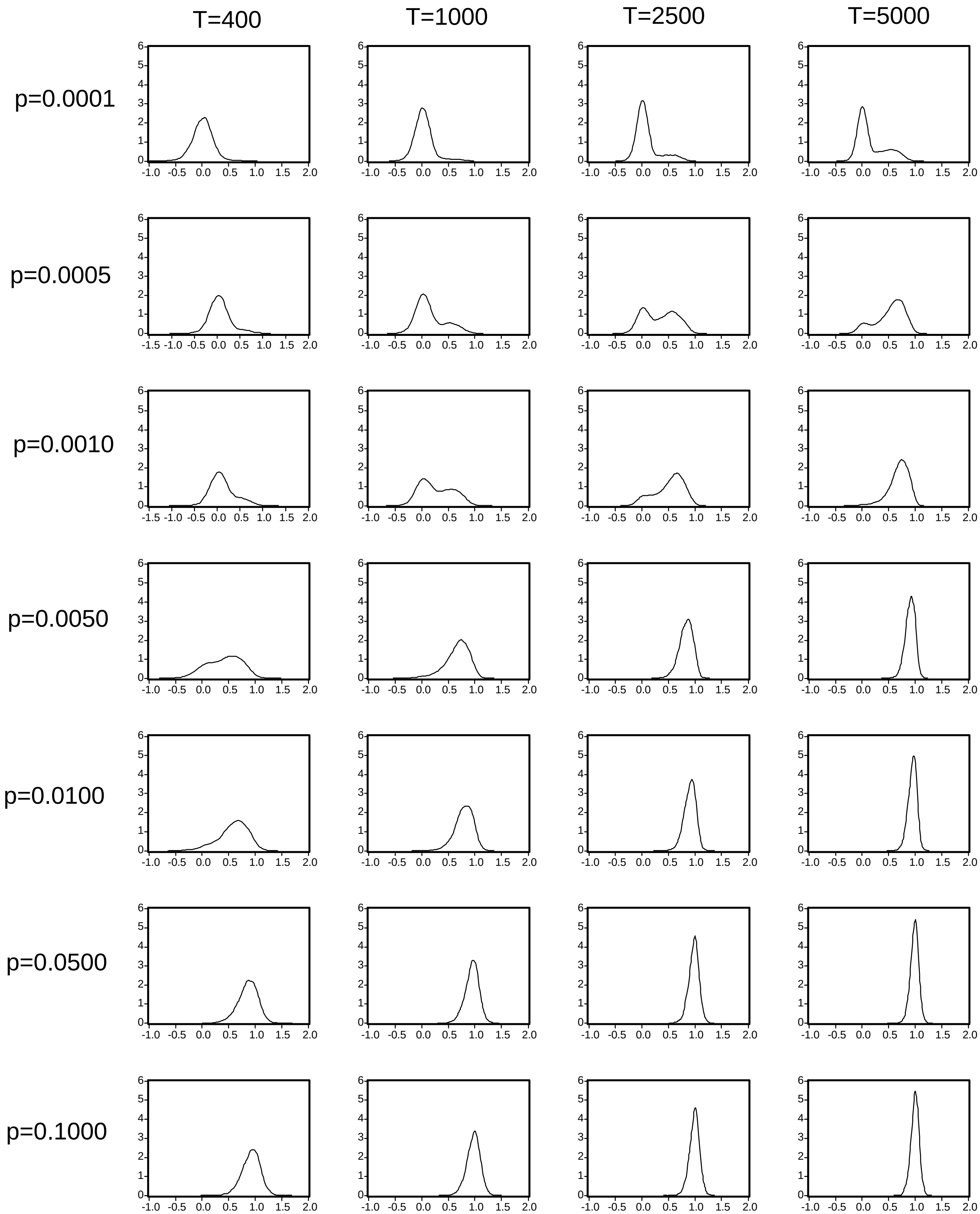

Figure 4
STOPBREAK Realization

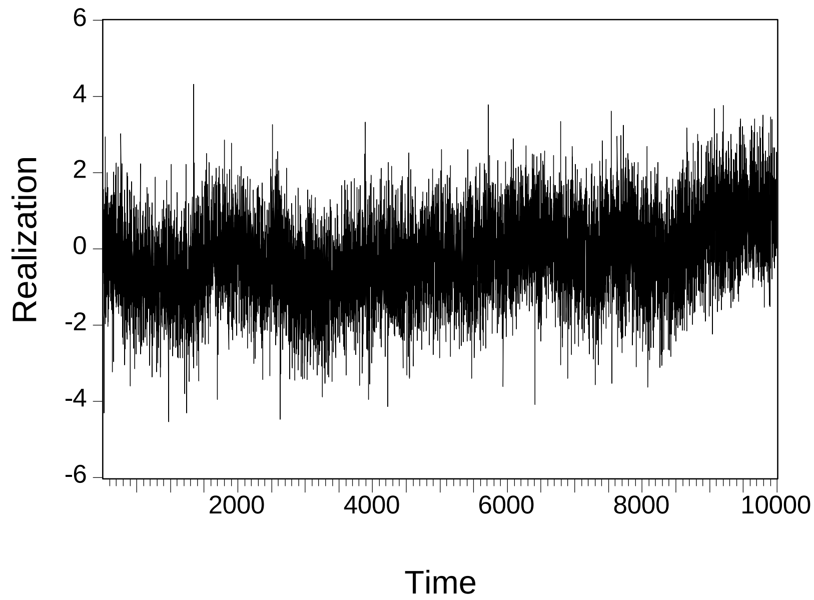

Figure 5
STOPBREAK Model
Low-Frequency Log Periodogram

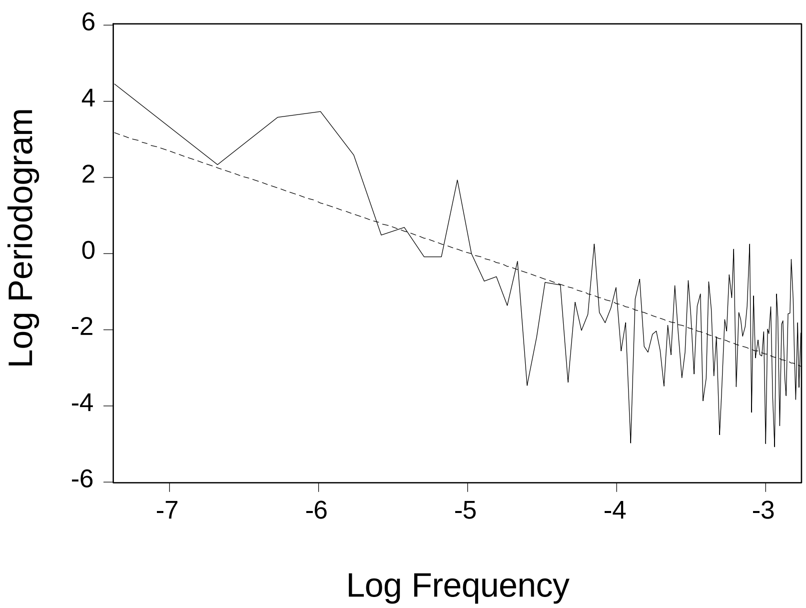

Figure 6
STOPBREAK Model
Distribution of Long-Memory Parameter Estimate

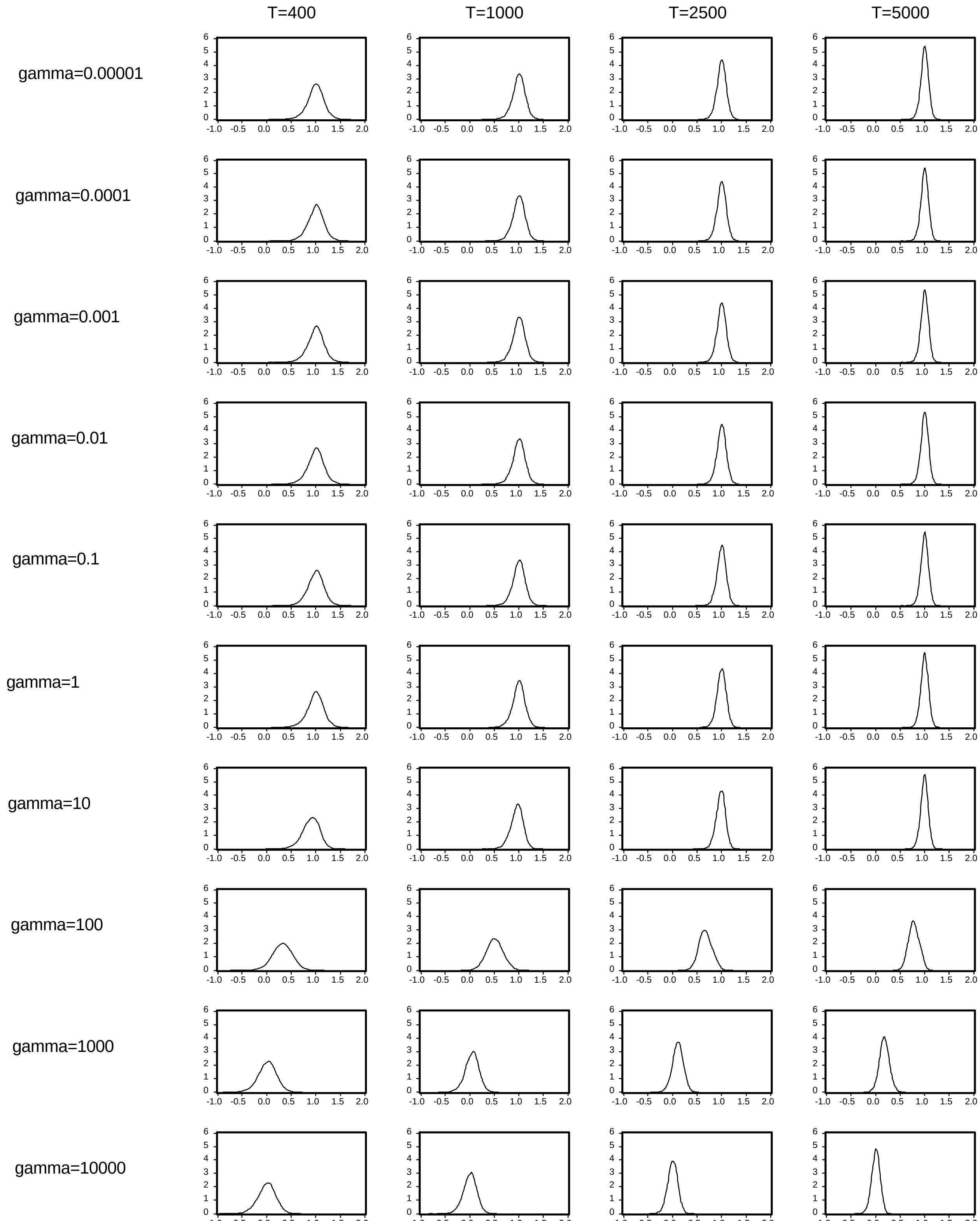

Figure 7
Markov Switching Realization

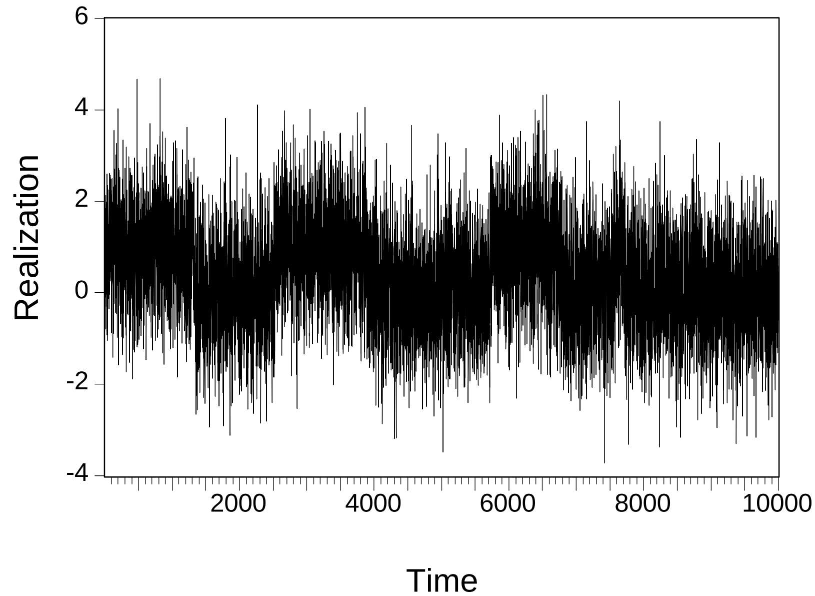

Figure 8
Markov Switching Model
Low-Frequency Log Periodogram

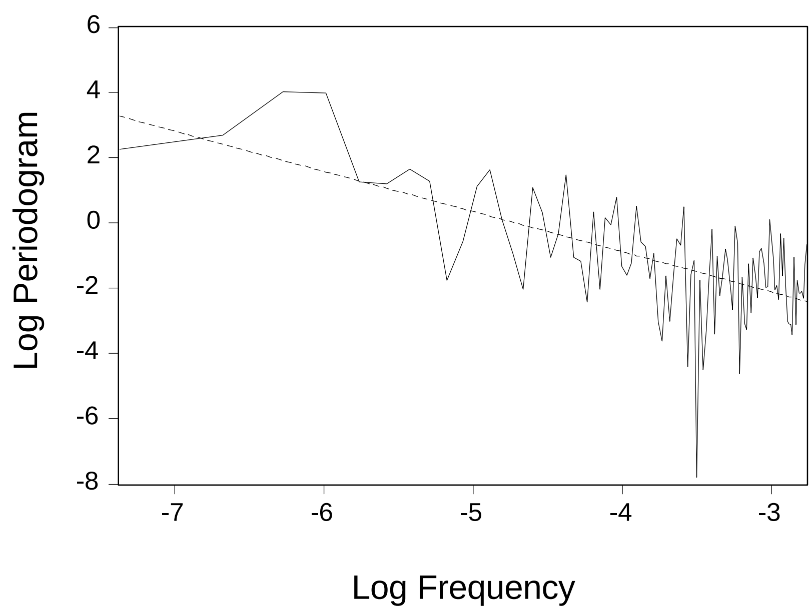

Figure 9
Markov Switching Model
Distribution of Long-Memory Parameter Estimate