



ELSEVIER

Journal of Econometrics 73 (1996) 61–77

**JOURNAL OF
Econometrics**

Varieties of long memory models

Clive W.J. Granger^{*,a}, Zhuangxin Ding^b

^a*Department of Economics, University of California, San Diego, La Jolla, CA 92093-0508, USA*

^b*Frank Russell Company, Tacoma, WA 98401, USA*

Abstract

Long memory is defined as a series having a slowly declining correlogram or, equivalently, an infinite spectrum at zero frequency. Fractional integrated processes have such properties but here it is pointed out that a number of other processes can also be long memory, including generalized fractionally integrated models arising from aggregation, time-changing coefficient models, and possibly nonlinear models. It seems that there are many classes of processes that deserve further study. The relevance of long memory is illustrated using absolute returns from a daily stock market index.

Key words: Fractional integration; Long memory; Nonlinear time series

JEL classification: C22; B23

1. The concept of long memory

The correlogram, the plot of the estimated autocorrelation between x_t and x_{t-k} against k , is a useful device for describing some of the linear properties of a single series in most cases. If the plot either declines exponentially or, alternatively, declines very slowly, Box and Jenkins (1970) suggested certain models as possible candidates for the generating mechanism of the series, either stationary ARMA or integrated ARIMA. However, not all correlogram plots have exactly these shapes and so new classes of models have to be derived as possible generating mechanisms. Sometimes the correlogram declines steadily but not exponentially, or it may start with $\hat{\rho}_1 = 0.4$, say, and then declines only slowly from this values. Amongst the models that have been suggested to generate

*Corresponding author.

The research was supported by NSF grants SES 90-23037 and 93-08295.

series having such properties are the fractionally integrated $I(d)$ models which are discussed below, together with some generalizations. It is important to remember that because a model produces series having a certain property P and one finds that the data also has P , then it is not correct to conclude that the model corresponds to the actual generating mechanism of the data. The model is a candidate for this generating mechanism, but other models may also produce series having P . This problem is enhanced by the fact that we never see the actual autocorrelations but only their estimates, and generally several shapes can fall within the confidence bands around these estimates.

In this paper a property called ‘long memory’ is defined, and then a variety of processes having this property will be discussed, including generalizations of fractionally integrated processes.

Let $h_j, j = 0, 1, \dots$, be an infinite sequence of constants and define a time-changing sequence of filters

$$h_t(B) = \sum_{j=0}^t h_j B^j, \quad (1.1)$$

where B is the backward operator. A series x_t can be generated by

$$x_t = h_t(B)e_t, \quad (1.2)$$

where the input series is assumed to have the form

$$e_t = m + \varepsilon_t, \quad (1.3)$$

where ε_t is a zero mean, constant variance white noise, so that $\text{corr}(\varepsilon_t, \varepsilon_s) = 0$, $t \neq s$, and with $\text{var}(\varepsilon) = \sigma_\varepsilon^2$. Denote $m(t) = mh_t(1)$,

$$v(t) = \sigma_\varepsilon^2 \sum_{j=0}^t h_j^2, \quad (1.4)$$

$$s_t = \sum_{j=0}^t h_j \varepsilon_{t-j}, \quad (1.5)$$

and

$$f_N(\omega) = E \left[N^{-1} \left| \sum_{t=1}^N s_t e^{it\omega} \right|^2 \right]. \quad (1.6)$$

If $f(\omega) = \lim_N f_N(\omega)$ exists for all $0 \leq \omega$, $f(\omega)$ will be called the (power) spectrum of x_t .

If $h_t(1)$ is monotonically increasing, $m(t)$ and $v(t)$ will be the (deterministic) ‘trends’ in mean and variance of x_t respectively, although these quantities may tend to constants. The properties of these trends are discussed in Granger (1988).

If t is large, so that $f_t(\omega)$ is a close approximation to $f(\omega)$ for almost all ω 's, the series x_t has been called 'long memory' if its spectrum has the following properties:

- (i) $f(\omega)$ tends to infinity as $\omega \rightarrow 0$

and

- (ii) $f(\omega)$ is bounded above for all but, at most, a finite number of other ω values.

A necessary condition for (i) is that h_j is a divergent sequence. Examples of series having these properties are discussed in the following sections. If $h_j = A_j^{d-1}$, the series generated is (effectively) the $I(d)$ process discussed by Granger and Joyeux (1980) and Hosking (1981). This process has spectrum which is proportional to ω^{-2d} for ω small. If d is not an integer, one has a fractionally integrated process. It is called 'long memory' because a shock ε_t at time t continues to influence future x_{t+k} for a longer horizon k than would be the case for the standard stationary ARMA process for which h_j declines exponentially as θ^j with $|\theta| < 1$.

There are few practical implications of a series having the long-memory property except that quantities such as the sample mean converge slowly for $d < 1/2$ and are not consistent for $d \geq 1/2$.

An alternative definition, which attempts to capture a more general feature of data, may be called 'extended memory'. Let $g_{h,n}(I_n) = I[x_{n+h}|x_{n-j}, j \geq 0]$ be the optimum least-squares forecast of x_{n+h} based on the information set $I_n: x_{n-j}, j \geq 0$. If, for all n , $g_{h,n}(I_n)$ does not tend to a constant as h becomes large, then x_t is said to have extended memory. The definition has more impact if deterministic processes such as chaotic and series with limit cycles are excluded from consideration. The two definitions are rather different, as long memory considers just the linear properties of the series, whereas extended memory can include optimum nonlinear forecasts. Some properties of extended memory process are considered in Granger (1995).

2. An example using stock market data

As an empirical example, let us look at the long memory property of S&P daily 500 stock market returns as discussed in Ding, Granger, and Engle (1993). In that paper, they found that although the returns themselves contain little serial correlation, the absolute value of returns has significantly positive serial correlation up to 2,700 lags which is more than 10 years! To illustrate this more clearly, Figs. 1 and 2 give the plot of stock market return r_t and absolute return $|r_t|$ from January 4, 1928 to August 30, 1990, with 17,054 daily observations. It is

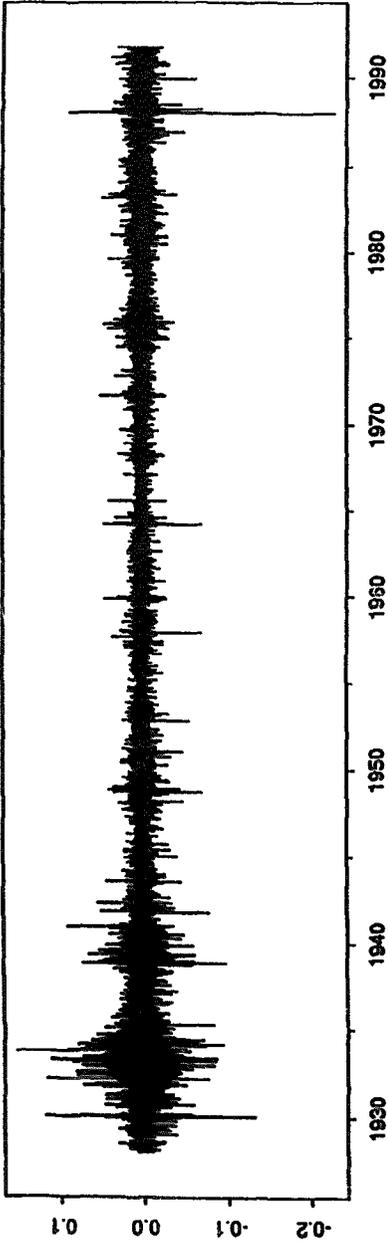


Fig. 1. Standard & Poor 500 daily returns, 01/04/28-08/30/91.

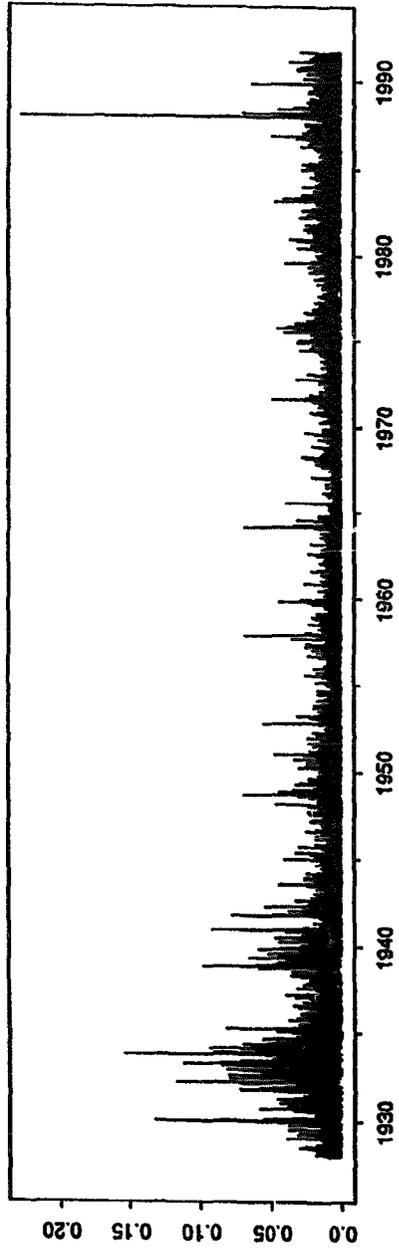


Fig. 2. Standard & Poor 500 daily absolute returns, 01/04/28-08/30/91.

seen that large absolute returns are more likely than small absolute returns to be followed by a large absolute return. The largest absolute return comes at the Black Monday stock market crash in 1987 which is more than 20%.

Fig. 3 gives the plot of $\rho_k(r_t) = \text{corr}(r_t, r_{t-k})$ and $\rho_k(|r_t|) = \text{corr}(|r_t|, |r_{t-k}|)$ up to lag 2,500. The dotted lines are 95% confidence interval for the autocorrelations of an i.i.d. process. It is seen that except for the first two lags $\rho_k(r_t)$ are very small, and most of them stay within the 95% confidence interval. But the autocorrelation structure for $|r_t|$ is totally different. Not only are the autocorrelations of $|r_t|$ all significantly different from zero, but also they are all positive. $\rho_k(|r_t|)$ decreases very fast for the first 40 lags or so, and then decreases very slowly, which suggests that an exponentially decreasing autocorrelation function like that of an ARMA model cannot produce the empirical autocorrelation pattern here. The dashed line in Fig. 3 is the curve fitting using such a autocorrelation function $\rho_k = \alpha\beta^k$. It is readily seen that the fitted curve decreases too slowly at the beginning, and then too fast at the end.

The discussion in this paper suggests that the autocorrelation structure of the fractionally integrated model might be able to give a good approximation to the sample autocorrelation found here. Fig. 4 again plots $\rho_k(|r_t|)$ for $k = 1, \dots, 2,500$, and then the fitted autocorrelation function (dotted line) of a fractionally integrated series

$$\rho_\tau = \frac{\Gamma(1-d)}{\Gamma(d)} \frac{\Gamma(\tau+d)}{\Gamma(\tau+1-d)},$$

where $d = 0.474$ is as estimated below. Comparing Fig. 3 and Fig. 4, it is seen that the shape of the autocorrelation function of a fractionally integrated model better fits the sample autocorrelations, although it has been scaled up for the reason mentioned below.

To estimate d , we assume $|r_t|$ is generated by the Fractionally Integrated ARCH model as in Granger and Ding (1996):

$$r_t = \sigma_t e_t, \quad e_t \sim \text{i.i.d. } (0, 1),$$

$$\sigma_t = (1 - (1 - B)^d) \frac{|r_t|}{\mu} = \frac{1}{\mu} \sum_{j=1}^{\infty} \frac{d}{\Gamma(1-d)} \cdot \frac{\Gamma(j-d)}{\Gamma(j+1)} |r_{t-j}|,$$

where $\mu = E|e_t|$ and $0 < d < 1/2$.

For this model one has $\text{corr}(r_t, r_{t-k}) = 0$ for all $k > 0$ but $\text{corr}(|r_t|, |r_{t-k}|) = \{[\Gamma(1-d)/\Gamma(d)]\} \{F(k+d)/\Gamma(k+1-d)\}$ which is the same as the correlation function of a fractionally integrated series. This can also be seen by rewriting the conditional standard deviation equation in the form

$$(1 - B)^d |r_t| = \sigma_t (|e_t| - \mu) = \varepsilon_t,$$

where $\varepsilon_t = \sigma_t (|e_t| - \mu)$ is a mean zero process with $\text{cov}(\varepsilon_t, \varepsilon_{t-k}) = 0$ for $k > 0$. For a more general class of models and a detailed discussion, see Ding (1994)

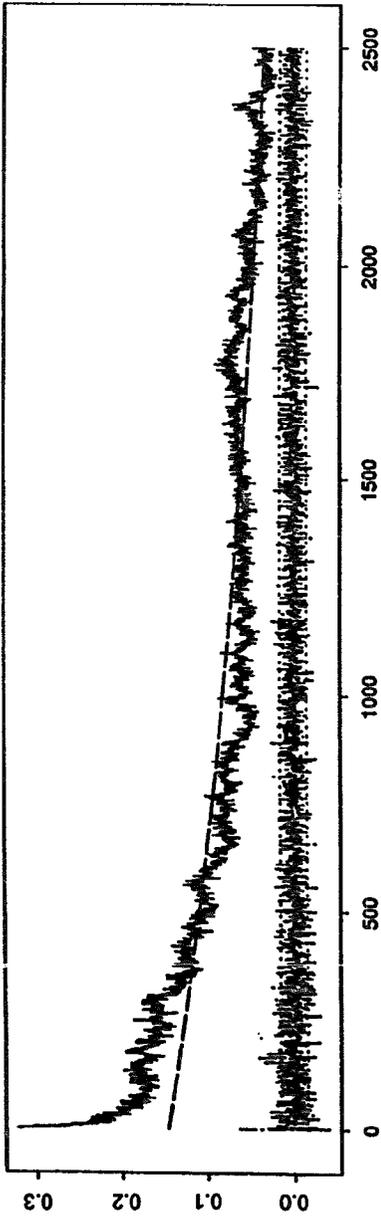


Fig. 3. Sample autocorrelations for r_t and its fitted value.

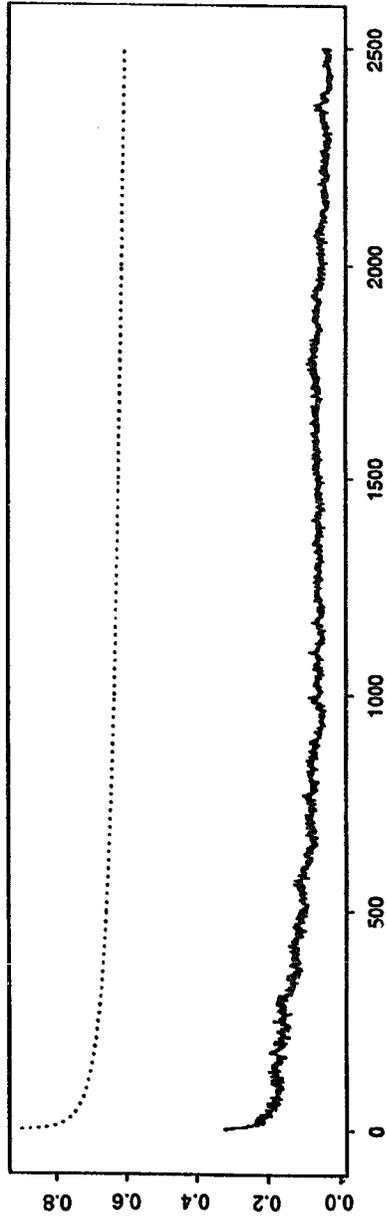


Fig. 4. Sample autocorrelations for $|r_t|$ and its fitted value.

and Ding and Granger (1996). The spectral density function of ε_t is denoted by $f_\varepsilon(\lambda)$ which is assumed to be finite, bounded away from zero and continuous on the interval $[-\pi, \pi]$. The spectral density function of $|r_t|$ is then

$$f(\lambda) = \frac{\sigma^2}{2\pi} \{4 \sin^2(\lambda/2)\}^{-d} f_\varepsilon(\lambda). \quad (2.5)$$

For simplicity, we used the estimation method of Geweke and Porter-Hudak (1983) (hereafter GPH) which is based on the simple linear regression of the log periodogram on a deterministic regressor since from (2.5) one has

$$\ln\{f(\lambda)\} = \ln\{\sigma^2 f_\varepsilon(0)/2\pi\} - d \ln\{4 \sin^2(\lambda/2)\} + \ln\{f_\varepsilon(\lambda)/f_\varepsilon(0)\}. \quad (2.6)$$

Suppose that a sample of $|r_t|$ of size T is available. Let $\lambda_{j,T} = 2\pi j/T$ ($j = 1, \dots, T-1$) denote the harmonic ordinates and $I(\lambda_{j,T})$ denote the periodogram at these ordinates, evaluate (2.6) at $\lambda_{j,T}$, and rearrange to obtain

$$\begin{aligned} \ln\{I(\lambda_{j,T})\} &= \ln\{\sigma^2 f_\varepsilon(0)/2\pi\} - d \ln\{4 \sin^2(\lambda_{j,T}/2)\} \\ &\quad + \ln\{f_\varepsilon(\lambda_{j,T})/f_\varepsilon(0)\} + \ln\{I(\lambda_{j,T})/f(\lambda_{j,T})\}. \end{aligned}$$

Geweke and Porter-Hudak (1983) show that, when attention is confined to frequencies near zero, $\ln\{f_\varepsilon(\lambda_{j,T})/f_\varepsilon(0)\}$ can be ignored and the differencing parameter d can be consistently estimated from the least squares regression

$$\ln\{I(\lambda_{j,T})\} = c - d \ln\{4 \sin^2(\lambda_{j,T}/2)\} + \mu_j, \quad j = 1, \dots, n, \quad (2.7)$$

where $n = g(T) \ll T$.

If the least squares estimator \hat{d} is significantly different from zero, then $|r_t|$ is fractionally integrated. In our estimation, we use $n = g(T) = \sqrt{T}$ and the Schuster periodogram which is defined as follows (see Granger and Newbold, 1986):

$$\begin{aligned} I(\lambda_{j,T}) &= \frac{1}{T^2} \left\{ \left[\sum_{t=1}^T (|r_t| - \bar{|r|}) \cos(\lambda_{j,T} t) \right]^2 \right. \\ &\quad \left. + \left[\sum_{t=1}^T (|r_t| - \bar{|r|}) \sin(\lambda_{j,T} t) \right]^2 \right\}. \end{aligned}$$

Here $T = 17,054$, so that $n = \sqrt{T} = 130$. The estimated result for $|r_t|$ is as follows:

$$\begin{aligned} \ln\{I(\lambda_{j,T})\} &= -21.505 - 0.474 \ln\{4 \sin^2(\lambda_{j,T}/2)\} + \mu_j, \quad R^2 = 0.34. \quad (2.8) \\ &\quad (-44.85) \quad (-8.15) \end{aligned}$$

The estimated d is 0.474 with a t -statistic of 8.15. It seems reasonable to say at this point that $|r_t|$ is a stationary long-memory process.

As mentioned earlier, when we use this estimated d to fit the sample autocorrelation, the fitted autocorrelation is much bigger than the actual one. This can occur if the GPH method overestimates the fractional integration parameter d since the autocorrelation function is an increasing function of d . If instead d is estimated directly by minimizing $\sum_{\tau=1}^{2500} [\rho_{\tau}(|r_{\tau}|) - \rho_{\tau}]^2$, one gets $\hat{d} = 0.358$ which gives an autocorrelation function as in Fig. 5 (dotted line). Comparing Fig. 5 with Figs. 3 and 4 it is seen that this model gives a much better approximation than those two models in terms of autocorrelation curve fitting. The major difference between ρ_{τ} and the sample autocorrelation occurs at the first 20 lags, which suggests some short-run effect is missing if use is only made of a fractionally integrated model.

The series $|r_{\tau}|$ is seen to have long memory in that $\text{corr}(|r_{\tau}|, |r_{\tau+k}|)$ remains positive and significant for many values of k . It is shown in a companion paper, Granger and Ding (1996), that if one defines

$$\rho_k(\theta) = \text{corr}(|r_{\tau}|^{\theta}, |r_{\tau-k}|^{\theta}),$$

then for a wide range of θ values the series $|r_{\tau}|^{\theta}$ is long memory, but

$$\rho_k(\theta) < \rho_k(1), \quad \text{all } k,$$

for θ both less than or greater than one, a result we call the Taylor effect.

The evident ‘long-memory’ property of daily absolute returns from a stock index might be explained as an $I(d)$, with d a fraction near but just below 0.5, as seen here. Results to be presented elsewhere suggest that absolute returns from other speculative markets have a similar property. What the remaining sections of this paper will emphasize is that the $I(d)$ model is not the only one available that produces the long-memory properties, and so it should not be accepted too readily. We will just discuss alternative specifications. As the estimation and choice of some of these alternative models needs elaboration, we do not try to fit them to the data used in this section, which should be viewed merely as an illustration of the existence of the long-memory property.

3. Generalized fractionally integrated processes

It is convenient to start with two definitions. If $f_1(\omega)$ and $f_2(\omega)$ are spectra corresponding to a pair of long-memory series, then x_{1t} is said to ‘dominate’ x_{2t} if

$$\lim_{\omega \rightarrow 0} \frac{f_1(\omega)}{f_2(\omega)} \rightarrow \infty. \quad (3.1)$$

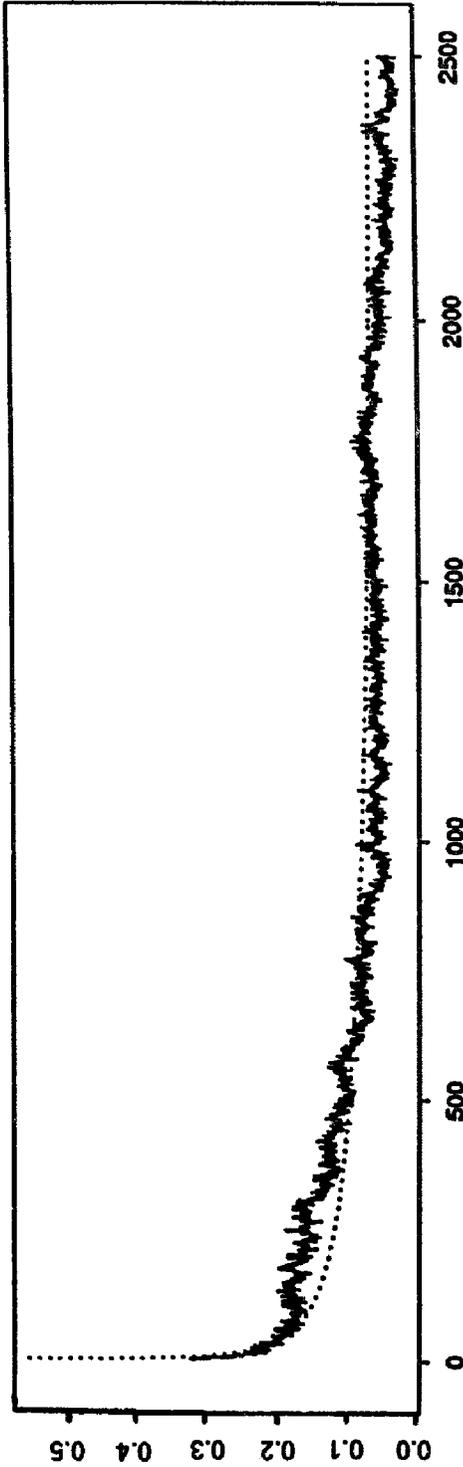


Fig. 5. Sample autocorrelations for $|r|$ and its adjusted fitted value.

Secondly, a positive measurable function $L(x)$ is 'slowly varying' (denoted *sv*) if for any $k \neq 0$,

$$\lim_{x \rightarrow \infty} \frac{L(kx)}{L(x)} = 1.$$

These functions are discussed in Seneta (1976), examples being $|\log x|^q$ and $[|\log|\log x| |]^q$.

Consider a process generated by

$$x_t = (1 - B)^{-d} [L((1 - B)^{-1})]^{-q} \varepsilon_t, \tag{3.2}$$

where $L(x)$ is *sv* and ε_t is a zero mean white noise. For ease of exposition only a particular example will be considered.

$$x_t = (1 - B)^{-d} [\log(1 - B)^{-1}]^{-q} \varepsilon_t, \tag{3.3}$$

with $d > 0, q > 0$. This will be called a generalized integrated model of order d, q , denoted $x_t \sim \text{GI}(d, q)$. If $L(x) = 1$, all x , or if $q = 0$, one has an integrated process of order $d, x_t \sim \text{I}(d)$. Provided $d > 0$ or if $d = 0$ but $q > 0$, the process will have the long-memory property. In all cases, the function of B in (3.2) or (3.3) should be considered as being truncated after t terms, and t is assumed large so that the truncated power series provides a good approximation to the function. From the properties of *sv* functions it is found that a series that is $\text{I}(d + \delta)$ dominates one that is $\text{GI}(d, q)$ which dominates one that is $\text{I}(d - \delta)$ for any $\delta > 0$. The $\text{GI}(d, q)$ series has spectrum

$$|1 - z|^{-2d} |\log(1 - z)^{-1}|^{-2q} \sigma_\varepsilon^2, \quad z = e^{i\omega}. \tag{3.4}$$

For small ω the spectrum is proportional to $\omega^{-2d} (\log \omega)^{2q}$. GI processes are discussed in Granger (1987, 1988), and their properties explored in Lin (1991) who directs particular attention to the case $d = 0, q > 0$, which is denoted $\text{GI}(0^+)$. From the dominance property, it is clear that the variance $v(t)$ of $\text{GI}(d, q)$ tends to a finite constant if $d < \frac{1}{2}$. Lin (1991) proves that if \bar{x}_n is the sample mean from an $\text{GI}(0^+)$ series, then $(n/\log n)^{1/2}(\bar{x}_n - \mu)$ tends asymptotically to a normal distribution with mean zero, for some μ , and a finite constant variance.

No specific formula is available for the autocorrelation sequence of a GI series, but a $\text{GI}(0^+)$ has autocorrelations close to being proportional to the inverse of the lag, for large lags. In general the properties will not be greatly different from those of $\text{I}(d)$ processes, as discussed in Granger and Joyeux (1980) and Hosking (1981), and the generalization discussed by Gray, Zhang, and Woodward (1989).

If in (3.3) the inputs have a constant, nonzero mean m , the series generated will have a trend component $mh(t)$, where

$$h(t) = t^{2d} (\log t)^{2q}$$

as well as the $GI(d, q)$ component previously discussed, as shown in Granger (1988). Thus, in these cases, a long-memory process with a specific low-frequency shape will correspond to a particular trend in mean if m is a nonzero constant. Clearly, if m changes with time, this simple relationship is lost. For a sequence h_j the trend and spectrum were defined in the previous section. However, it is *not* possible to go from *any* trend to a long-memory process. If the trend is $mh(t)$, then h_t is well approximated by $dh(t)/dt$. Consider the trend $h(t) = \exp(at)$, so that $h_t = a \exp(at)$, then some algebra shows that $f_N(\omega)$ defined by (1.6) does not converge for any ω as N increases, so that the spectrum does not obey the second condition for a long-memory process. It is worth noting that $e^{-2aN} f_N(\omega)$ does converge for $\omega = 0$.

It is noted in Granger (1988) that if the further requirement is made that h_j is decreasing with j , so that less weight is given to the distant past than to recent shocks, this limits trends to those dominated by the linear trend.

4. Aggregation results

It is shown in Granger (1980) that if y_{jt} is AR(1) with autoregressive parameter α_j and if x_t is the sum of independent y_{jt} 's over a large number of components with the α_j 's drawn from a particular Beta distribution on $(0, 1)$, then asymptotically $x_t \sim I(d)$, with $d < \frac{1}{2}$. Thus, aggregation of stationary series with autoregression parameters up to, but less than one, can lead to a fractionally integrated series. Lin (1991) shows that similarly if the α_j are drawn from the probability density function $cx(1 - x^2)$, x in $[0, 1]$, then asymptotically the sum is $GI(0^+)$.

One can also aggregate $I(d)$ series. Lin (1991) shows that if $y_{jt} \sim I(d_j)$ with the d_j drawn from the p.d.f. $c(1 - x)^{-1/2}$, x in $[0, d]$, then the sum of a large number of independent components will be $GI(d)$. $S(\omega)$, the spectrum of the sum, x_t , will be the sum of the components spectra, so that if d_j is drawn from the p.d.f. $g(x)$, at low frequencies one has

$$S(\omega) = N \int \omega^{-2x} g(x) dx. \quad (4.1)$$

Writing the integral as $\int \exp(x \log \omega^{-2}) g(x) dx$, then if $M(\tau)$ is the moment generating function of the p.d.f. $g(x)$, so that

$$M(\tau) = \int \exp(\tau x) \cdot g(x) dx, \quad (4.2)$$

one gets

$$S(\omega) = NM(\log \omega^{-2}), \quad (4.3)$$

for ω small. It seems that a variety of spectral shapes can be generated in this fashion just by selecting $g(x)$ and looking up the corresponding $M(\tau)$ in tables, as

in Lukacs (1970). For example, if $g(x)$ is the normal $N(\mu, \sigma^2)$ distribution, then one gets

$$S(\omega) = N\omega^{(-2\mu + 2\sigma^2 \log \omega)} \quad (4.4)$$

If $\sigma^2 > 0$, a process having this spectrum will dominate any $I(d)$ process. As a further example, Lukacs (1970, p. 251) refers to Levy (1937) who proves that if $P(x)$ is a real polynomial with all coefficients positive, then $M(\tau) = \exp(P(e^\tau) - P(1))$ is a moment generating function, which gives a spectrum

$$S(\omega) = N \exp \left[\sum_{j=1}^P \beta_j (\omega^{-2j} - 1) \right], \quad \text{all } \beta_j > 0. \quad (4.5)$$

For example, if $P = 1$,

$$S(\omega) = Nc \exp(\beta\omega^{-2}), \quad (4.6)$$

which corresponds to a long-memory process. This moment generating function corresponds to a Poisson variable with parameter β , so that if $y_{jt} \sim I(d_j)$ with d_j drawn from a $P_0(\beta)$ distribution, then the large-sample sum will have spectrum (4.6). Such a series will also dominate an $I(d)$ series, for any d , and may be thought of as $I(\infty)$.

It is interesting to note that not all moment generating functions lead to the spectrum of a long-memory process; examples include the Laplace distribution, for which $M(\tau) = (1 - \tau^2)^{-1}$, and the Gamma distribution, for which $M(\tau) = (1 - \tau/\theta)^{-\lambda}$, $\theta, \lambda > 0$. However, for these examples τ is bounded above, and so (4.3) cannot be used for small ω .

5. Time-changing coefficient models

There is plenty of evidence throughout economics of structural changes, time-varying parameters, or regime switches, and so it is interesting to ask if such changes occur with long-memory models and what are the effects. To illustrate the possibility, we reconsidered the example of Section 2.

Sixty-four years with 17,054 daily observations is a rather long time period, and it is very possible there is structural change over time. To investigate this, we divided the whole sample period to ten equal subperiods (the last four observations were deleted here) and still use the GPH estimation procedure. The estimated d 's and their standard errors and t -statistics are shown in Table 1. Fig. 6 also plots \hat{d} for ten different sample periods and their one standard deviation lines.

There is strong evidence from the estimated result that d is changing over time. \hat{d} reaches its smallest value of 0.156 at period 5, which is approximately from 1954 to 1960. After that \hat{d} increases towards its maximum value of 0.714 at

Table 1
Estimated d for ten subperiods

Period	Time	d	Std. err.	t -stat.	$d - \sigma$	$d + \sigma$
1	1928–1934	0.358	0.133	2.70	0.225	0.419
2	1934–1940	0.405	0.131	3.08	0.274	0.536
3	1941–1947	0.438	0.097	4.52	0.341	0.535
4	1947–1953	0.336	0.144	2.33	0.192	0.480
5	1954–1960	0.156	0.106	1.47	0.050	0.262
6	1960–1966	0.445	0.113	3.93	0.332	0.558
7	1967–1973	0.518	0.088	5.89	0.430	0.606
8	1973–1979	0.714	0.105	6.79	0.608	0.818
9	1980–1986	0.436	0.110	3.98	0.326	0.546
10	1986–1991	0.352	0.070	5.00	0.282	0.422

period 8, which is approximately from 1973 to 1979. Two out of ten estimated d 's are in the nonstationary range (≥ 0.5). Since the estimated standard deviation for period 8 is 0.105 and \bar{d} is 0.714, we can say with high confidence that $|r_t|$ during the time period 1973–1979 is nonstationary.

Although we do not plan to produce a model that exactly reproduces these results, in this section a variety of time-varying parameter models are discussed. The obvious model would just have d in $I(d)$ replaced by $d_t = d(z_t)$, where z_t is some exogenous, observed variable, such as the start of the business cycle. However, if such a driving variable is not available, the following model may be of interest. A stochastic fractional differencing model can be defined as

$$x_t = \exp[-d_t \log(1 - B)] \varepsilon_t, \quad (5.1)$$

where ε_t is zero mean white noise and d_t is a stochastic process, for example,

$$(d_t - \bar{d}) = \alpha(d_{t-1} - \bar{d}) + u_t, \quad (5.2)$$

where u_t is zero mean i.i.d., so that d_t is AR(1) with mean \bar{d} . If the variance of u_t is small, x_t will appear to be $GI(\bar{d})$. However, note that d_t is now being generated with a temporal structure compared to how $GI(d)$ processes were being generated by aggregation in the previous section.

A rather different situation is when d_t switches between two regimes. Suppose that the two processes x_{1t} and x_{2t} are generated from the same input, so that

$$x_{1t} = a_1(B)\varepsilon_t, \quad x_{2t} = a_2(B)\varepsilon_t.$$

The filters $a_j(B)$, $j = 1, 2$, may be such that x_{1t} is $I(d_1)$ and x_{2t} is $I(d_2)$, say, $d_1 \neq d_2$. Let λ_t be a Markov switching process taking either the value 0 or 1 with

$$P_{ij} = \text{prob}(\lambda_t = i, \lambda_{t+1} = j), \quad i, j = 0, 1,$$

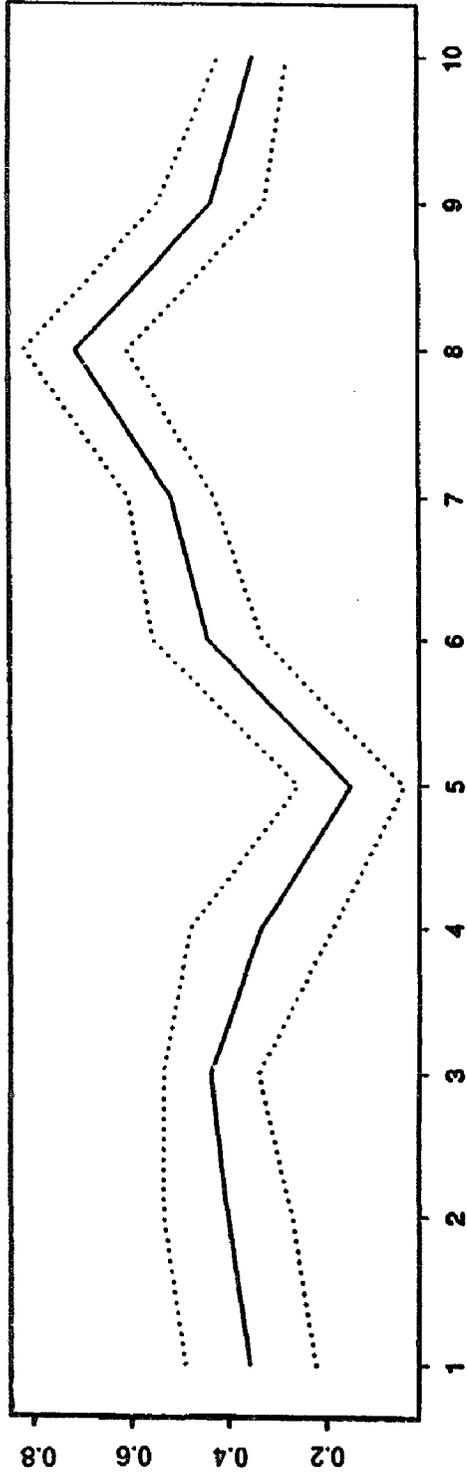


Fig. 6. Estimated d for ten equal sample periods and their one standard deviation lines.

and define

$$y_t = \lambda_t x_{1t} + (1 - \lambda_t) x_{2t}.$$

Let

$$\underline{P} = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix},$$

and denote

$$\underline{P}^K = \begin{bmatrix} P_{00}^{(K)} & P_{01}^{(K)} \\ P_{10}^{(K)} & P_{11}^{(K)} \end{bmatrix},$$

so that $P_{ij}^{(K)} = \text{prob}(\lambda_t = i, \lambda_{t+K} = j)$, $i, j = 0, 1$. It is easy to see that

$$\begin{aligned} E[y_{t+K} y_t] &= P_{11}^{(K)} E[x_{1,t+K} x_{1,t}] + P_{01}^{(K)} E[x_{1,t+K} x_{2,t}] \\ &\quad + P_{00}^{(K)} E[x_{2,t+K} x_{2,t}] + P_{10}^{(K)} E[x_{2,t+K} x_{1,t}]. \end{aligned} \quad (5.3)$$

Denote by $\bar{P}_0, \bar{P}_1 (1 - \bar{P}_0)$ the marginal probabilities that $\lambda_t = 0, 1$, respectively. Then, for K large, $P_{11}^{(K)} = \bar{P}_1, P_{00}^{(K)} = \bar{P}_0$, and $P_{10}^{(K)} = P_{01}^{(K)} = 0$. Further, denoting $P_{00} = p$ and $P_{11} = q$ it is easily shown that

$$\bar{P}_0 = (1 - q)/\theta, \quad \bar{P}_1 = (1 - p)/\theta,$$

where $\theta = (2 - p - q)$. Thus, if $p = q$, it follows that $\bar{P}_0 = \bar{P}_1 = \frac{1}{2}$. Convergence of $P_{00}^{(K)}$ is fairly rapid; for example, if $p = q = 0.8$, then $P_{00}^{(K)} = P_{11}^{(K)}$ is 0.565 for $K = 4$ and 0.523 for $K = 6$. From this analysis it is seen that the autocovariance of y_t , which experiences regime switching, is just a weighted average of the autocovariance of the series in the two regimes. Clearly, a variety of correlogram shapes can arise, producing different long-memory processes.

6. Nonlinear models

Long memory, defined by properties of the correlogram and spectrum, is concerned essentially with linear properties of a process. There has been very little work considering these properties for processes generated by nonlinear models. An indication that new forms of long-memory process may be produced nonlinearly comes from results on the instantaneous transformation of random walks, as discussed by Granger and Hallman (1991) and Ermini and Granger (1993). If x_t is generated by $x_t = x_{t-1} + \varepsilon_t$, where ε_t is i.i.d. $N(0, 1)$, let $y_t = G(x_t)$ for some function G . Provided $G(x)$ is monotonic nondecreasing, then y_t will have the extended-memory property discussed in the first section. However, if $G(x)$ is $G(x) = (1 + \exp(-x))^{-1}$, the autocorrelations of y_t do decline, but slowly, suggesting that a long-memory model may provide a reasonable linear approximation.

The general topic of long- or extended-memory series generated by nonlinear models has been little explored and could be an interesting area of research in the future. As an illustration of the possibilities, consider a fractional differenced process $x_t \sim I(d)$, generated by

$$x_t = \sum_{j=0}^{\infty} c_j(d) \varepsilon_{t-j},$$

where

$$c_j(d) = \Gamma(j+d)/(\Gamma(d)\Gamma(j+1)) = A j^{d-1}.$$

Now consider a series x_t generated by a Volterra series expansion of the form

$$x_t = \sum c_j(d_1) \varepsilon_{t-j} + \sum c_j(d_2) \varepsilon_{t-j}^2 + \sum c_j(d_3) \varepsilon_{t-j}^3 + A \sum (j+k)^{d-1} \varepsilon_{t-j} \varepsilon_{t-k} \text{ etc.}$$

It will clearly be long memory, but of a complicated type. In theory, there is no reason to stay with just linear and quadratic terms.

7. Conclusions

The property of 'long memory' as defined in the time series literature relates to a series having a finite spectrum at (essentially) all frequencies, but infinite at the zero frequency. The fractionally integrated processes have that property, but in this paper it is pointed out that many other generating mechanism can produce such processes, at least in theory. These new classes will often be of more theoretical interest than of practical relevance, but we have shown that on some occasions at least time-varying parameter models are worth consideration.

Our data in economics and elsewhere have interesting and complicated properties, and one task of statisticians and econometricians is to produce models that have these properties. On some occasions it is difficult to find any simple model that has all of the properties and that makes an interesting challenge. Some economic series appear to contain a property that looks like 'long memory', at least through the eyes of statistical tools that we use to view the series. For example, if a monthly macro series has fitted to it an $I(d)$ model, then d may well be significantly different from zero or one, and lie in that region, say, but that is because its true generating mechanism is something quite different and is not being considered within the class of models being estimated. We believe that it is correct to search over a specific set of simple parsimonious models, but that one should realize that the model so achieved is, at best, an approximation to the truth. The models fitted may be on slightly better footing if they are suggested by some convincing theory, using realistic assumptions, but this does not seem to be the case for fractionally integration models, so far.

In the future, we believe the emphasis will turn to groups of properties, such as long memory and the marginal distribution of a stationary $I(d)$ series (which should not be assumed to be normal) or the fractional integrated relatedness between the mean and variance of a process. It may well be found that the choice of appropriate models becomes more difficult as more properties are added, but this also makes the task more interesting.

References

- Box, G.E.P. and G.M. Jenkins, 1970, *Time series, forecasting and control* (Holden-Day, San Francisco, CA).
- Ding, Z., 1994, *Time series analysis of speculative returns*, Ph.D. dissertation (Economics Department, University of California, San Diego, CA).
- Ding, Z. and C.W.J. Granger, 1996, Modeling volatility persistence of speculative returns: A new approach, *Journal of Econometrics*, this issue.
- Ding, Z., C.W.J. Granger, and R.F. Engle, 1993, A long memory property of stock market returns and a new model, *Journal of Empirical Finance* 1, 83–106.
- Ermini, L. and C.W.J. Granger, 1993, Some generalizations of the algebra of $I(1)$ processes, *Journal of Econometrics* 58, 369–384.
- Geweke, J. and S. Porter-Hudak, 1983, The estimation and application of long memory time series models, *Journal of Time Series Analysis* 4, 221–238.
- Granger, C.W.J., 1980, Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics* 14, 227–238.
- Granger, C.W.J., 1987, *Generalized integrated processes*, Discussion paper (Economics Department, University of California, San Diego, CA).
- Granger, C.W.J., 1988, Models that generate trends, *Journal of Time Series Analysis* 9, 329–343.
- Granger, C.W.J., 1995, Non-linear relationships between non-stationary processes, *Econometrica* 63, 265–279.
- Granger, C.W.J. and Z. Ding, 1996, Some properties of absolute return: An alternative measure of risk, *Annales d'Economie et de Statistique*, forthcoming.
- Granger, C.W.J. and J. Hallman, 1991, Nonlinear transformation of integrated time series, *Journal of Time Series Analysis* 13, 207–224.
- Granger, C.W.J. and R. Joyeux, 1980, An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis* 1, 15–29.
- Granger, C.W.J. and P. Newbold, 1986, *Forecasting economic times series*, 2nd ed. (Academic Press, San Francisco, CA).
- Granger, C.W.J. and N. Swanson, 1992, *An introduction to stochastic unit root processes*, Discussion paper (Economics Department, University of California, San Diego, CA).
- Gray, H.L., N.F. Zhang, and W.A. Woodward, 1989, On generalized fractional processes, *Journal of Time Series Analysis* 10, 233–258.
- Hosking, J., 1981, Fractional differencing, *Biometrika* 68, 165–176.
- Levy, P., 1937, Sur les exponentielles des polynoms, *Annales Scientifiques Ecole Normale Superieure (3e serie)* 73, 231–292.
- Lin, Jin-Lung, 1991, Generalized integrated processes and the aggregation of dynamic time series, *Academia Economic Papers* 19, 207–226.
- Lukacs, E., 1970, *Characteristic functions*, 2nd ed. (Griffin, London).
- Seneta, E., 1976, *Regularly varying functions*, Lecture notes in mathematics no. 508 (Springer-Verlag, New York, NY).