ELSEVIER

Economics Letters 50 (1996) 305–313

economics letters

Fractional integration and interval prediction

Francis X. Diebold[a,b,*], Peter Lindner[c]

[a]University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297, USA
[b]NBER, Cambridge, MA 02138, USA
[c]Lehman Brothers, 3 World Financial Center, New York, NY 10285-1100, USA

Received 12 August 1994; accepted 16 June 1995

Abstract

The motivation for fractional integration in terms of low-frequency spectral behavior and long-lag autocorrelation behavior is well known. Using results on the rate of growth of variances of sums of integrated random variables, we provide additional and complementary time-domain motivation for fractional integration in terms of the long-horizon behavior of (1) the variance–time function, and (2) confidence intervals for predictions. The results are illustrated with an empirical application to real interest rate forecasting.

Keywords: Forecasting; Long memory; Prediction interval; Real interest rates

JEL classification: C10

1. Introduction and statement of results

The frequency-domain motivation for the generality afforded by fractional integration, in terms of spectral behavior near frequency zero, is well known. Similarly, the time-domain motivation, in terms of the slow decay of autocovariances, is also well known (see, for example, Granger and Joyeux (1980)). The simple and straightforward objective of this paper is closely related to these. We provide complementary time-domain motivation for fractional integration in terms of:

(1) the long-horizon behavior of the variance–time function, used to study long memory at least since Mandelbrot (1972) and popularized recently by Cochrane (1988), Lo (1991) and Faust (1992); and

(2) the long-horizon behavior of prediction intervals, the importance of which is emphasized by Stock and Watson (1988), and which can be very important in practical applications.

* Correspondence to: F.X. Diebold, Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297, USA. Fax: (215) 573-2057.

0165-1765/96/$12.00 © 1996 Elsevier Science S.A. All rights reserved
SSDI 0165-1765(95)00772-5

Our original interest in this research program was in characterizing and illustrating the behavior of long-range prediction intervals in long-memory models. The variance–time function (that is, the behavior of $\mathrm{var}(x_t - x_{t-k})$ as a function of k), however, turns out to be closely related to the behavior of prediction intervals and is of substantial interest in its own right.[1] Understanding the variance–time function and the behavior of prediction intervals, in turn, requires an understanding of the behavior of variances of the partial sums of random variables; that is, the behavior of $\mathrm{var}(\Sigma x_t)$ as a function of the number of terms summed.

The results for the rates of growth of variances of M-fold partial sums are straightforward and for the most part well known. The rate of growth of the variance of M-fold partial sums of white noise ($X_t = \varepsilon_t$) or $I(0)$ ($X_t = \psi(L)\varepsilon_t$) processes is the same, $\mathrm{var}(S_M) = O(M)$. Similarly, the rate of growth of the variance of the sums of random walk ($\Delta X_t = \varepsilon_t$) or $I(1)$ ($\Delta X_t = \psi(L)\varepsilon_t$) processes is the same, $\mathrm{var}(S_M) = O(M^3)$. Finally, the rate of growth of the variance of the sums of fractional noise ($X_t = (1-L)^{-d}\varepsilon_t$, $-1/2 < d < 1/2$) or $I(d)$ ($X_t = \psi(L)(1-L)^{-d}\varepsilon_t$, $-1/2 < d < 1/2$) processes is the same, $\mathrm{var}(S_M) = O(M^{2d+1})$. All the results in this paper stem from these simple facts on the rates of variance growth, which make clear the rigid nature of $I(0)$ and $I(1)$ processes. To obtain a continuously varying range of rates of variance growth, we need to allow for a continuously varying range of orders of integration.

The results for the variance–time function are as follows. Let $\Delta_k X_{t+k} \equiv X_{t+k} - X_t$. Then, in the white noise and $I(0)$ cases we have $\mathrm{var}(\Delta_k X_{t+k}) = O(1)$, and in the random walk and $I(0)$ cases we have $\mathrm{var}(\Delta_k X_{t+k}) = O(k)$. The covariance stationary fractional case parallels the $I(0)$ case; more precisely, in the cases of covariance stationary pure fractional noise ($X_t = (1-L)^{-d}\varepsilon_t$, $-1/2 < d < 1/2$) and covariance stationary $I(d)$ ($X_t = \psi(L)(1-L)^{-d}\varepsilon_t$, $-1/2 < d < 1/2$) we have $\mathrm{var}(\Delta_k X_{t+k}) = O(1)$. The fractional cases for which $1/2 < d < 3/2$ are more interesting; then in both the pure fractional noise and general $I(d)$ cases we have $\Delta_k X_{t+k} = O(k^{2d-1})$. This generality in long-lag behavior of the variance–time function is a key time-domain analog of the well-known frequency-domain result on low-frequency spectral density behavior.

Now let us consider prediction intervals. As is well known, the width of the k-step-ahead prediction interval for a Gaussian covariance stationary (and hence strictly stationary) process x_t approaches $c\sigma_x$ (from below) as $k \to \infty$, where σ_x is the unconditional standard deviation of x, and c is again a constant determined by the desired confidence level (e.g. a 95% prediction interval requires $c = 1.96$). That is, it is bounded by $c\sigma_x$, which it approaches in the limit. We can think of this as the prediction-interval width eventually increasing like k^0, for large k.

As is also well known, the width of the k-step-ahead prediction interval for a Gaussian $I(1)$ series x_t grows like $c\sigma_\varepsilon k^{1/2}$ as $k \to \infty$, where σ_ε is the standard deviation of the innovation to x, and c is a constant determined by the desired confidence level. That is, the width of the prediction interval grows without bound, for large k, *and at the particular rate* $k^{1/2}$.

The upshot is obvious. Why should forecast error uncertainty necessarily propagate with the forecast horizon like k^0 or $k^{1/2}$? What about the general case k^d, d being a positive fraction? That is where fractional integration comes in, enabling substantially more generality and flexibility in the construction of k-step-ahead prediction intervals than do standard $I(0)$ or $I(1)$ parameterizations.

[1] These issues were mentioned, but not pursued, in Diebold and Nerlove (1990).

The results are as follows. In the white noise and $I(0)$ cases we have $\mathrm{var}(X_{t+k} - E_t X_{t+k}) = O(1)$, and in the random walk and $I(1)$ cases we have $\mathrm{var}(X_{t+k} - E_t X_{t+k}) = O(k)$. As with the variance–time function, the covariance stationary fractional case parallels the $I(0)$ case; more precisely, in the cases of covariance stationary pure fractional noise and covariance stationary $I(d)$ processes, we have $\mathrm{var}(X_{t+k} - E_t X_{t+k}) = O(1)$. The non-stationary fractional cases for which $1/2 < d < 3/2$ are again the most interesting; then in both the pure fractional noise and general $I(d)$ cases we have $\mathrm{var}(X_{t+k} - E_t X_{t+k}) = O(k^{2d+1})$.

2. Proofs

Now we provide proofs of the above-made claims. Let $B(L)$ be an infinite-ordered lag operator polynomial, let $\psi(L)$ be a ratio of finite-ordered lag operator polynomials, and let $\gamma_z(\tau)$ denote the autocovariance function of a time series z at displacement τ.

2.1. Partial sums

Lemma. For the fractional noise $X_t = (1 - L)^{-d}\varepsilon_t$, $-1/2 < d < 1/2$, $\sum_{\tau=0}^{M-1} \gamma_u(\tau) = O(M^{2d})$.

Proof. From Sowell (1990), we have

$$\sum_{\tau=0}^{M-1} \gamma_u(\tau) = \sigma_\varepsilon^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \sum_{\tau=0}^{M-1} \frac{\Gamma(d+\tau)}{\Gamma(1-d+\tau)} .$$

Changing the summation index yields

$$\sigma_\varepsilon^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \sum_{\tau=1}^{M} \frac{\Gamma(d-1+\tau)}{\Gamma(-d+\tau)} ,$$

which may be rewritten as

$$\sigma_\varepsilon^2 \frac{\Gamma(1-2d)}{2d\Gamma(d)\Gamma(1-d)} \left[\frac{\Gamma(M+d)}{\Gamma(M-d)} - \frac{\Gamma(d)}{\Gamma(-d)} \right] = O(M^{2d}) .$$

(a) *White noise* ($X_t = \varepsilon_t$). In the white noise case, $S_M = \sum_{i=0}^{M-1} X_{t+i}$ is just the sum of M innovations, $\sum_{i=0}^{M-1} \varepsilon_{t+i}$. Thus,

$$\mathrm{var}(S_M) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \mathrm{cov}(\varepsilon_{t+i}, \varepsilon_{t+j}) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \gamma_\varepsilon(i-j) .$$

However, the white noise property implies that $\gamma_\varepsilon(i-j) = 0$, $\forall i \neq j$. Thus, $\mathrm{var}(S_M) = \sum_{i=0}^{M-1} \sigma_\varepsilon^2 = M\sigma_\varepsilon^2 = O(M)$.

(b) $I(0)$ $(X_t = \psi(L)\varepsilon_t = u_t)$. In parallel with the white noise case, $S_M = \Sigma_{i=0}^{M-1} X_{t+i} = \Sigma_{i=0}^{M-1} u_{t+i}$. Thus,

$$\mathrm{var}(S_M) = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} \mathrm{cov}(u_{t+i}, u_{t+j}) + \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} \gamma_u(i-j) = \sum_{i=0}^{M-1} O(1)\,,$$

because the autovariances of an $I(0)$ process have a finite sum. Thus, $\mathrm{var}(S_M) = O(M)$.

(c) *Random walk* $(\Delta X_t = \varepsilon_t)$. Immediately, $S_M = \Sigma_{i=0}^{M-1} X_{t+i} = \Sigma_{i=0}^{M-1}\Sigma_{j=0}^{i}\varepsilon_{t+j}$. Thus, $\mathrm{var}(S_M) = \Sigma_{i=0}^{M-1}\ \Sigma_{j=0}^{M-1}\ (M-i)(M-j)\ \mathrm{cov}(\varepsilon_{t+i},\ \varepsilon_{t+j})$. Rearranging,

$$\sum_{i=0}^{M-1} (M-i) \sum_{j=0}^{M-1} (M-j)\gamma_\varepsilon(i-j) = \sigma_\varepsilon^2 \sum_{i=0}^{M-1} (M-i)^2 = \sigma_\varepsilon^2 \tfrac{1}{6}[(M^2+M)(2M+1)] = O(M^3)\,.$$

(d) $I(1)$ $(\Delta X_t = \psi(L)\varepsilon_t = u_t)$. In parallel with the random walk case, $S_M = \Sigma_{i=0}^{M-1} X_{t+i} = \Sigma_{i=0}^{M-1}\ \Sigma_{j=0}^{i}\ u_{t+j} = \Sigma_{i=0}^{M-1}\ (M-i)u_{t+i}$. Evaluating the variance gives

$$\begin{aligned}\mathrm{var}(S_M) &= \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} (M-i)(M-j)\,\mathrm{cov}(u_{t+i}, u_{t+j}) = \sum_{i=0}^{M-1} (M-i)\left[\sum_{j=0}^{M-1} (M-j)\gamma_u(i-j)\right]\\ &= \sum_{i=0}^{M-1} (M-i)[O(M)] = O(M^3)\,.\end{aligned}$$

(e) *Fractional noise* $(X_t = (1-L)^{-d}\varepsilon_t = u_t,\ -1/2 < d < 1/2$. $S_M = \Sigma_{i=0}^{M-1}\ X_{t+i} = \Sigma_{i=0}^{M-1}\ u_{t+i}$. Thus, $\mathrm{var}(S_M) = \Sigma_{i=0}^{M-1}\ \Sigma_{j=0}^{M-1}\ \mathrm{cov}(u_{t+i},u_{t+j}) = \Sigma_{i=0}^{M-1}\ \Sigma_{j=0}^{M-1}\ \gamma_u(i-j)$. However, by the lemma, this is just $\Sigma_{i=0}^{M-1}\ O(M^{2d}) = O(M^{2d+1})$.

(f) $I(d)$ $(X_t = \psi(L)(1-L)^{-d}\varepsilon_t = u_t,\ \ -1/2 < d < 1/2)$. As usual, $S_M = \Sigma_{i=0}^{M-1}\ X_{t+i} = \Sigma_{i=0}^{M-1}\ u_{t+i}$. Thus,

$$\mathrm{var}(S_M) = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} \mathrm{cov}(u_{t+i}, u_{t+j}) = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} \gamma_u(i-j) = \sum_{i=0}^{M-1} O(M^{2d})$$

(because the long-memory component dominates) $= O(M^{2d+1})$.

2.2. The variance–time function

(a) *White noise* $(X_t = \varepsilon_t)$. $X_{t+k} = \varepsilon_{t+k}$, so that the kth difference is $\Delta_k X_{t+k} \equiv X_{t+k} - X_t = \varepsilon_{t+k} - \varepsilon_t$. Thus, $\mathrm{var}(\Delta_k X_{t+k}) = \mathrm{var}(\varepsilon_{t+k} - \varepsilon_t) = 2\sigma_\varepsilon^2 = O(1)$.

(b) $I(0)$ $(X_t = \psi(L)\varepsilon_t = u_t)$. $X_{t+k} = \psi(L)\varepsilon_{t+k} = u_{t+k}$, has kth difference $\Delta_k X_{t+k} = u_{t+k} - u_t$. Thus, $\mathrm{var}(\Delta_k X_{t+k}) = \mathrm{var}(u_{t+k}) + \mathrm{var}(u_t) - 2\mathrm{cov}(u_{t+k},\ u_t) = O(1)$. In particular, we have $\lim_{k\to\infty} \mathrm{var}(\Delta_k X_{t+k}) = 2\sigma_x^2$ (by the square summability of $\psi(L)$).

(c) *Random walk* $(\Delta X_t = \varepsilon_t)$. The process is $X_{t+k} = X_{t+k-1} + \varepsilon_{t+k} = X_t + \Sigma_{i=1}^{k}\ \varepsilon_{t+i}$. Thus, $\mathrm{var}(\Delta_k X_{t+k}) = \mathrm{var}(\Sigma_{i=1}^{k}\ \varepsilon_{t+i}) = k\sigma_\varepsilon^2 = O(k)$.

(d) $I(1)$ ($\Delta X_t = \psi(L)\varepsilon_t = u_t$). The process is $\Delta_k X_{t+k} = \Sigma_{j=1}^{k} \Delta X_{t+j} = \Sigma_{j=1}^{k} u_{t+j}$. Thus,

$$\mathrm{var}(\Delta_k X_{t+k}) = \mathrm{var}\left(\sum_{j=1}^{k} u_{t+j}\right) = \sum_{j=1}^{k} \mathrm{var}(u_{t+j}) + \sum_{i=1}^{k}\sum_{j=1}^{k} \mathrm{cov}(u_{t+i}, u_{t+j}) = O(k) ,$$

by square summability of $\psi(L)$.

(e) *Fractional noise* ($X_t = (1-L)^{-d}\varepsilon_t = B(L)\varepsilon_t = u_t$, $-1/2 < d < 1/2$). The proof parallels exactly that for the $I(0)$ case, with $B(L)$ replacing $\psi(L)$, resulting in a variance–time function that is $O(1)$.

(f) $I(d)$ ($X_t = \psi(L)(1-L)^{-d}\varepsilon_t = B(L)\varepsilon_t = u_t$, $-1/2 < d < 1/2$). The proof parallels exactly that for the $I(0)$ case, with $B(L)$ replacing $\psi(L)$, resulting in a variance–time function that is $O(1)$.

(g) *Fractional noise* ($\Delta X_t = (1-L)^{-d'}\varepsilon_t = u_t$, $d' = d-1$, $1/2 < d < 3/2$). $\Delta_k X_{t+k} = \Sigma_{j=1}^{k} \Delta X_{t+j} = \Sigma_{j=1}^{k} u_{t+j}$, so that

$$\begin{aligned}\mathrm{var}(\Delta_k X_{t+k}) &= \sum_{i=1}^{k}\sum_{j=1}^{k} \mathrm{cov}(u_{t+i}, u_{t+j}) = \sum_{i=1}^{k}\left[\sum_{j=1}^{k} \gamma_u(i-j)\right] = \sum_{i=1}^{k} O(k^{2d'}) \text{ (by the lemma)} \\ &= O(k^{2d'+1}) = O(k^{2d-1}) .\end{aligned}$$

(h) $I(d)$ ($\Delta X_t = (1-L)^{-d'}\psi(L)\varepsilon_t = B(L)\varepsilon_t = u_t$, $d' = d-1$, $1/2 < d < 3/2$). $\Delta_k X_{t+k} = \Sigma_{j=1}^{k} \Delta X_{t+j} = \Sigma_{j=1}^{k} u_{t+j}$, so that

$$\begin{aligned}\mathrm{var}(\Delta_k X_{t+k}) &= \sum_{i=1}^{k}\sum_{j=1}^{k} \mathrm{cov}(u_{t+i}, u_{t+j}) = \sum_{i=1}^{k}\left[\sum_{j=1}^{k} \gamma_u(i-j)\right] = \sum_{i=1}^{k} [O(k^{2d'})] \text{ (by the lemma)} \\ &= O(k^{2d'+1}) = O(k^{2d-1}) .\end{aligned}$$

2.3. Prediction error variance

(a) *White noise* ($X_t = \varepsilon_t$). The future value of the process is $X_{t+k} = \varepsilon_{t+k}$, so that $E_t X_{t+k} = 0$ and $(X_{t+k} - E_t X_{t+k}) = \varepsilon_{t+k}$. Thus, $\mathrm{var}(X_{t+k} - E_t X_{t+k}) = \mathrm{var}(\varepsilon_{t+k}) = \sigma_\varepsilon^2 = O(1)$. In particular, $\lim_{k\to\infty} \mathrm{var}(X_{t+k} - E_t X_{t+k}) = \sigma_\varepsilon^2$.

(b) $I(0)$ ($X_t = \psi(L)\varepsilon_t = u_t$). The future value is $X_{t+k} = \psi(L)\varepsilon_{t+k}$, so that $E_t X_{t+k} = \psi_k \varepsilon_t + \psi_{k+1}\varepsilon_{t-1} + \cdots$ and $(X_{t+k} - E_t X_{t+k}) = \Sigma_{i=0}^{k-1} \psi_i \varepsilon_{t+k-i}$. Thus,

$$\mathrm{var}(X_{t+k} - E_t X_{t+k}) = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} \psi_i \psi_j \,\mathrm{cov}(\varepsilon_{t+k-i}, \varepsilon_{t+k-j}) = \sigma_\varepsilon^2 \sum_{i=0}^{k-1} \psi_i^2 = O(1) .$$

In particular, $\lim_{k\to\infty} \mathrm{var}(X_{t+k} - E_t X_{t+k}) = \sigma_\varepsilon^2 \Sigma_{i=0}^{\infty} \psi_i^2 = \sigma_x^2$.

(c) *Random walk* ($\Delta X_t = \varepsilon_t$). The future value of the process is $X_{t+k} = X_t + \Sigma_{i=0}^{k-1} \varepsilon_{t+k-i}$, so that $E_t X_{t+k} = X_t$ and $(X_{t+k} - E_t X_{t+k}) = \Sigma_{i=0}^{k-1} \varepsilon_{t+k-i}$. Thus, $\mathrm{var}(X_{t+k} - E_t X_{t+k}) = O(k)$, by our earlier theorem on the partial sums of white noise series. In particular, $\mathrm{var}(X_{t+k} - E_t X_{t+k}) = \Sigma_{i=0}^{k-1} \Sigma_{j=0}^{k-1} \mathrm{cov}(\varepsilon_{t+k-i}, \varepsilon_{t+k-j}) = k\sigma_\varepsilon^2$, so that $\lim_{k\to\infty} \mathrm{var}(X_{t+k} - E_t X_{t+k}) = \infty$.

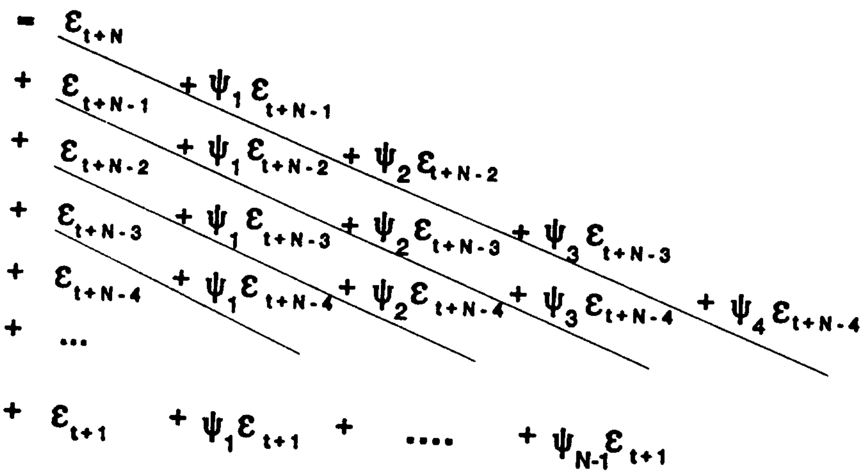

Fig. 1. Prediction error for the $I(1)$ process.

(d) $I(1)$ $(\Delta X_t = \psi(L)\varepsilon_t = u_t)$. The future value of the process is $X_{t+k} = X_t + \Sigma_{i=0}^{k-1} \psi(L)\varepsilon_{t+k-i}$, so that $E_t X_{t+k} = X_t + (\Sigma_{i=1}^{\infty} \psi_i)\varepsilon_t + (\Sigma_{i=2}^{\infty} \psi_i)\varepsilon_{t-1} + \cdots$ and $(X_{t+k} - E_t X_{t+k}) = \Sigma_{j=0}^{k-1}(\Sigma_{i=0}^{j} \psi_i)\varepsilon_{t+k-j}$. Writing this as a triangular array as in Fig. 1, and examining the various diagonals, it is apparent that the k-step-ahead prediction error converges to $S_k = \Sigma_{i=0}^{k-1} \psi(L)\varepsilon_t$. Thus, $\mathrm{var}(X_{t+k} - E_t X_{t+k}) = O(k)$, by our theorem on partial sums of $I(0)$ series. In particular,

$$\mathrm{var}(X_{t+k} - E_t X_{t+k}) = \sum_{j=0}^{k-1} \sum_{k=0}^{k-1} \left(\sum_{i=0}^{j} \psi_i\right)\left(\sum_{i=0}^{k} \psi_i\right) \mathrm{cov}(\varepsilon_{t+k-j}, \varepsilon_{t+k-k}) = \sigma_\varepsilon^2 \sum_{j=0}^{k-1} \left(\sum_{i=0}^{j} \psi_i\right)^2 ,$$

so that $\lim_{k\to\infty} \mathrm{var}(X_{t+k} - E_t X_{t+k}) = \infty$.

(e) *Fractional noise* $(X_t = (1-L)^{-d}\varepsilon_t = B(L)\varepsilon_t,\ -1/2 < d < 1/2)$. The future value of the process is $X_{t+k} = B(L)\varepsilon_{t+k}$, so that $E_t X_{t+k} = b_k\varepsilon_t + b_{k+1}\varepsilon_{t-1} + \cdots$ and $(X_{t+k} + E_t X_{t+k}) = \Sigma_{i=0}^{k-1} b_i\varepsilon_{t+k-i}$. Thus, $\mathrm{var}(X_{t+k} - E_t X_{t+k}) = \sigma_\varepsilon^2 \Sigma_{i=0}^{k-1} b_i^2 = O(1)$, and $\lim_{k\to\infty} \mathrm{var}(X_{t+k} - E_t X_{t+k}) = \sigma_x^2$.

(f) $I(d)$ $(X_t = (1-L)^{-d}\psi(L)\varepsilon_t = B(L)\varepsilon_t,\ -1/2 < d < 1/2)$. The proof parallels precisely the covariance stationary fractional noise case above.

(g) *Fractional noise* $(\Delta X_t = (1-L)^{-d'}\varepsilon_t = B(L)\varepsilon_t,\ d' = d-1,\ 1/2 < d < 3/2)$. The future value is $X_{t+k} = X_t + \Sigma_{i=1}^{k} B(L)\varepsilon_{t+i}$, so that $E_t X_{t+k} = X_t + (\Sigma_{i=1}^{\infty} b_i)\varepsilon_t + (\Sigma_{i=2}^{\infty} b_i)\varepsilon_{t-1} + \cdots$ and $(X_{t+k} - E_t X_{t+k}) = \Sigma_{j=0}^{k-1} (\Sigma_{i=0}^{j} b_i)\varepsilon_{t+k-j}$. Using the same argument as in the $I(1)$ case, we see that this k-step-ahead prediction error converges to $S_k = \Sigma_{i=0}^{k-1} B(L)\varepsilon_t$. Thus, $\mathrm{var}(X_{t+k} - E_t X_{t+k}) = O(k^{2d'+1}) = O(k^{2d-1})$ and $\lim_{k\to\infty} \mathrm{var}(X_{t+k} - E_t X_{t+k}) = \infty$. Specifically, note that $(1-L)(1-L)^{d-1}X_t = \varepsilon_t$, so that

$$\mathrm{var}(X_{t+k} - E_t X_{t+k}) = \sum_{j=0}^{k-1} \sum_{k=0}^{k-1} \left(\sum_{i=0}^{j} b_i\right)\left(\sum_{i=0}^{k} b_i\right) \mathrm{cov}(\varepsilon_{t+k-j}, \varepsilon_{t+k-k}) = \sigma_\varepsilon \sum_{j=0}^{k-1} \left(\sum_{i=0}^{j} b_i\right)^2 .$$

(h) $I(d)$ $(\Delta X_t = (1-L)^{-d'}\psi(L)\varepsilon_t = B(L)\varepsilon_t,\ d' = d-1,\ 1/2 < d < 3/2)$. The proof parallels precisely the non-stationary pure fractional noise case (G) above.

3. An empirical example: Real interest rate forecasting

We illustrate our results with a substantive application to real interest rate forecasting.[2] Using Mishkin's (1992) data, generously supplied by Ric Mishkin, we estimate a long-memory forecasting model for the one-month real interest rate, and we compare its out-of-sample interval forecasts to those from a random walk.[3]

The nominal interest rate, the real interest rate, and the expected inflation rate are assumed to satisfy the Fisher equation, $i_t = rr_t + \pi_t^e$, where i_t is the nominal rate, rr_t is the real rate, and π_t^e is the expected inflation rate. We construct an ex ante real rate series by estimating an integrated moving average model of order one for the inflation rate (determined by experimentation), generating the expected inflation rate as the prediction from that model, and subtracting the expected inflation rate from the nominal interest rate.[4] Finally, using parameters estimated through December 1986, we generate the 'actual' 48-month path of the ex ante real rate from 1987.01–1990.12.

We compare the results of forecasting the ex ante real rate with two models: a random walk and a pure fractional noise. We consider extrapolation forecasts of the four-year future path of the real interest rate, 1987.01–1990.12. The random walk forecast of the future path is simply the December 1986 value, 3.19%. The 95% Gaussian prediction interval for the random walk forecast is $3.19 \pm 1.96 * 0.88 * k^{0.5}$, where k is the number of steps ahead.

Construction of the forecast from the fractional noise model is slightly more complicated. First we estimate the fractional-integration parameter (d) for the ex-ante real interest rate using the Geweke, Porter and Hudak (1983) (GPH) procedure, 1953.01–1986.12. The evidence against a unit root and in favor of mean-reverting long memory appears strong ($d = 0.61$, s.e. $= 0.17$). Then we forecast the real rate by casting the fitted fractional model in (infinite) autoregressive form, truncating the infinite autoregression as necessary at the beginning of the sample, and applying Wold's chain rule. To compute the associated prediction interval, we use the usual estimate of the k-step-ahead prediction error variance (call it $\hat{\sigma}_k^2$) based on the MA($k - 1$) representation of the k-step-ahead prediction error. The 95% Gaussian confidence interval is centered at the point forecast, with width $1.96 * \hat{\sigma}_k$.

We show the point and interval forecasts from the fractional noise and random walk models in Fig. 2. The random walk prediction interval spans values from approximately −9% to +15%, whereas the fractional noise prediction interval is much tighter. Moreover, it seems that the random walk prediction interval is poorly calibrated – it appears much too wide. The traditional time-series framework, which allows only for integer differencing, does not acknowledge the potential for the mean reversion of series that are not covariance stationary (e.g. fractional noise with $1/2 < d < 1$), and the resulting prediction intervals may substantially overstate real interest rate forecast uncertainty.

[2] Throughout, 'real interest rate' refers to the ex ante real interest rate. The ex post real interest rate will be explicitly referenced as such.

[3] We use the monthly one-month US treasury bill rate and inflation rate data (annualized), derived from those of the Center for Research in Security Prices at the University of Chicago; for details see Huizinga and Mishkin (1986). We use 1953.01–1986.12 for estimation, and we use 1987.01–1990.12 for forecast comparison.

[4] This is obviously not the only way to approximate ex ante real rates. Fama and Gibbons (1982), for example, estimate the real rate using signal-extraction procedures, under the assumption that the real rate has a unit root. Here we impose no such assumption.

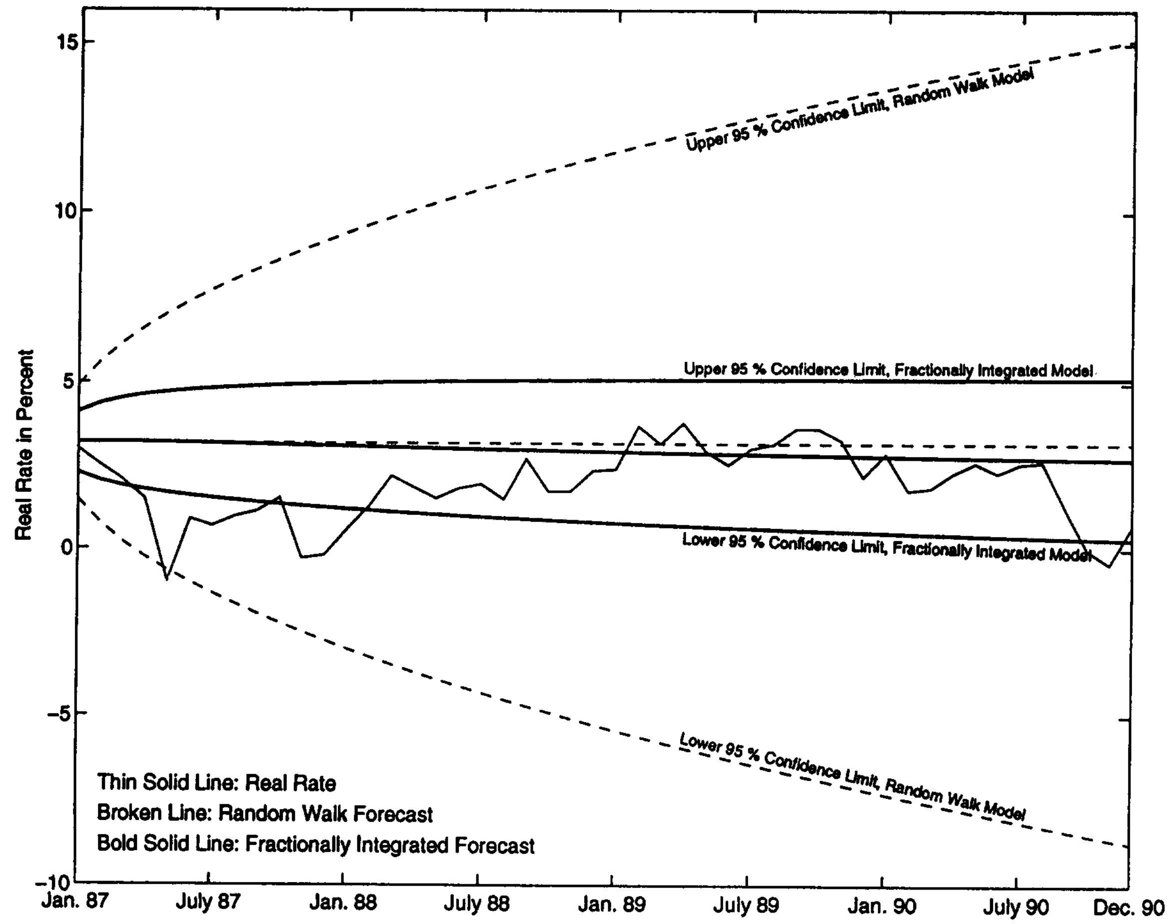

Fig. 2. Real rate forecasts; fractionally integrated vs. random walk model.

4. Concluding remarks

We have characterized the behavior of fractionally integrated processes in terms of long-horizon variance ratios and long-horizon prediction intervals, and we have shown that the behavior of long-horizon variance ratios and prediction intervals can be very different in long- vs. short-memory models.

We also provided an illustrative application to real interest rate forecasting. The application, of course, involved many debatable choices and assumptions, ranging from the model selected for forecasting inflation, to the choice of estimator for the long-memory parameter, to the absence of short-memory components in the forecasting models. However, our main point – that the behavior of long-horizon interval forecasts is very different in long- vs. short-memory models, and that differences in such behavior may have important implications for applied work – emerged clearly.

A variety of research avenues remain open. For example, prediction of long-memory processes via the Kalman filter appears impossible, as state-space representations evidently do not exist for long-memory models, because the spectral density function is not rational. Weiner–Kolmogorov prediction is straightforward conceptually, but the truncation that must be adopted in practice may be serious in smaller sample sizes, due to the long memory. A number of alternatives more sophisticated than simple truncation may be entertained, including the Durbin–Levinson innovations, and Wilson algorithms, as well as 'backcasting'. In addition, it will certainly be of interest to account for coefficient uncertainty (e.g. Sampson, 1991) in addition to innovation uncertainty when constructing point and interval forecasts in long-memory environments.

Acknowledgements

Parts of this work were completed at the University of Chicago, whose hospitality is gratefully acknowledged, as is the support from the Sloan Foundation, the National Science Foundation, and the University of Pennsylvania Research Foundation. The views expressed here are not necessarily those of Lehman Brothers or its staff.

References

Cochrane, J.H., 1988, How big is the random walk in GNP?, Journal of Political Economy, 96, 893–920.

Diebold, F.X. and M. Nerlove, 1990, Unit roots in economic time series: A selective survey, in: T.B. Fomby and G.F. Rhodes, eds., Advances in econometrics: Co-integration, spurious regressions, and unit roots, (JAI Press, Greenwich, CT) 3–69.

Fama, E.F. and M.R. Gibbons, 1982, Inflation, real returns and capital investment, Journal of Monetary Economics 9, 297–323.

Faust, J., 1992, When are variance ratio tests for serial dependence optimal?, Econometrica 60, 1215–1226.

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. and R. Joyeux, 1980, An introduction to long-memory time series models and fractional differencing, Journal of Time Series Analysis 1, 15–39.

Huizinga, J. and F.S. Mishkin, 1986, Monetary policy regime shifts and unusual behavior of real interest rates, Carnegie–Rochester Conference Series on Public Policy 24, 231–274.

Lo, A.W., 1991, Long-term memory in stock market prices, Econometrica 59, 1279–1313.

Mandelbrot, B.B., 1972, Statistical methodology for nonperiodic cycles: From the covariance to R/S analysis, Annals of Econometric and Social Measurement 1, 259–290.

Mishkin, F.S., 1992, Is the Fisher effect for real? A re-examination of the relationship between inflation and interest rates, Journal of monetary economics 30, 195–215.

Sampson, M., 1991, The effect of parameter uncertainty on forecast variance and confidence intervals for unit root and trend stationary time-series models, Journal of Applied Econometrics 6, 67–76.

Sowell, F., 1990, The fractional unit root distribution, Econometrica 50, 495–505.

Stock, J.H. and M.W. Watson, 1988, Variable trends in economic time series, Journal of Economic Perspectives 2, 147–174.