Macroeconomic Dynamics, **14** (Supplement 1), 2010, 59–87. Printed in the United States of America.
doi:10.1017/S1365100509991052

APPARENT LONG MEMORY IN TIME SERIES AS AN ARTIFACT OF A TIME-VARYING MEAN: CONSIDERING ALTERNATIVES TO THE FRACTIONALLY INTEGRATED MODEL

RICHARD A. ASHLEY AND DOUGLAS M. PATTERSON
Virginia Tech

Structural breaks and switching processes are known to induce apparent long memory in a time series. Here we show that any significant time variation in the mean renders the sample correlogram (and related spectral estimates) inconsistent. In particular, smooth time variation in the mean—i.e., even a weak trend, either stochastic or deterministic—induces apparent long memory. This apparent long memory can be eliminated by either high-pass filtering or by detrending. Here we demonstrate the effectiveness in this regard of nonlinear detrending via penalized-spline nonparametric regression. A time-varying mean can be of economic interest in its own right. This suggests that isolating out and separately examining both a local mean (i.e., a nonlinear trend or the realization of a stochastic trend) and deviations from it is preferable as a modeling strategy to simply estimating a fractionally integrated model. We illustrate the superiority of this strategy using stock return volatility data.

Keywords: Long Memory, Fractional Integration, Detrending, Stock Returns, Volatility

1. INTRODUCTION

The concept of "long memory" in a time series, usually coupled with fractional integration, has received much attention since it was first introduced into the literature by Granger (1980) and Granger and Joyeux (1980) to provide a theoretical explanation for the slow decay of sample correlograms in certain empirical contexts. The most relevant aspects of this literature are briefly summarized in Section 2 below; Beran (1994), Baillie (1996), and others provide detailed expositions.

In recent years a number of authors have suggested that the apparent long memory observed in some sample data is not generated by fractional integration but

Address correspondence to: Richard A. Ashley, Economics Department (0316), Virginia Tech, Blacksburg, VA 24061, USA; e-mail: ashleyr@vt.edu.

is instead an artifact of unmodeled nonlinear serial dependence and/or structural shifts in the generating mechanisms for these time series. Salient examples include Granger and Hyung (1999), Granger and Teräsvirta (1999), Diebold and Inoue (2001), Gourieroux and Jasiak (2001), Mikosch and Stărică (2004), Jensen and Liu (2006), Perron and Qu (2006), and Charfeddine and Guégan (2007). In contrast, Baillie and Kapetanios (2005) identify a number of cases where apparent long memory is still observed, even after nonlinear serial dependence is modeled.

In this paper we suggest a broader interpretation of apparent long-memory behavior: we take apparent long memory as being generically caused by time variation in the population mean of the series. The literature alluded to above has focused on sudden variations in the mean, which then persist for a lengthy period: what one might call "structural breaks." Here we suggest that apparent long memory is also—and probably more commonly—caused by long, smooth variation in the mean.[1] We call such variation—whether linear or nonlinear, stochastic or deterministic—"trendlike" behavior.

In Section 2 we provide a simple proof of the proposition that *any* time variation in the mean renders the sample autocorrelation estimates inconsistent.[2] This result reconciles the observance of sample long-memory behavior—in the sense of nonnegligible sample autocorrelations at extremely long lags—with the common-sense notion that fluctuations in the distant past are hardly likely to engender current fluctuations in a time series. That section further discusses what meaning ought sensibly be attached to the term "trendlike" and provides a second, also very simple, theoretical result demonstrating that arbitrarily smooth trendlike behavior yields an arbitrarily slowly decaying sample correlogram.

The analysis and isolation of trendlike behavior, using a variety of moving average filters, was a central feature of time series analysis prior to the 1970's; Kendall et al. (1983, Chapter 46) provide a detailed summary of this standard literature. Thus, one approach—employed in Ashley and Patterson (2007)—is to observe that high-pass filtering of data that appear to exhibit long memory eliminates the sample evidence for fractional integration in the series.[3]

Here, in contrast, we focus on directly modeling whatever trendlike behavior exists in a time series using nonparametric regression methods—in particular, penalized spline regression techniques, which are briefly reviewed in Section 3.

This nonparametric detrending method is applied in Section 4 to several artificially generated time series, each of which exhibits either weak trendlike behavior or a mild structural break and, concomitantly, significant long-memory behavior. The apparent long-memory behavior is in each case assessed using both the Geweke/Porter-Hudak (1983) and the Robinson (1995) tests, both of which estimate the fractional integration exponent based on the low-frequency behavior of the estimated spectrum of the time series. We find in each case that the apparent long-memory behavior disappears once a nonlinear trend is removed from the data.

Similar results are obtained in an empirical example using weekly stock return volatility data in Section 5.[4] Finally, in Section 6, we conclude by noting that the nonlinear detrending framework described here is likely a more fruitful modeling approach than routinely restricting one's attention to fractionally integrated formulations. This is because—although it is impossible to distinguish a deterministic trend statistically from the realization of a stochastic trend in any single sample, even a very long one—an explicit consideration of the time variation in the local mean may itself be of substantial economic interest.

2. FRACTIONAL INTEGRATION

A time series that is nonstationary [in the sense in which Box and Jenkins (1976) use the term] is said to be "integrated of order d" if the dth difference of the series is a stationary process. Such time series are referred to in the literature as I(d) processes and the underlying autoregressive—moving average process is in such cases designated as an ARIMA(p, d, q) model. A more general form of these models, called *fractionally integrated*, allows the differencing order d to take noninteger values. Such fractionally integrated processes are designated as ARFIMA(p, d, q) models. The concept of fractionally integrated processes, first introduced by Granger (1980) and Granger and Joyeux (1980), has attracted a good deal of subsequent attention. For example, see the reviews by Beran (1994) and Baillie (1996) cited in the preceding section.

If the parameter d is positive but less than one-half, then the fractionally integrated process is stationary in the mean—i.e., has a constant unconditional mean—but exhibits what is called "long-memory" behavior. This behavior can be defined in terms of slowly decaying autocorrelation at lag k,

$$\rho_k \propto k^{2d-1} \quad \text{as} \quad k \to \infty, \tag{1}$$

or—as Beran (1994) has shown is equivalent under certain conditions—in terms of an exploding spectral density at zero frequency,

$$s(\omega) \propto \omega^{-2d} \quad \text{as} \quad \omega \to 0^{+}. \tag{2}$$

In the remainder of this section the analysis will focus primarily on the slowly decaying autocorrelation aspect of long-memory behavior, but it is worth noting that the fractional integration exponent (d) is most conveniently estimated and tested via equation (2), by examining the slope of the logarithm of the estimated spectrum, as in the Geweke/Porter-Hudak (1983) or "GPH" test and the test later proposed by Robinson (1995).

The fact that ρ_k decays very slowly as the lag k becomes large implies that the current value of the series is linearly (albeit weakly) related to its own distant past. This follows intuitively from the fact that a fractionally integrated process can be viewed as the limit of a very large-order AR(p) or MA(q) process whose weights decline very slowly. For example, Hamilton (1994, pp. 448–449) provides the MA(∞) expansion for $(1 - B)^d y_t = u_t$ as

$$y_t = \sum_{j=0}^{\infty} h_j u_{t-j}, \qquad (3)$$

where $h_j = (d + j - 1)(d + j - 2) \cdots (d + 1)(d)/(j!)$, so that $h_{j+1}/h_j = (j + d)/(j + 1)$. This MA($\infty$) expansion implies that the mean value of y_t, conditional on its past, evolves very slowly and very smoothly through time.

Granger and Joyeux (1980) originally motivated the concept of fractional integration by considering the aggregation of k independent cross-sectional AR(1) processes; if there is no fortuitous cancellation of lag operators, this aggregate is an ARMA($k, k - 1$) process, which approaches a fractionally integrated process as k becomes large.[5] But it is fair to say that the continuing and growing interest in fractionally integrated processes stems from the observation that a number of macroeconomic and financial time series exhibit what appears to be long-memory behavior in their sample autocorrelation functions and in sample estimates of their power spectra.

3. TRENDLIKE BEHAVIOR IN A TIME SERIES AND HOW IT INDUCES APPARENT LONG MEMORY

3.1. Trendlike Behavior in a Time Series

Prior to Box and Jenkins (1976) it was commonplace to decompose a nonseasonal time series into a "trend" component and an "irregular" component. As noted in Section 1, Kendall et al. (1983, Chapter 46) provide an extensive review of this standard literature. In this section we define trendlike behavior in a time series either as any smooth deterministic variation in its mean or as any stochastic variation in its mean that is so smooth, relative to the length of the sample available, that one has no choice but to condition upon it. [We suggest reserving the term "structural break" for such variations which are not smooth. Because such structural breaks are already known to induce the appearance of long-memory behavior in a time series—e.g., Jensen and Liu (2006) and Perron and Qu (2006)—this section focuses on the role of trendlike behavior, as defined above.]

The following climatological example clarifies this definition of trendlike behavior. Figure 1 is a time plot of $\delta^{18}O_t$—the percentage of total oxygen in the form of the heavier isotope (O^{18})—in a typical set of ocean sediment core samples. These particular data are from a DSDP (Deep Sea Drilling Project) site located in the North Atlantic, roughly halfway between New York and Madrid. This isotopic percentage is of interest because it correlates directly with the volume of seawater tied up in ice sheets and thus correlates inversely with the average planetary temperature. The data are sampled at equal intervals of length 2 kiloyears; the sample consists of 1,634 such observations, spanning a sample period from 3,264 kiloyears B.C. to 2 kiloyears A.D.[6]

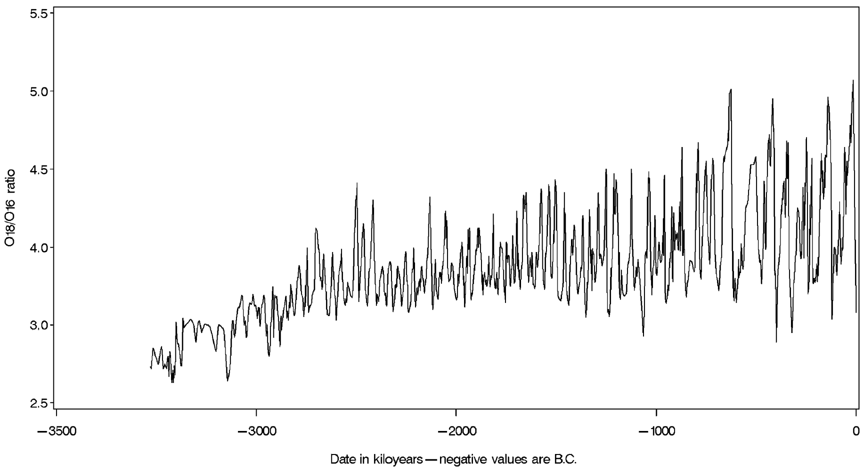

FIGURE 1. Time plot of $\delta^{18}O_t$: The percentage of total oxygen in the form of O^{18} 3,264 kiloyears B.C. to 2 kiloyears A.D., sampled at 2-kiloyear intervals.[7]

Evidently, the Earth has been cooling off for the last three million years or so: this time series is clearly dominated by an upward trend, whose slope changed fairly abruptly around 2,700 kiloyears ago in what one might reasonably term a "structural break." A closer examination of these data reveals the presence of irregular cycles—recently more pronounced—which climatologists call "ice ages." These ice ages vary in length, on a time scale of 50 to 150 kiloyears. Note, however, that if our data set were instead sampled every century over a sample period of just 100 kiloyears or so, then our data would only reflect one of these irregular cycles and the same variation that appears to be a realization of a stochastic process in the full data set would appear to be a trend in our 100 kiloyear–long subsample.

Moreover, note that the Earth's oceans are currently thought to be well over 4,000,000 kiloyears old, so this entire 3,264 kiloyear-long data set is actually less than one-tenth of one percent of the potential sample. Thus, the "obvious" upward trend alluded to above may be part of a much, much longer (albeit nonlinear) upward trend or could equally well be a realization of a seemingly smooth fluctuation in a stochastic process that evolves in a noisy and/or oscillatory manner over a time scale of tens or hundreds of millions of years. Absent a vastly longer sample—or a perhaps-naive faith in a theoretical model—these two possibilities are observationally indistinguishable and an analyst using only the data plotted in Figure 1 has no real choice but either to treat this apparent upward tendency in the time series as a deterministic time trend or to, in any case, analyze these data conditional on this observed trend.[8]

This example illustrates the fact there is no non-faith-based way to distinguish realized sample variation in a stochastic process that fluctuates on a time scale similar to that of the sample length from the sample variation of a deterministic trend. Indeed, this distinction is operationally meaningful only where one can in some well-defined sense obtain repeated samples of the process over the same time interval: in that case, and in that case only, can one directly observe whether the "trend" varies substantially across the repeated samples.[9] For an ergodic time series, one can indirectly observe whether the trend varies across repeated samples by examining the time variation of the series over a sample period quite substantially longer than the purported trend variations themselves, but that is precisely the kind of sample length that is not available in this instance.

Therefore—because there is really no choice in the matter—we define any variation in a time series that is so smooth that only one or two "fluctuations" can be discerned over the entire sample as trendlike behavior and propose to treat it as deterministic variation in the mean of the time series—either because it actually *is* deterministic or because we are forced to condition our analysis on it due to the limited length of the available sample.

3.2. Implications of Trendlike Behavior for the Sample Correlogram and for Apparent Long Memory

This section provides two simple theorems that illuminate the impact of trendlike behavior on the sample correlogram and indicate how such behavior can cause a time series to appear to be fractionally integrated.

THEOREM 1. *If* $\{X_t, t = 1, T\}$ *is any time series for which* $E[X_t]$ *varies over time, then the sample autocorrelation at lag* (r_k) *is not a consistent estimator of the population autocorrelation at lag* $k(\rho_k)$.

Proof. The proof of this proposition is fundamental and follows immediately from the definition of the sample autocorrelation of X_t with X_{t-k}—i.e., r_k—and the definition of the population autocorrelation of X_t with X_{t-k}—i.e., ρ_k:

$$r_k = \frac{\sum_{t=k+1}^{x} (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^{T} (X_t - \bar{X})^2} \tag{4}$$

and

$$\rho_k = \frac{E[(X_t - E[X_t])(X_{t-k} - E[X_{t-k}])]}{E[(X_t - E[X_t])^2]}. \tag{5}$$

If $E[X_t]$ and $E[X_{t-k}]$ are not equal, then the probability limit of $\bar{X}$ cannot possibly equal both $E[X_t]$ and $E[X_{t-k}]$; therefore, $\text{plim}(r_k)$ plainly cannot equal ρ_k.

Theorem 1 explains why the sample correlogram is an unreliable reflection of the population correlogram of a series that has trendlike behavior, even in very large samples. Obviously, spectral estimates not specially adapted for $E[X_t]$ variation will fare no better.

Theorem 2 shows how smooth, trendlike behavior yields sample correlogram estimates that necessarily exhibit apparent long-memory behavior:

THEOREM 2. *Suppose that*

$$X_t = \tau_t + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{MA}(q), \tag{6}$$

where ε_t is an MA(q) process with autocovariance function γ_k and τ_t is a deterministic trend. (In other words, because X_t has a trendlike component, it is observed conditional on this component.) Then

$$\mathrm{plim}[\hat{\mathrm{cov}}(X_t, X_{t-k})] = \gamma_k + \lim_{T\to\infty} \{\hat{\mathrm{cov}}(\tau_t, \tau_{t-k})\} \tag{7}$$

for all values of k, where T is the sample length, "$\hat{\mathrm{cov}}$" denotes the sample covariance, and "$\hat{\mathrm{cov}}$" is replaced by "$\hat{\mathrm{var}}$" for k equal to zero.

Proof. An elementary proof is given in the Appendix. In an obvious extension, the same result obtains if τ_t is a covariance stationary stochastic trend uncorrelated with $\varepsilon_{t\pm k}$, with the population autocovariance of τ_t at lag k in that case replacing the limit of the sample autocovariance in equation (7). If $[\tau_1, \ldots, \tau_T]$ is not a realization of a covariance stationary stochastic process, but is instead simply a fixed sequence, then Theorem 2 implicitly requires that the extrapolation of this sequence into the future must be such that its sample variance and sample autocovariances (viewed, in that case, as descriptive statistics) all approach finite limits.

Because ε_t is an MA(q) process, its autocovariances (γ_k) are zero for all lags k exceeding q. Thus, Theorem 2 implies that the large-lag decay behavior of the sample correlogram of X_t is completely determined, for large samples, by the rate at which the sample correlations of the trendlike component decay as the lag increases. But, by definition, any trendlike component in X_t varies so smoothly across the sample that its sample autocorrelations remain substantial even at lags as long as a substantial fraction of the total sample length. Thus, Theorem 2 implies that the existence of such a trendlike component necessarily induces apparent long memory in any long realization of the time series.

4. ELIMINATION OF APPARENT LONG MEMORY BY NONLINEAR DETRENDING: ARTIFICIAL DATA EXAMPLES

The trendlike behavior discussed above is clearly a low-frequency phenomenon. That is, it reflects sample variation that fluctuates so slowly that very few of these sample variations—perhaps only one or two—can occur during the course of the sample period. This observation immediately suggests that this trendlike behavior (and the apparent long memory that Theorem 2 shows it to cause) can be eliminated by using a filter designed to attenuate variation at low frequencies relative to variation at high frequencies—i.e., a high-pass filter.

As noted in Section 1, the removal of trendlike behavior via high-pass filtering has a lengthy and distinguished history in time series analysis; its application to removal of the apparent long memory caused by trendlike behavior is considered in Ashley and Patterson (2007). It is worth noting here, however, that the fractional difference operator is itself a crude high-pass filter. As noted above in equation (3) of Section 2, the lag operator $(1-B)^d$ is actually equivalent to a particular family of MA(∞) filters with smoothly (and very slowly) declining weights. This filter does indeed primarily remove the low-frequency variation in a time series, tending to eliminate any trendlike component and whatever apparent long memory it might be causing. Moreover, as Granger (1980) originally pointed out, this filter might be appropriate if the trendlike variation in the series were due to the aggregation of a large number of independent AR(1) components whose coefficients were independent draws from a particular form of the beta distribution. However, absent compelling evidence that aggregation artifacts of this particular form are the determining feature of the data-generating process, it seems unlikely that any analyst would have chosen this particular way to detrend the data. Rather, the fractional difference is now widely considered solely because fractional integration and, recently, structural breaks and the like are the only models that have been proposed to explain the apparent long-memory behavior actually observed in a number of financial and economic time series.[10]

The focus here is on demonstrating that one can equally well eliminate apparent long memory in a time series by directly detrending the data using now-standard nonparametric regression methods. This approach is more straightforward and also eliminates the argument that such high-pass filtering might be distorting the short-term or medium-term dynamics of the time series.

A number of nonparametric regression approaches have been proposed and implemented in the literature—e.g., kernel regression, nearest-neighbor regression, local polynomial regression, and penalized spline regression. Based on results in Ashley (2008) with similar data sets, nonlinear detrending models were identified and estimated using penalized spline regression methods.

The penalized spline method approximates the mean of the dependent variable by a spline function of the set of explanatory variables—i.e., by a piecewise-continuous patchwork of polynomials—to minimize an objective function that is the weighted sum of the squared fitting errors plus a penalty depending on an estimate of the average smoothness of the spline function. This smoothness penalty is parameterized as a linear function of the average value of the mth derivative of the spline function; the value of m is set to the minimum value possible, but must exceed half the number of explanatory variables in the model. Most implementations of penalized spline regression—e.g., Kauermann et al. (2008) and Krivobokova and Kauermann (2007)—restrict the analysis either to bivariate regression models or to additively separable models; the penalized spline regression implementation used here [Wahba (1990), implemented in SAS as procedure TPSPLINE] imposes neither of these restrictions. The effective number of degrees of freedom consumed in the fitting process increases sharply as the number of explanatory variables rises, however. This increase is reflected in an adjusted residual variance reported by the routine and in a resulting limitation on

TABLE 1. Data generation models[12]

Bilinear Trend	$x_t = u_t + 0.2u_{t-1} + 1.2\|(10{,}000 - 2\,\mathrm{mod}\,(t, 10{,}000) - 2)/10{,}000\|$
Random-Phase Sinusoidal Trend	$x_t = u_t + 0.2u_{t-1} + 1.1\sin[\pi/2 + y_t + (2\pi t)/20{,}000]$ $y_t = 0.7y_{t-1} + \sqrt{0.7}v_t$
Random Walk Mixture	$x_t = u_t + 0.2u_{t-1} + 0.01y_t$ $y_t = y_{t-1} + v_t$
Squared Random Walk Mixture	$x_t = (u_t + 0.2u_{t-1} + 0.01y_t)^2(1 + 0.0001t)$ $y_t = y_{t-1} + v_t$
Structural Break	$x_t = u_t + 0.2u_{t-1}, \quad t \leq 5{,}330$ $x_t = u_t + 0.2u_{t-1} + 0.3, \quad t > 5{,}330$
Fractional Integration (ARFIMA)	$(1 - B)^{0.3}x_t = u_t$

Note: In each case u_t and v_t are independent unit-normal variates. Each series was scaled to have sample variance equal to that generated by the ARFIMA model; each series except the Squared Random Walk Mixture was centered on its sample mean.

how many explanatory variables can be included in the model for a given sample length.

The smoothness penalty weight parameter is chosen here to minimize the ("one left out") generalized cross-validation function. The trend regression function is specified as a general nonlinear function of time and p lags of the dependent variable, with p chosen to minimize the adjusted residual variance of the resulting penalized spline model.

In the remainder of this section, this penalized spline nonlinear trend regression model is applied to several generated time series, each of which exhibits either mild trendlike behavior or a modest structural break, and each of which also exhibits statistically significant apparent long memory. In each case, filtering out a general nonlinear trend in this manner eliminates the apparent long memory in the time series. This demonstrates that there is nothing uniquely special about the fractional difference operator (or the fractionally integrated model) in this regard: in fact, any sufficiently flexible high-pass filter or nonlinear detrending method that one finds convenient and appealing will suffice to eliminate the apparent long memory caused by a time-varying mean—either trendlike behavior or a structural break—in a time series.

Table 1 lists the six data-generating models considered here. In each case the model was used to generate a single realization of length 10,000 observations.

The Bilinear Trend model in Table 1 was included as an example of a purely deterministic trend mechanism; the Random-Phase Sinusoidal Trend model was included as a first example of a mechanism with a stochastic trend. The Random Walk Mixture is a stochastically trended process of particular interest: this is a completely stochastic model in which a weak MA(1) process is mildly contaminated (or driven) by an I(1) covariate. The Squared Random Walk Mix process is the square of the Random Walk Mixture process plus a weak trend. It is of interest because its time plot resembles that of a typical financial returns volatility series—such as the weekly root-mean-square daily return to a broad stock market index analyzed in the next section. The generated Structural Break series is included both to remind the reader that such shifts in the mean are known to induce the appearance of long memory in a time series and to illustrate how nonlinear detrending eliminates this artifact also.

Finally, the fractionally integrated (ARFIMA) process in Table 1 is actually approximated by the MA(∞) process given in equation (3), truncated after 1,000,000 terms. [An MA(100,000) truncation did not yield an adequate approximation to the theoretical spectrum of the fractionally integrated process at the lowest frequencies.[11]] Of course, as would necessarily always be the case in practice, the empirical analysis proceeded conditional on the weak trend which this process induced in the sample.

The results obtained with the data generated from these six models are displayed in Figures 2 through 7. Each figure contains a time plot, a plot of the sample correlogram, and a plot of the estimated spectrum for one of these generated time series; in Figure 7 the theoretical spectrum for the fractionally integrated process is plotted with the symbol "+." The time plot in each case exhibits the fact that these data are, at worst, mildly trended; the sample correlogram in each case yields an informal indication of apparent long-memory behavior, in that the sample correlations do not promptly decay to lie within the usual Bartlett 95% confidence intervals. (One might note, however, that the Bartlett standard error estimates assume normality, which is certainly not the case for the data generated using the Squared Random Walk model.) In each case the spectral estimates are plotted only for the lowest frequencies—in the range from 0.0002 to 0.0050 cycles per observation—corresponding to fluctuations with periods ranging from 5,000 to 200 observations in length.[12]

We note that the estimated spectra of the data generated by all of these processes generally increase sharply for the low frequencies plotted here; the estimated low-frequency spectrum of the data generated as a mixture of a random walk and I(0) noise is increasing for low frequencies but appears to dip at the very lowest frequencies.[13]

Table 2 summarizes more formal evidence for fractional integration in each of these six time series, based on Geweke/Porter-Hudak (1983) and Robinson (1995) estimates of the fractional integration exponent, d, along with the p-values at which the null hypothesis H_o: $d = 0$ can be rejected on a one-tailed test against the relevant alternative hypothesis H_a: $d > 0$.[14]

Estimates and p-values are in each case given for the usual range of tuning exponent (θ) values for these tests: both the GPH and Robinson estimates are based on log-periodogram regressions [see equation (2)] utilizing only the first T^{θ} frequency ordinates. Typical values for θ are 0.5 (for the GPH estimates) and 0.8 (for the Robinson estimates).

Turning to the results displayed in Table 2, note that all of these estimates of d (and test p-values) would lead one to conclude the each of these time series is fractionally integrated.[15]

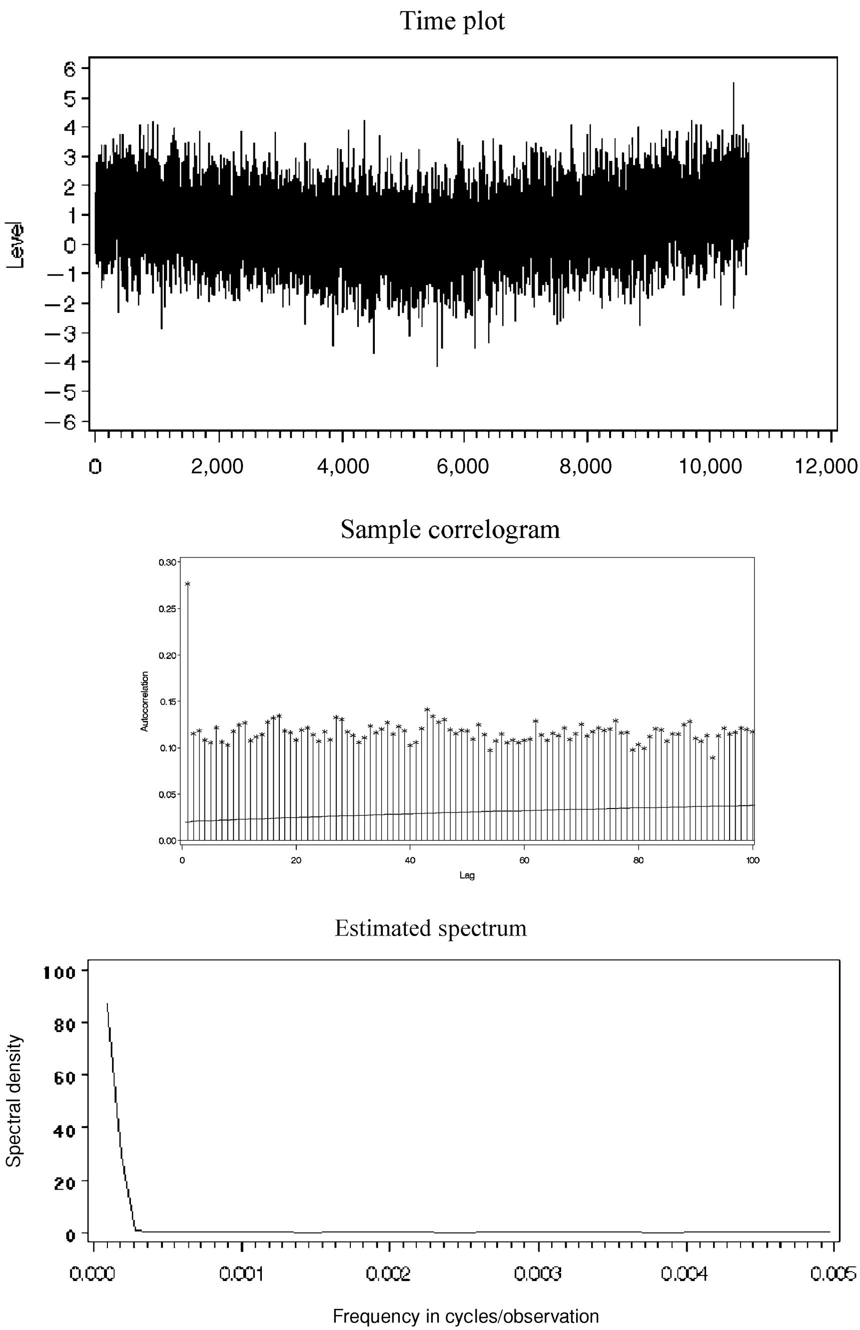

FIGURE 2. Results on data generated as MA(1) noise plus a bilinear trend.

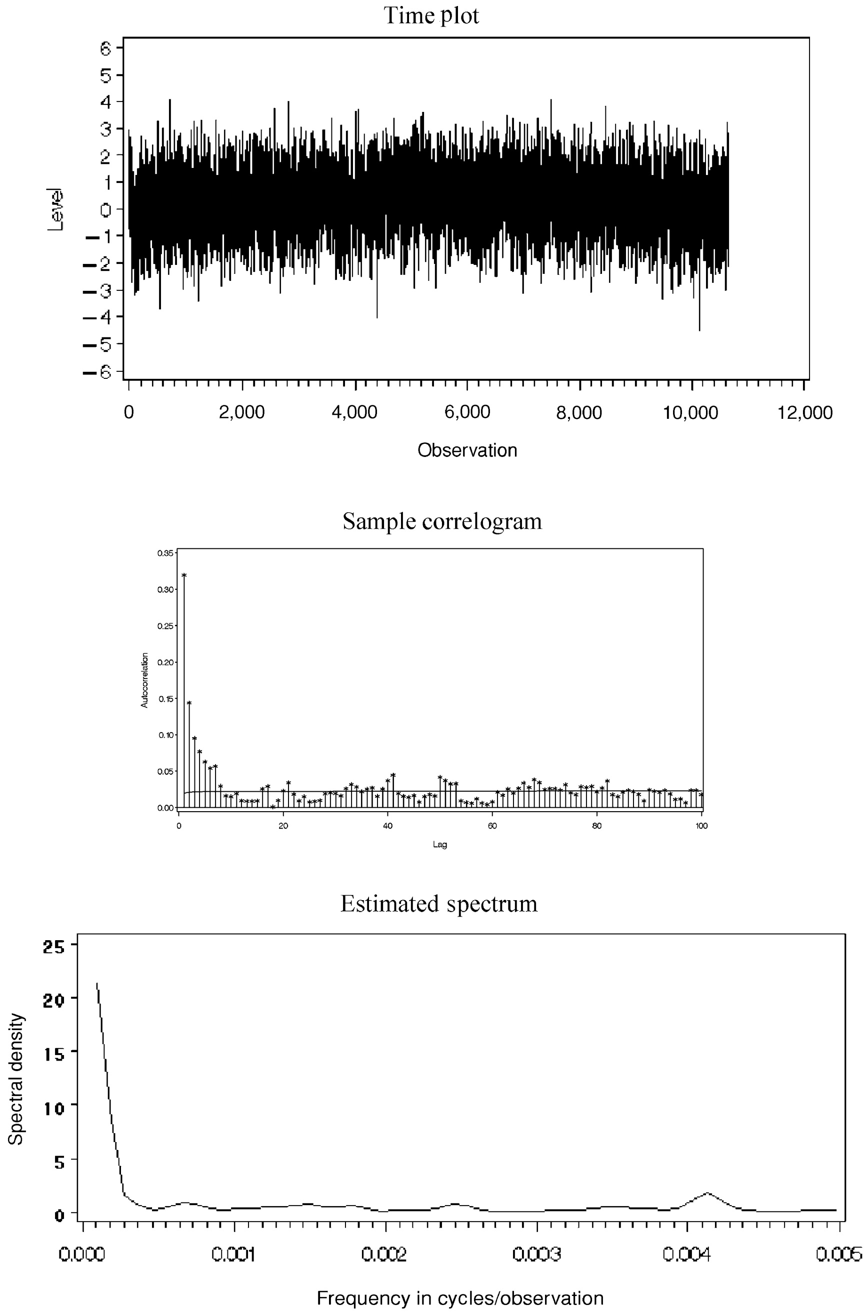

FIGURE 3. Results on data generated as MA(1) noise plus a random-phase sinusoidal trend.

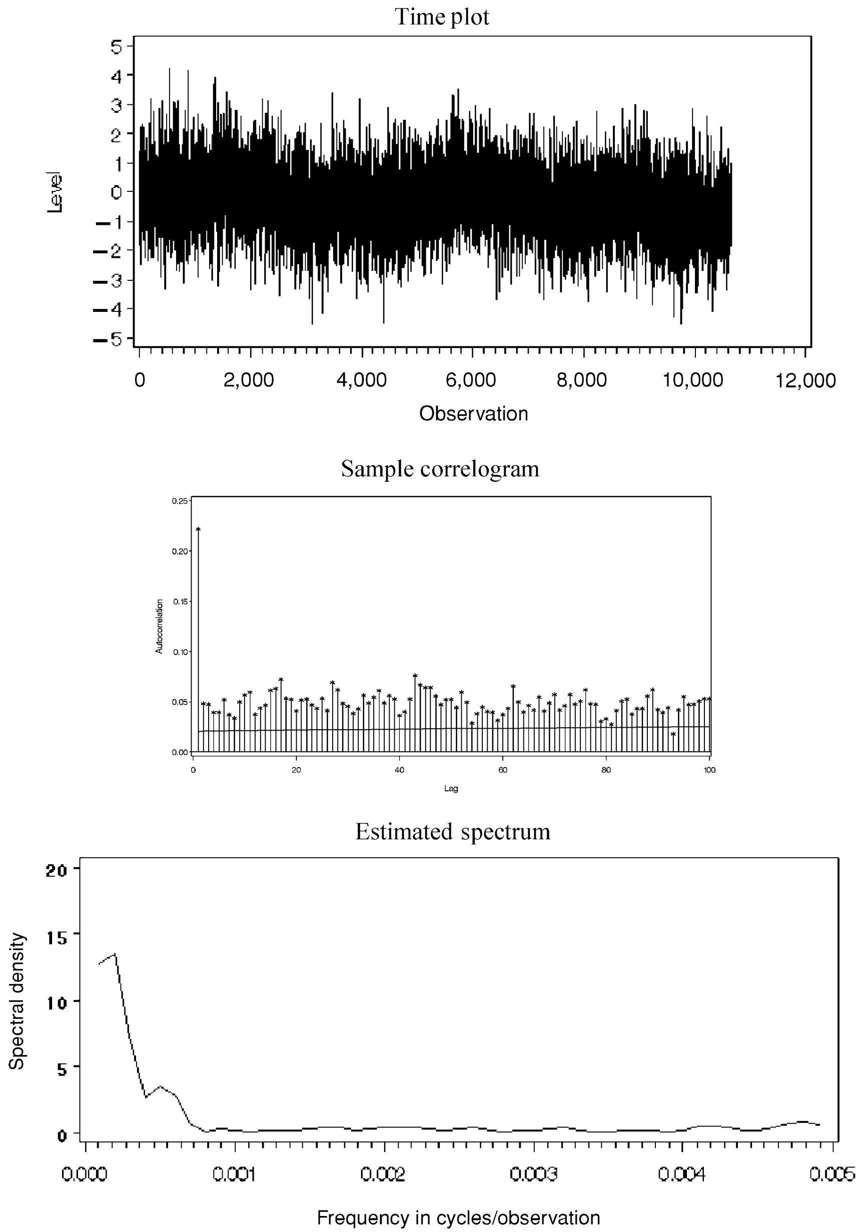

FIGURE 4. Results on data generated as MA(1) noise plus a (random walk) stochastic trend.

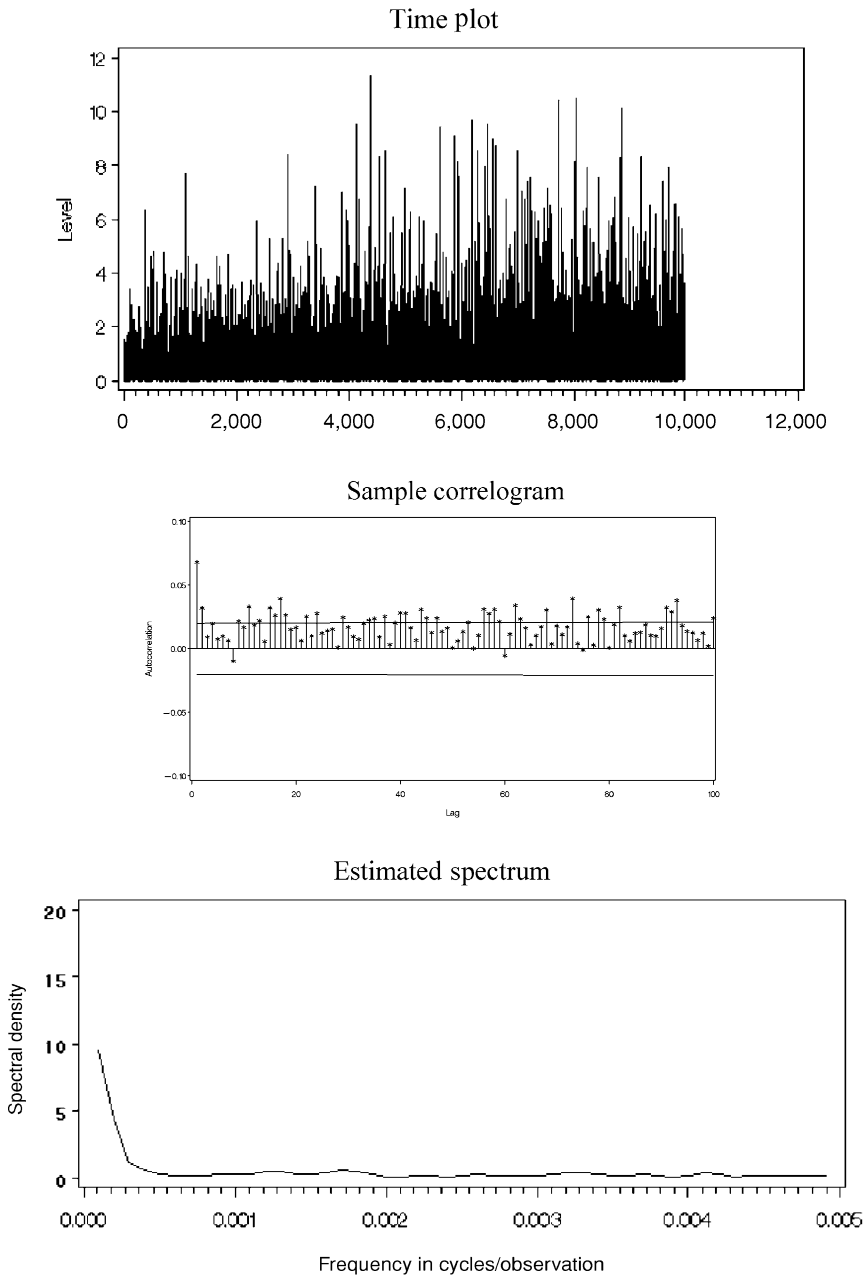

FIGURE 5. Results on data generated as the squared value of the sum of MA(1) noise and a (random walk) stochastic trend.

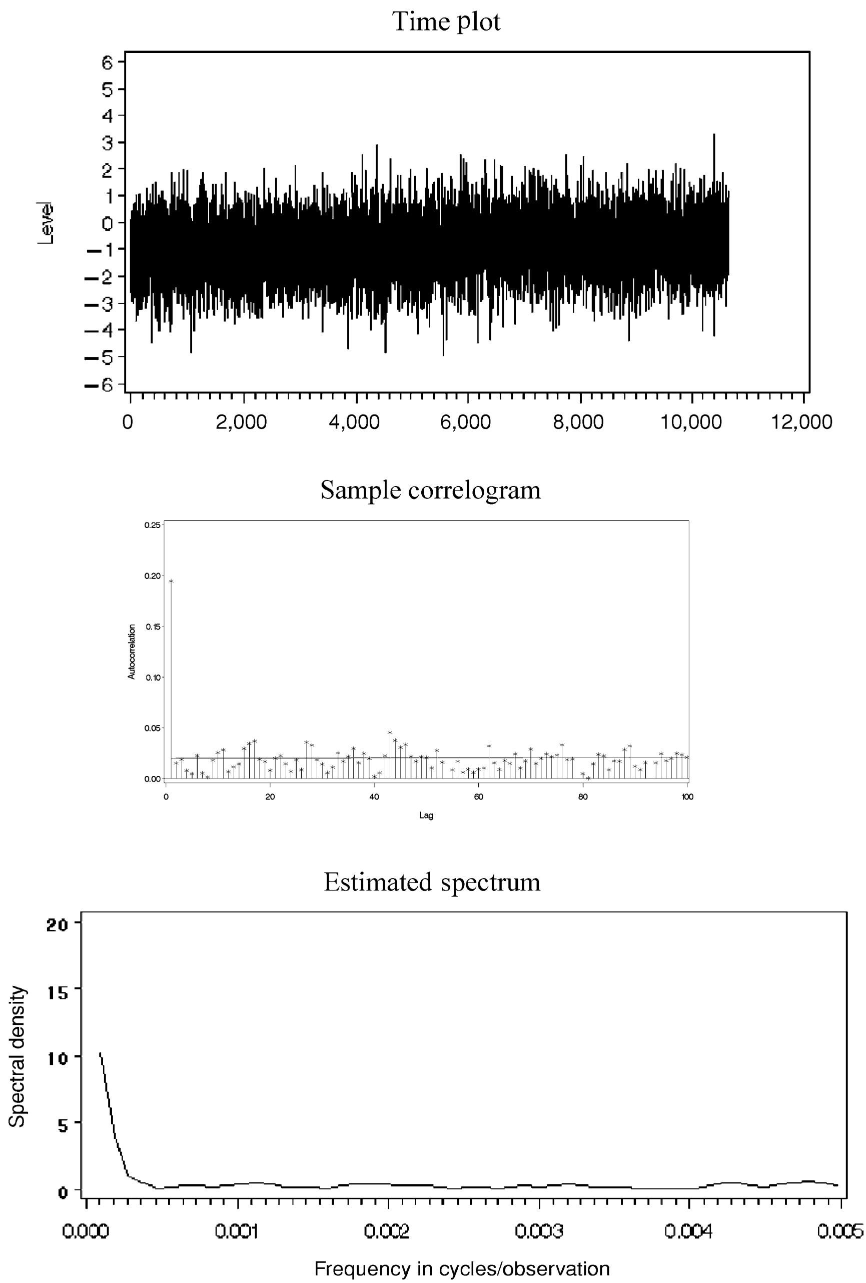

FIGURE 6. Results on data generated as MA(1) noise with a shift in the mean in the middle of the sample period.

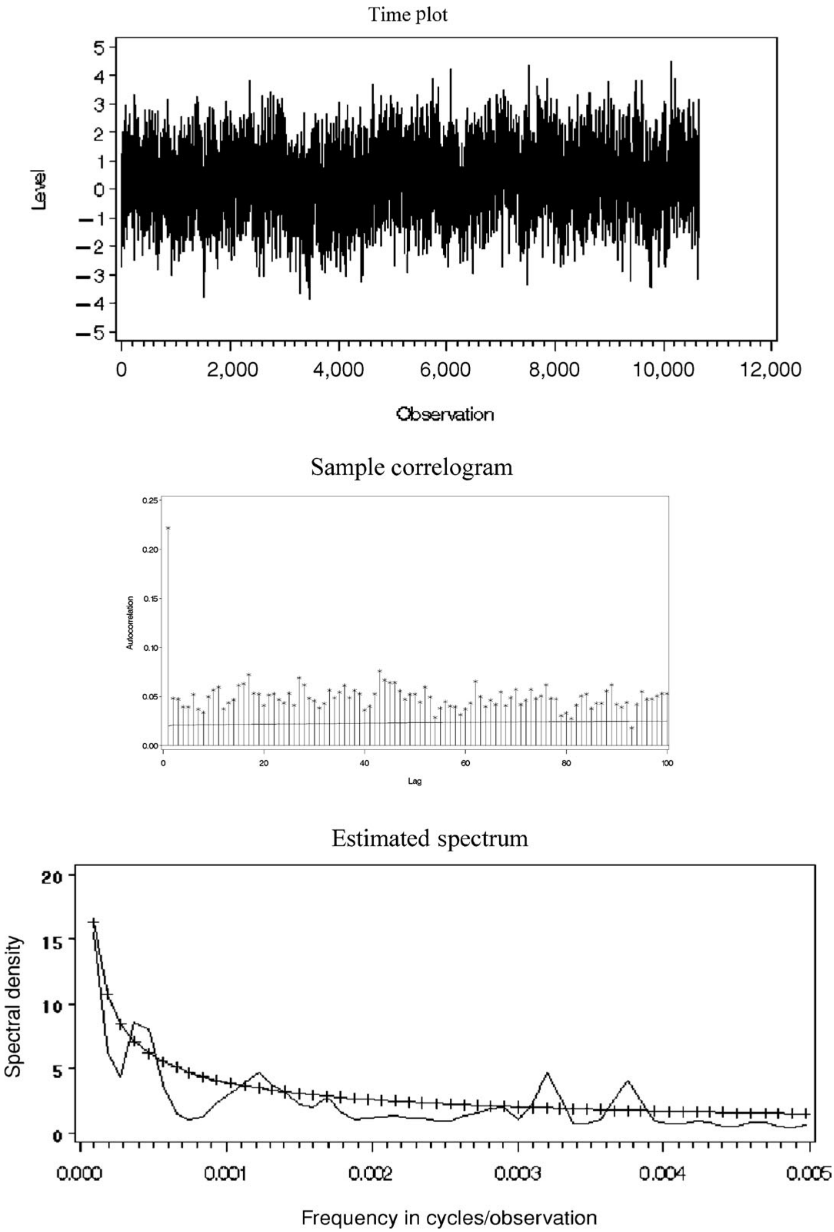

FIGURE 7. Results on data generated from fractionally integrated model.

TABLE 2. Fractional integration exponent estimates and inference results on one-tailed test of H_o: $d = 0$

Tuning exponent[a] (θ)	Bilinear Trend		Random-Phase Sinusoidal Trend		Random Walk Mixture		Squared Random Walk Mixture		Structural Break		Fractional Integration (ARFIMA)	
θ	d-hat	p-value	d-hat	p-value	d-hat	p-value	d-hat	p-value	d-hat	p-value	d-hat	p-value
					Geweke Porter-Hudak							
0.4	0.402	.005	0.417	<.001	0.581	<.001	0.428	.001	0.231	.023	0.310	.008
0.5	0.212	.004	0.147	.017	0.282	<.001	0.240	.001	0.190	.004	0.376	<.001
0.6	0.094	.013	0.047	.124	0.182	<.001	0.123	.002	0.106	.006	0.304	<.001
0.7	0.044	.045	0.027	.129	0.079	.001	0.061	.010	0.050	.029	0.296	<.001
0.8	0.037	.012	0.124	<.001	0.064	<.001	0.030	.033	0.033	.021	0.279	<.001
					Robinson							
0.4	0.402	.005	0.401	<.001	0.581	<.001	0.433	.001	0.261	.013	0.304	.007
0.5	0.210	.050	0.147	.017	0.283	<.001	0.250	.001	0.190	.004	0.376	<.001
0.6	0.094	.013	0.048	.120	0.182	<.001	0.123	.003	0.103	.007	0.304	<.001
0.7	0.044	.044	0.027	.146	0.078	.002	0.061	.013	0.050	.030	0.296	<.001
0.8	0.037	.011	0.124	<.001	0.063	<.001	0.030	.035	0.033	.021	0.276	<.001

[a]The GPH and Robinson estimates and tests are based on log-periodogram regressions, each using the lowest T^{θ} frequency ordinates; typical values for θ are 0.5 for GPH and 0.8 for Robinson.

Table 3 displays analogous results for these same data in detrended form, where a nonlinear trend was removed using a nonparametric model, identified, and fitted using penalized spline regression as described above. Note that essentially all evidence for fractional integration (and "long memory" has evaporated.[16]

Indeed, the detrending even eliminates all evidence of long memory in the data that were generated from an MA(1,000,000) model closely approximating a fractionally integrated process. The fact of the matter is that the entire impact of actual fractional integration on a time series is to induce a weak, slow stochastic trend in the sample data. Because only a single realization (albeit quite a long one) is available, this single (realized) trend is indistinguishable from a deterministic trend.[17]

It is all well and good that detrending eliminates spurious detection of fractional integration. But one might object that also eliminating the detection of authentic fractional integration—should one believe that fluctuations from the remotely distant past can ever truly impact the present—is a bad thing. The empirical example given in the next section illustrates why one should nevertheless detrend one's data—and examine both the detrended series *and* the trend estimate—rather than blithely assuming a fractionally integrated process whenever long memory appears to be present.

5. ELIMINATION OF APPARENT LONG MEMORY BY NONLINEAR DETRENDING: WEEKLY VOLATILITY OF STOCK RETURNS

The observed volatility in the rate of returns to corporate stock frequently exhibits apparent long-memory behavior. Consequently, such time series are commonly modeled as fractionally integrated processes. For example, in this section we examine the weekly volatility of U.S. stock returns, as measured by the root-mean-square (rms) value of the daily returns to the CRSP value-weighted stock index during the week.[18] The sample period used here consists of the 2,705 weeks from the first week in March 1956 to the last week in December 2007. A time plot of these data (Figure 8) is very similar in appearance to that of other volatility measures; it is also similar to that of the generated Squared Random Walk Mixture data (Figure 5) considered in Section 4.

As is typical of stock return volatility time series, the sample autocorrelations of the rms CRSP index returns do not promptly decay to lie within the usual Bartlett 95% confidence intervals (Figure 9).[19]

More formally, Table 4 summarizes statistical evidence for fractional integration in this time series, based on Geweke/Porter-Hudak (1983) and Robinson (1995) estimates of the fractional integration exponent, d, along with the p-values at which the null hypothesis H_o: $d = 0$ can be rejected on a one-tailed test against the relevant alternative hypothesis H_a: $d > 0$.[20]

Based on these results, most analysts would conclude that there is very strong evidence for fractional integration in this volatility time series. On the other hand, a plot of the estimated spectrum for this time series—like that of the data generated

TABLE 3. Fractional integration exponent estimates and inference results on one-tailed test of H_o: $d = 0$ (detrended data)

Tuning exponent[a] (θ)	Bilinear Trend		Random-Phase Sinusoidal Trend		Random Walk Mixture		Squared Random Walk Mixture		Structural Break		Fractional Integration (ARFIMA)	
θ	d-hat	p-value	d-hat	p-value	d-hat	p-value	d-hat	p-value	d-hat	p-value	d-hat	p-value
					Geweke Porter-Hudak							
0.4	−0.129	.859	−1.917	>.999	0.074	.338	−0.537	>.999	−0.075	.736	−2.110	>.999
0.5	−0.004	.523	−2.337	>.999	0.046	.281	−0.301	>.999	0.018	.404	−2.450	>.999
0.6	0.036	.210	−3.015	>.999	0.072	.058	−0.076	.963	0.042	.171	−2.975	>.999
0.7	0.016	.271	−3.230	>.999	−0.010	.641	−0.029	.869	0.015	.283	−3.075	>.999
0.8	−0.002	.546	−2.364	>.999	−0.019	.873	−0.014	.797	−0.000	.503	−2.269	>.999
					Robinson							
0.4	−0.105	.809	−1.916	>.999	0.074	.338	−0.537	>.999	−0.043	.642	−2.120	>.999
0.5	−0.004	.523	−2.337	>.999	0.046	.281	−0.301	>.999	0.018	.402	−2.450	>.999
0.6	0.034	.224	−3.015	>.999	0.073	.057	−0.076	.947	0.040	.184	−2.976	>.999
0.7	0.016	.271	−3.225	>.999	−0.110	.656	−0.030	.860	0.015	.283	−3.070	>.999
0.8	−0.002	.540	−2.337	>.999	−0.019	.873	−0.013	.792	−0.000	.504	−2.242	>.999

[a] The GPH and Robinson estimates and tests are based on log-periodogram regressions, each using the lowest T^{θ} frequency ordinates; typical values for θ are 0.5 for GPH and 0.8 for Robinson.

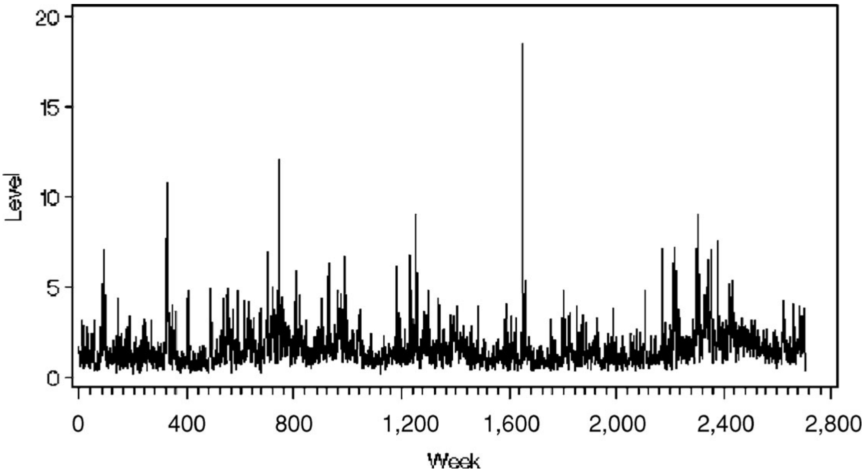

FIGURE 8. Time plot of weekly RMS volatility of returns to the CRSP value-weighted stock index.

as the sum of a random walk and an I(0) noise series examined in the previous section[21]—actually dips, rather than exploding, at the very lowest frequencies (see Figure 10).

Moreover, when this volatility time series is detrended—using a nonparametric model estimated using the penalized spline method, as described in Section 4—all evidence for fractional integration (and "long memory") completely evaporates. In particular, Table 5 summarizes the statistical evidence for fractional integration in

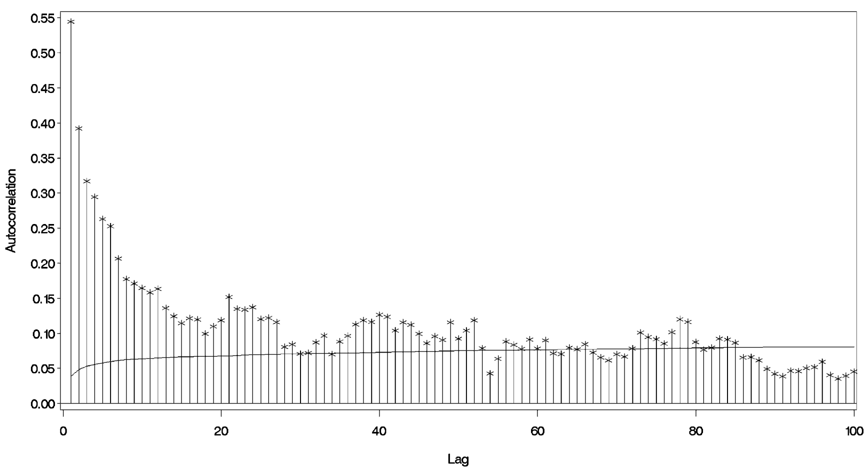

FIGURE 9. Sample correlogram of weekly RMS volatility of returns to the CRSP value-weighted stock index.

TABLE 4. Fractional integration exponent estimates and inference results for one-tailed test of H_o: $d = 0$ on volatility data

Tuning exponent (θ)[a]	d-hat	p-value
Geweke/Porter-Hudak results		
0.4	0.343	.049
0.5	0.224	.017
0.6	0.255	<.001
0.7	0.313	<.001
0.8	0.326	<.001
Robinson results		
0.4	0.343	.043
0.5	0.208	.022
0.6	0.256	<.001
0.7	0.309	<.001
0.8	0.321	<.001

[a] The GPH and Robinson estimates and tests are based on log-periodogram regressions, each using the lowest T^{θ} frequency ordinates; typical values for θ are 0.5 for GPH and 0.8 for Robinson.

this detrended time series, based on Geweke/Porter-Hudak (1983) and Robinson (1995) estimates of the fractional integration exponent, d, along with the p-values at which the null hypothesis H_o: $d = 0$ can be rejected on a one-tailed test against the relevant alternative hypothesis H_a: $d > 0$.

Note, however, that something quite interesting emerges from an explicit consideration of the estimated nonlinear trend in the weekly stock return volatility data.[22] First, looking at the estimated spectrum of nonlinear trend, it is very

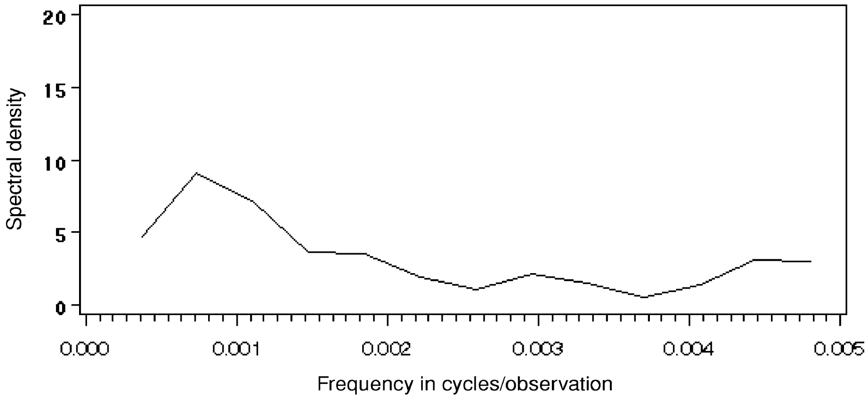

FIGURE 10. Estimated spectrum of weekly RMS volatility of returns to the CRSP value-weighted stock index.

TABLE 5. Fractional integration exponent estimates and inference results for one-tailed test of H_o: $d = 0$ on nonlinearly detrended volatility data

Tuning exponent $(\theta)^a$	d-hat	p-value
Geweke/Porter-Hudak results		
0.4	−2.019	>.999
0.5	−2.538	>.999
0.6	−2.874	>.999
0.7	−2.574	>.999
0.8	−1.768	>.999
Robinson results		
0.4	−2.019	>.999
0.5	−2.561	>.999
0.6	−2.873	>.999
0.7	−2.564	>.999
0.8	−1.727	>.999

[a] The GPH and Robinson estimates and tests are based on log-periodogram regressions, each using the lowest T^{θ} frequency ordinates; typical values for θ are 0.5 for GPH and 0.8 for Robinson.

similar in shape at low frequencies to the estimated spectrum of the volatility series itself (see Figure 11). Thus, it is reasonable to conclude that the nonparametric detrending procedure is not substantially distorting the time variation in what we have here called the trendlike behavior of the volatility time series.

Second, Figure 12 displays a time plot of the estimated trend in the volatility time series, with an expanded vertical scale to emphasize its time variation. Note

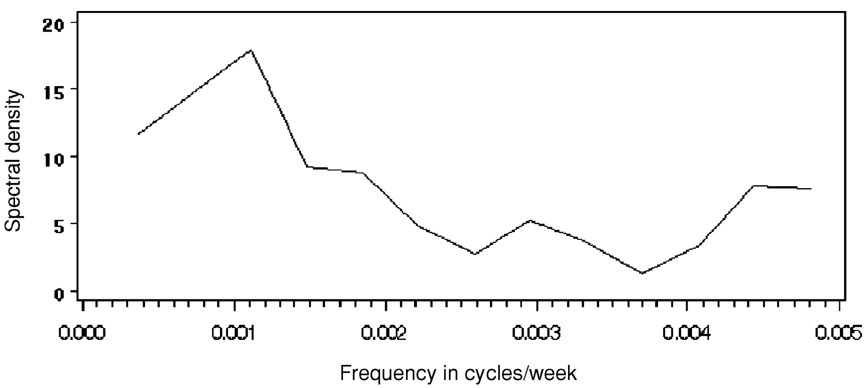

FIGURE 11. Estimated spectrum of nonlinear trend in weekly RMS volatility of returns to the CRSP value-weighted stock index.

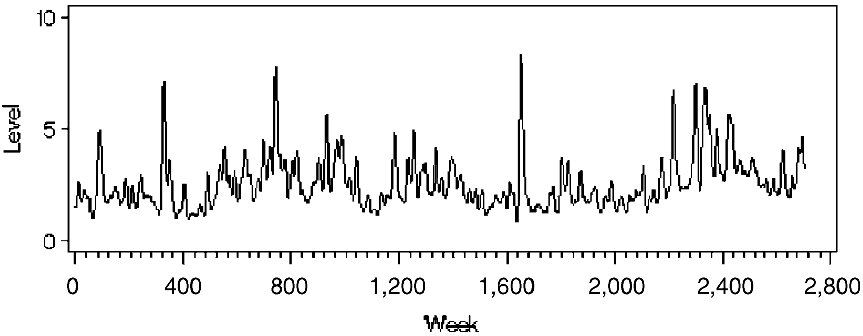

FIGURE 12. Estimated nonlinear trend in weekly RMS volatility of returns to the CRSP value-weighted stock index.

that it is evident that the time variation in the local mean of this volatility series is composed of a handful of sudden shifts ("structural breaks") set against a backdrop—over a 50-year period—of a substantial number of smooth variations ("local trends"). In particular, the peak around week 1,600 is clearly related to the crash of 1987. We also note that the volatility of the index rose to a broad peak around week number 2,400 (corresponding to the stock market bubble period of 1998–2001) and began rising in the last 80 or so weeks of the sample.

One could choose to ignore these patterns as meaningless random variations in a stochastic trend and therefore to model just the detrended data—whether the trend is removed by nonparametric modeling (as above), or by a moving-average high-pass filter [as in Ashley and Patterson (2007)], or by some other form of high-pass filtering.[24] That might be a reasonable choice for some. But the patterns in the nonlinear trend of this particular time series, to us, call out for further investigation and modeling. In particular, these features in the trendlike portion of the data move us to inquire why the local trends in aggregate stock return volatility varied in the way they did. We are pursuing these questions and expect to find answers which will inform insights into the behavior of the markets underlying these data. Surely, we conclude, any responsible analysis of data like these ought to include at least observing and investigating patterns such as these—even if one, in the end, declines to model them—in contrast to the practice of simply estimating an ARFIMA model when statistical tests indicate the presence of long memory.

6. CONCLUSIONS

The upshot of this paper is that the observation of long-memory behavior in a time series could be due to fractional integration—e.g., the result of a fairly particular family of weak but very high-order ARMA processes induced by certain forms of aggregation. But this sample statistical behavior could just as easily be due to occasional structural breaks or, as emphasized here, to the presence of a weak (perhaps nonlinear) trend, as might be caused by a slowly trending or I(1) covariate. Moreover, it is a consideration of this latter set of causes for sample long-memory behavior that is arguably most likely to produce progress in modeling the underlying economic/financial phenomena that generated the data.

Our view, therefore, is that one ordinarily should extract an estimate of this time-varying mean from the sample data—either to eliminate its influence on the sample statistics or for further (perhaps multivariate) analysis in its own right—using some kind of low-pass filter. This filter might be a sophisticated nonlinear bandpass filter—e.g., as in Baxter and King (1999)—or it might be a simple moving average, as in the "moving mean" model of Ashley and Patterson (2007), or it could be a nonparametric nonlinear trend regression, as used in Sections 4 and 5 above.[24] However, as illustrated by the empirical example given in Section 5, using the weekly volatility of the daily returns to the CRSP value-weighted stock price index, routinely estimating a fractionally integrated model for such data, and leaving it at that would seem to risk missing out on modeling potentially interesting and important aspects of the time series.

In other words, this paper suggests that it would be useful for the profession to shift its focus away from a consideration of fractionally integrated processes *per se* and toward what seem to us more crucial modeling questions for time series exhibiting apparent long-memory behavior:

- What is causing this failure of the sample correlogram to decay "properly"? Are structural breaks or a weak (perhaps nonlinear) trend important in this case?
- What kind of filtering or detrending method does the best job of eliminating the problem in this particular case? Are any of the conclusions we most care about sensitive to the choice?
- If, as we expect to often be the case in practice, the apparent long-memory behavior is due to some sort of weak trend, is this trend itself worthy of serious analysis rather than simply being filtered away?

More broadly, we suggest that apparent long memory in a time series X_t generically results from the fact that the sample autocorrelation at lag k is an inconsistent estimator of the population autocorrelations whenever $E[X_t]$ varies over time—fundamentally because $\bar{\mathbf{X}}$ cannot in that case consistently estimate $E[X_t]$. Such time variation in $E[X_t]$ could be due to structural breaks, to structural drift, to regime switching (as in SETAR or Markov switching models), to the evolution of an ordinary unit root process corrupted by substantial measurement error, or simply to the unfolding of a multivariate generating mechanism for X_t in which one or more driving variables drift slowly over time. This observation suggests that apparent long memory does not imply fractional integration but rather should be viewed as an invitation to consider the data's generating process more deeply.

NOTES

1. See Bhattacharya, et al. (1983) for an analogous insight in the context of the Hurst effect.

2. Spectral estimates are, of course, similarly corrupted by time variation in the population mean.

3. In fact, one could view the fractional difference operator as simply an example of such a high-pass filter. The expansion of the inverse of the fractional difference operator does correspond to an MA(∞) filter/process, one that Granger (1980) has argued might be appropriate for data in specific aggregation contexts. However, it seems reasonably evident that the fractional difference operator is not the high-pass filter most analysts would likely choose on any other grounds.

4. Weekly volatility is estimated using the root mean square of daily returns data from the Center for Research in Security Prices (CRSP) at the Graduate School of Business, University of Chicago.

5. Note, however, that the sum of k independent MA(q) processes is still an MA(q) process and that an AR(1) process, in contrast to an MA(q) process, is already serially correlated at large lags, albeit only weakly: aggregation merely lowers the rate at which these serial correlations decay with lag length.

6. See Raymo (1997) or Patterson and Ashley (2000, Chapter 11) for details.

7. Source: Raymo (1997).

8. The interested reader might find it useful to note that climatologists attribute the 50- to 150-kiloyear fluctuations as due to a nonlinear interaction between the Earth's climate and its average reflectivity—e.g., more extensive ice sheets reflect more sunlight back into space, lowering the mean temperature, increasing the ice sheet coverage, etc. The climatological fluctuations on a 1,000- to 3,000-kiloyear time scale are thought to arise from cyclical variations in the Earth's orbit and in the tilt of its rotational axis relative to its orbital plane.

9. A large-T panel data set on individuals differing only in, say, their regression equation intercept would provide such an opportunity.

10. Abadir et al. (2005) consider models that include both a fractional difference and a linear time trend; Beran and Ocker (2001) analyze models that combine both fractional integration and nonparametric trend estimation. We view the trend estimation in these models as arbitrarily assigning to the trend only that portion of the low-frequency variation that does not happen to correspond to the particular MA(∞) representation of a fractional difference operator.

11. Note that, were these daily data, a lag of 100,000 business days corresponds to ca. 500 years; in contrast, simulations of ARMA models rarely require lags in excess of a handful of years.

12. The six generated series were rescaled to have equal sample variances in order to ensure comparability of the spectral estimates. Also, in each case, the estimated spectrum was smoothed using a triangular filter with a base width of three spectral estimates. Results using a base width of five spectral estimates are not materially different.

13. The weak MA(1) term in each of the non-ARFIMA generating mechanisms makes a negligible contribution to the increase in the spectrum at low frequencies.

14. The one-tailed test is appropriate because one would not interpret a negative estimate of d as evidence for long memory in the time series.

15. Except for θ equal to 0.6 and 0.7 for the Random-Phase Sinusoidal Trend series.

16. The negative fractional integration exponent estimates in Table 3 simply indicate that the log-periodograms for these series are decaying rather than increasing as the frequency declines toward zero.

17. Thus, had we generated additional 10,000 period–long realizations of the "bilinear trend" and the "structural break" series, we would have observed the *same* trendlike pattern in each realization. In contrast, additional realizations of any of the four other processes would have yielded similar evidence for long memory, but different trendlike patterns for each realization.

18. The data are from the Center for Research in Security Prices (CRSP) at the Graduate School of Business, University of Chicago.

19. It should be noted that the Bartlett results explicitly assume Gaussianity in the data, however, which is certainly not the case here.

20. The one-tailed test is appropriate because one would not interpret a negative estimate of d as evidence for long memory in the time series.

21. See Table 1 and Figure 4.

22. Inclusion of lagged values of the volatility does not reduce the adjusted standard deviation of the model fitting errors, so the model consists solely of a nonlinear function of time plus an innovation term. Thus, the trend in this case is identical to the predicted value from the nonparametric model.

23. Such as a pass-band filter—e.g., Baxter and King (1999)—or even a fractional difference. See Hinich et al. (2008) for a critical discussion of the use of such filters.

24. Or one could even use a low-pass filter based on the fractional difference operator—i.e., $1 - (1 - B)^d$—for this purpose, although we strongly doubt that most analysts would choose that operator for this purpose.

REFERENCES

Abadir, K.M., W. Distaso, and L. Giriatis (2005) Semiparametric Estimation and Inference for Trending I(d) and Related Processes. Mimeo, Department of Economics, Imperial College London.

Ashley, R. (2008) On the Origins of Conditional Heteroscedasticity in Time Series. Mimeo, Economics Department, Virginia Tech. http://ashleymac.econ.vt.edu/working_papers/origins_of_conditional_heteroscedasticity.pdf.

Ashley, R. and D.M. Patterson (2007) Apparent Long Memory in Time Series as an Artifact of a Time-Varying Mean: A "Local Mean" Filtering Alternative to the Fractionally Integrated Model. Mimeo, Economics Department, Virginia Tech.

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

Baillie, R.T. and G. Kapetanios (2005) Testing for Neglected Nonlinearity in Long Memory Models. Unpublished manuscript, Department of Economics, Michigan State University.

Baxter, M. and R. King (1999) Measuring business cycles: Approximate band-pass filters for economic time series. *Review of Economics and Statistics* 81, 575–593.

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

Beran, J. and D. Ocker (2001) Volatility of stock-market indexes: An analysis based on SEMIFAR models. *Journal of Business and Economic Statistics* 19, 103–116.

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

Box, G.E.P. and G.M. Jenkins (1976) *Time Series Analysis*. San Francisco: Holden-Day.

Charfeddine, L. and D. Guégan (2007) Which Is Best for the US Inflation Series: A Structural Change Model or a Long Memory Process? CES-AC working paper 01-2007.

Diebold, F.X. and A. Inoue (2001) Long memory and regime switching. *Journal of Econometrics* 105, 131–159.

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.

Gourieroux, C. and J. Jasiak (2001) Memory and infrequent breaks. *Economics Letters* 70, 29–41.

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

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.

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

Hamilton, James D. (1994) *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hinich, Melvin J., John Foster, and Philip Wild (2008) An Investigation of the Cycle Extraction Properties of Several Bandpass Filters Used to Identify Business Cycles. School of Economics Discussion Paper No. 358, School of Economics, The University of Queensland, Australia.

Jensen, M.J. and M. Liu (2006) Do long swings in the business cycle lead to strong persistence in output? *Journal of Monetary Economics* 53(3), 597–611.

Kauermann, G., T. Krivobokova, and W. Semmler (2008) Filtering Time Series with Penalized Splines. Unpublished manuscript, New School for Social Research.
Kendall, M., A. Stuart, and J.K. Ord (1983) *The Advanced Theory of Statistics*, 4th ed., vol. 3. London: Charles Griffen.
Krivobokova, T. and G. Kauermann (2007) A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association* 102(440), 1328–1337.
Mikosch, T. and Cătălin Stărică (2004) Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics* 86(1), 378–390.
Patterson, D.M. and R. Ashley (2000) *A Nonlinear Time Series Workshop*. Boston: Kluwer Academic.
Perron, P. and Z. Qu (2006) An Analytical Evaluation of the Log-Periodogram Estimate in the Presence of Level Shifts and Its Implications for Stock Returns Volatility. Mimeo, Department of Economics, Boston University.
Raymo, M. (1997) The timing of major climate terminations. *Paleoceanography* 12(4), 577–585.
Robinson, P.M. (1995) Log-periodogram regression of time series with long range dependence. *Annals of Statistics* 23(3), 1048–1072.
Wahba, G. (1990) *Spline Models for Observational Data*. Philadelphia: Society for Industrial and Applied Mathematics.

APPENDIX: PROOF OF THEOREM 2

Suppose that

$$X_t = \tau_t + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{MA}(q),$$

where ε_t is an MA(q) process with autocovariance function γ_k and τ_t is a deterministic trend. (In other words, because X_t has a trendlike component, it is observed conditional on this component.) Then

$$\mathrm{plim}[\hat{\mathrm{cov}}(X_t, X_{t-k})] = \gamma_k + \lim_{T\to\infty}\{\hat{\mathrm{cov}}(\tau_t, \tau_{t-k})\}$$

for all values of k, where "côv" denotes the sample covariance and "côv" is replaced by "vâr" for k equal to zero.

Proof.

$$\hat{\mathrm{cov}}(X_t, X_{t-k}) = \frac{1}{T}\sum_{t=k+1}^{T}(X_t - \bar{X})(X_{t-k} - \bar{X}).$$

But

$$\bar{X} = \frac{1}{T}\sum_{j=1}^{T} X_j = \frac{1}{T}\sum_{j=1}^{T}(\tau_j + \varepsilon_j) = \bar{\tau} + \bar{\varepsilon},$$

so that

$$\begin{aligned}\hat{\mathrm{cov}}(X_t, X_{t-k}) &= \frac{1}{T-k}\sum_{t=k+1}^{T}([\tau_t + \varepsilon_t] - [\bar{\tau} + \bar{\varepsilon}])([\tau_{t-k} + \varepsilon_{t-k}] - [\bar{\tau} + \bar{\varepsilon}]) \\ &= \frac{1}{T-k}\sum_{t=k+1}^{T}([\tau_t - \bar{\tau}] + [\varepsilon_t - \bar{\varepsilon}])([\tau_{t-k} - \bar{\tau}] + [\varepsilon_{t-k} - \bar{\varepsilon}]),\end{aligned}$$

$$\begin{aligned}\hat{\text{cov}}(X_t, X_{t-k}) &= \frac{1}{T-k}\sum_{t=k+1}^{T}([\tau_t - \bar{\tau}] + [\varepsilon_t - \bar{\varepsilon}])([\tau_{t-k} - \bar{\tau}] + [\varepsilon_{t-k} - \bar{\varepsilon}]) \\ &= \hat{\text{cov}}(\varepsilon_t, \varepsilon_{t-k}) + \hat{\text{cov}}(\tau_t, \tau_{t-k}) \\ &\quad + \frac{1}{T-k}\sum_{t=k+1}^{T}[\tau_t - \bar{\tau}][\varepsilon_{t-k} - \bar{\varepsilon}] \\ &\quad + \frac{1}{T-k}\sum_{t=k+1}^{T}[\tau_{t-k} - \bar{\tau}][\varepsilon_t - \bar{\varepsilon}].\end{aligned}$$

Hence,

$$\begin{aligned}\text{plim}\{\hat{\text{cov}}(X_t, X_{t-k})\} &= \gamma_k + \lim_{T\to\infty}\{\hat{\text{cov}}(\tau_t, \tau_{t-k})\} \\ &\quad + \text{plim}\left\{\frac{1}{T-k}\sum_{t=k+1}^{T}[\tau_t - \bar{\tau}][\varepsilon_{t-k} - \bar{\varepsilon}]\right\} \\ &\quad + \text{plim}\left\{\frac{1}{T-k}\sum_{t=k+1}^{T}[\tau_{t-k} - \bar{\tau}][\varepsilon_t - \bar{\varepsilon}]\right\}.\end{aligned}$$

These last two probability limits are zero because each of these sample covariances converges in quadratic mean to zero:

$$\begin{aligned}&\lim_{T\to\infty}\left\{E\left(\frac{1}{T-k}\sum_{t=k+1}^{T}[\tau_t - \bar{\tau}][\varepsilon_{t-k} - \bar{\varepsilon}]\frac{1}{T-k}\sum_{s=k+1}^{T}[\tau_s - \bar{\tau}][\varepsilon_{s-k} - \bar{\varepsilon}]\right)\right\} \\ &= \lim_{T\to\infty}\left\{\frac{1}{(T-k)^2}E\left(\sum_{t=k+1}^{T}\sum_{s=k+1}^{T}[\tau_t - \bar{\tau}][\varepsilon_{t-k} - \bar{\varepsilon}][\tau_s - \bar{\tau}][\varepsilon_{s-k} - \bar{\varepsilon}]\right)\right\} \\ &= \lim_{T\to\infty}\left\{\frac{1}{(T-k)^2}\sum_{t=k+1}^{T}\sum_{s=k+1}^{T}[\tau_t - \bar{\tau}][\tau_s - \bar{\tau}]E\left([\varepsilon_{t-k} - \bar{\varepsilon}][\varepsilon_{s-k} - \bar{\varepsilon}]\right)\right\} \\ &= \lim_{T\to\infty}\left\{\frac{1}{(T-k)^2}\sum_{t=k+1}^{T}\sum_{s=k+1}^{T}[\tau_t - \bar{\tau}][\tau_s - \bar{\tau}]\gamma_{t-s}\right\}.\end{aligned}$$

Because $X_t \sim \text{MA}(q)$, γ_{t-s} is zero for $|t-s| > q$. Consequently, letting $j = t - s$, this double sum reduces to

$$\begin{aligned}&\lim_{T\to\infty}\left\{E\left(\frac{1}{T-k}\sum_{t=k+1}^{T}[\tau_t - \bar{\tau}][\varepsilon_{t-k} - \bar{\varepsilon}]\frac{1}{T-k}\sum_{s=k+1}^{T}[\tau_s - \bar{\tau}][\varepsilon_{s-k} - \bar{\varepsilon}]\right)\right\} \\ &= \lim_{T\to\infty}\left\{\frac{1}{(T-k)^2}\sum_{t=k+1}^{T}\sum_{s=k+1}^{T}[\tau_t - \bar{\tau}][\tau_s - \bar{\tau}]\gamma_{t-s}\right\}\end{aligned}$$

$$= \lim_{T\to\infty}\left\{\frac{1}{(T-k)^2}\sum_{t=k+1}^{T}\sum_{j=-q}^{q}[\tau_t-\bar{\tau}][\tau_{t-j}-\bar{\tau}]\gamma_j\right\}$$

$$= \lim_{T\to\infty}\left\{\frac{1}{(T-k)}\sum_{j=-q}^{q}\gamma_j \hat{\text{cov}}(\tau_t,\tau_{t-j})\right\}$$

$$= 0$$

so long as $\lim_{T\to\infty}\{\hat{\text{cov}}(\tau_t, \tau_{t-k})\}$ is finite. The proof that the other sample covariance converges in quadratic mean to zero is essentially identical. This completes the proof of the theorem.