Level Shifts and the Illusion of Long Memory in Economic Time Series

Aaron SMITH
Department of Agricultural and Resource Economics, University of California, Davis, CA 95616
(*adsmith@ucdavis.edu*)

When applied to time series processes containing occasional level shifts, the log-periodogram (GPH) estimator often erroneously finds long memory. For a stationary short-memory process with a slowly varying level, I show that the GPH estimator is substantially biased, and derive an approximation to this bias. The asymptotic bias lies on the (0, 1) interval, and its exact value depends on the ratio of the expected number of level shifts to a user-defined bandwidth parameter. Using this result, I formulate the modified GPH estimator, which has a markedly lower bias. I illustrate this new estimator via applications to soybean prices and stock market volatility.

KEY WORDS: Fractional integration; Mean shift; Structural break.

1. INTRODUCTION

The study of long memory in time series processes dates back at least to Hurst (1951), and Granger and Joyeux (1980) and Granger (1981) introduced it to econometrics. Long-memory models prove useful in economics as a parsimonious way of modeling highly persistent yet mean-reverting processes. They apply to stock price volatility, commodity prices, interest rates, aggregate output, and numerous other series. Recently, Diebold and Inoue (2001), Liu (2000), Granger and Hyung (1999), Granger and Ding (1996), Lobato and Savin (1998), Hidalgo and Robinson (1996), Breidt and Hsu (2002), and others have suggested that the apparent long memory in many time series is an illusion generated by occasional level shifts. If this suggestion is correct, then a few rare shocks induce the observed persistence, whereas most shocks dissipate quickly. In contrast, all shocks are equally persistent in a long-memory model. Thus distinguishing between long memory and level shifts could dramatically improve policy analysis and forecasting performance.

Despite much anecdotal evidence that implicates level shifts as the cause of long memory, the properties of long-memory tests in the context of level shifts are not well understood. In this article I show formally that a popular test for long memory is substantially biased when applied to short-memory processes with slowly varying means. This bias leads to the erroneous conclusion that these processes have long memory. I derive an approximation to the bias and show that it depends only on the ratio of the expected number of level shifts to a user-defined bandwidth parameter. This result illuminates the connection between long memory and level shifts, and leads directly to a simple method for bias reduction.

A long-memory process is defined by an unbounded power spectrum at frequency 0. The most common long-memory model is the fractionally integrated process, for which the elasticity of the spectrum at low frequencies measures the fractional order of integration, d. This feature motivates the popular log-periodogram, or GPH, regression (Geweke and Porter-Hudak 1983). Specifically, the GPH estimate of d is the slope coefficient in a regression of the log periodogram on two times the log frequency.

Figure 1 illustrates why estimators such as GPH may erroneously indicate long memory when applied to level-shift processes. Figure 1(a) shows the power spectra of a fractionally integrated long-memory process and a short-memory process with occasional level shifts. A sequence of Bernoulli trials determines the timing of the level shifts, and a draw from a distribution with finite variance determines each new level. The curves cover frequencies, ω, between 0 and $2\pi/\sqrt{1{,}000}$, which is the range recommended by a common rule of thumb for GPH regression with a sample of 1,000 observations. For all but the very lowest frequencies, the spectrum of the level-shift process closely corresponds to that of the fractionally integrated process.

Figure 1(b) shows the log spectrum evaluated at the log of the first 32 Fourier frequencies, that is, $\omega_j = 2\pi j/T, j = 1, 2, \ldots, 32, T = 1{,}000$. These points represent the actual observations that would be used in a GPH regression based on the aforementioned rule of thumb. I did not incorporate sample error into these graphs, but it is already apparent that $d = .6$ fits this level-shift process better than the true value of 0. The fact that the spectrum of the level-shift process is flat at frequency 0 only becomes evident at the very lowest Fourier frequencies. To obtain a GPH estimate near the true value of 0, one would need to focus on the far-left portion of the spectrum, either by including a very low number of frequencies in the regression or by increasing the sample size T.

Figure 1 demonstrates the dominant source of bias in the GPH estimator. The short-memory components create bias because the estimator includes parts of the spectrum away from 0. A second source of bias is induced by the use of the log periodogram, which is a biased estimator of the log spectrum. In Section 3 I demonstrate that the first source of bias clearly dominates the second source, even in small samples. Further, I show that when the shifts are rare relative to the sample size, the dominant component of the GPH bias converges to a value on the interval (0, 1) as the sample size grows. This result applies to a general class of mean-plus-noise processes, which I present in Section 2.

In Section 4 I use the asymptotic bias result to propose the modified GPH estimator, which has smaller bias than the GPH estimator. I derive the asymptotic properties of the proposed

Journal of Business & Economic Statistics
July 2005, Vol. 23, No. 3
DOI 10.1198/073500104000000280

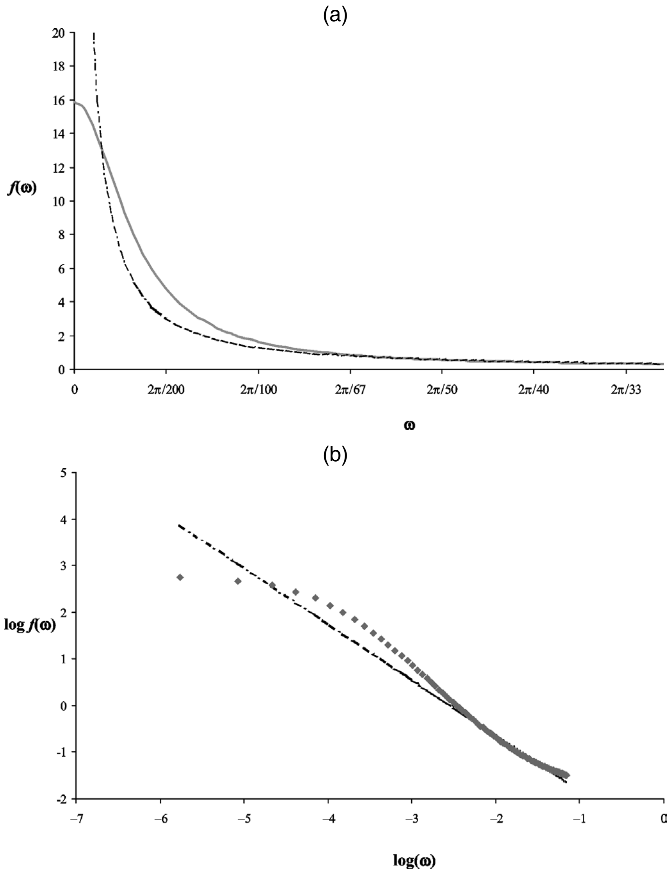

Figure 1. Spectra of Level Shift and Fractionally Integrated Processes. (a) Spectrum (—— level shifts; -·-· fractionally integrated). (b) Log spectrum at Fourier frequencies (◆ level shifts; -·-· fractionally integrated). The data-generating process for level shifts was $y_t = \mu_t + \varepsilon_t$, $\mu_t = (1 - s_t)\mu_{t-1} + s_t\xi_t$, $\varepsilon_t \sim iid(0,1)$, $s_t \sim iid$ Bernoulli(p), $p = .1$, $\xi_t \sim iid(0,1)$. The data-generating process for fractional integration $y_t = (1-L)^{-.6}u_t$, $u_t \sim iid(0, .3)$. (b) Shows a scatterplot of the log spectrum against frequencies $\omega = 2\pi j/T$, for $j = 1, 2, \ldots, T^{1/2}$, and $T = 1{,}000$.

method and illustrate it with applications to the relative price of soybeans to soybean oil and to volatility in the S&P 500. In Section 5 I provide concluding remarks, and in an Appendix I provide proofs of all theorems.

2. A GENERAL MEAN–PLUS–NOISE PROCESS

This section presents a general mean-plus-noise (MN) process with short memory, which can be written as

$$y_t = \mu_t + \varepsilon_t \tag{1}$$

and

$$\mu_t = (1-p)\mu_{t-1} + \sqrt{p}\eta_t, \tag{2}$$

where ε_t and η_t denote short memory random variables with finite nonzero variance. Without loss of generality, I assume that ε_t and η_t have mean 0. I also assume that ε_t and η_s and are independent of each other for all t and s. The parameter p determines the persistence of the level component, μ_t. If p is small, then the level varies slowly. To prevent the variance of μ_t from blowing up as p goes to 0, I scale the innovation in (2) by $\sqrt{p}$. Assuming Gaussianity, the GPH estimator is consistent and asymptotically normal when applied to this process (Hurvich, Deo, and Brodsky 1998). However, I show in Section 3 that the estimator is substantially biased when p is small, even for large T.

The MN process with small p and a fractionally integrated process with $d < 1$ are both highly dependent mean-reverting processes. Thus they compete as potential model specifications for persistent mean-reverting data. However, the application of GPH regression to non-mean-reverting processes with level shifts has received some attention. Granger and Hyung (1999) and Diebold and Inoue (2001) studied the random level-shift process of Chen and Tiao (1990). Diebold and Inoue (2001) also studied the stochastic permanent breaks (STOPBREAK) model of Engle and Smith (1999).

The MN process in (1) and (2) places few restrictions on how the level component evolves. For example, if η_t is Gaussian, then the level evolves continuously and the process is a linear ARMA process with autoregressive and moving average roots that almost cancel out. However, the MN process also encompasses nonlinear models with discrete level shifts. Two prominent examples are stationary random-level shifts (Chen and Tiao 1990) and Markov switching (Hamilton 1989), each of which I discuss in more detail in the following sections.

2.1 Random-Level Shifts

Consider the following random-level shift (RLS) specification for μ_t:

$$\mu_t = (1-s_t)\mu_{t-1} + s_t\xi_t, \tag{3}$$

where $s_t \sim$ iid Bernoulli(p) and ξ_t denotes a short-memory process with mean 0 and variance σ_ξ^2. In each period, the level either equals its previous value or is drawn from some distribution with finite variance. To see that (2) encompasses (3), rewrite (3) as $\mu_t = (1-p)\mu_{t-1} + \sqrt{p}\eta_t$, where $\sqrt{p}\eta_t \equiv s_t\xi_t + (p - s_t)\mu_{t-1}$ and $E(\eta_t^2) = (2-p)\sigma_\xi^2$. Note that the variance of μ_t equals σ_ξ^2, and thus it does not blow up as p goes to 0.

The stationary RLS process in (3) resembles the RLS process of Chen and Tiao (1990). The only difference is that Chen and Tiao specified the level as $\mu_t = \mu_{t-1} + s_t\xi_t$, which is not mean-reverting. As Chen and Tiao (1990, p. 85) stated, this model "provides a convenient framework to assess the performance of a standard time series method on series with level shifts." They studied ARIMA approximations to the RLS process and applied the model to variety store sales. McCulloch and Tsay (1993) used the Gibbs sampler to estimate a RLS model of retail gasoline prices. Such simulation methods are necessary to make inference with this model, because integrating over the 2^T possible sequences of the state process $\{s_t\}$ is infeasible, even for moderate sample sizes.

Chib (1998) and Timmermann (2001) analyzed a hidden Markov process that is observationally equivalent to the RLS process. However, they conditioned on the states, μ_t, and treated them as parameters to be estimated. They did not assume that the timing of the breaks was fixed, unlike in conventional deterministic break-point analysis (e.g., Bai and Perron 1998), which conditions on both the timing (s_t) and the level (ξ_t) of the breaks. When one conditions on s_t or ξ_t, the moments of y_t are time-varying, and it is unclear what the spectrum represents or if it exists. For this reason, I treat s_t and ξ_t as realizations from a stationary stochastic process to obtain a well-defined spectrum.

2.2 Markov Switching

In a Markov-switching process (Hamilton 1989), the level switches between a finite number of discrete values. Since Hamilton's seminal work, Markov-switching models have been applied extensively in economics and finance. For a two-state model, the level equation is

$$\mu_t = (1 - s_t)m_0 + s_t m_1, \tag{4}$$

where m_0 and m_1 denote finite constants and the state variable $s_t \in \{0, 1\}$ evolves according to a Markov chain with the transition matrix

$$\mathbf{P} = \begin{bmatrix} 1 - p_0 & p_1 \\ p_0 & 1 - p_1 \end{bmatrix}.$$

If the parameters p_0 and p_1 are small, then level shifts are rare.

Following Hamilton (1994, p. 684), we can express s_t as an AR(1) process,

$$s_t = p_0 + (1 - p_0 - p_1)s_{t-1} + \sqrt{p_0 + p_1}v_t, \tag{5}$$

where $E(v_t|s_{t-1}) = 0$ and

$$E(v_t^2) = \frac{p_0 p_1(2 - p_0 - p_1)}{(p_0 + p_1)^2} \equiv \sigma_v^2.$$

Combining (4) and (5) yields an AR(1) representation for μ_t,

$$\mu_t = p_0 m_1 + p_1 m_0 + (1 - p_0 - p_1)\mu_{t-1} + (m_1 - m_0)\sqrt{p_0 + p_1}v_t. \tag{6}$$

Without loss of generality, we can set $p_0 m_1 + p_1 m_0 = 0$, and it follows that the MN process in (2) encompasses (6). This illustration can be easily extended to allow for more than two states. Thus the MN specification incorporates models with discrete as well as continuous state space.

3. BIAS IN THE GPH ESTIMATOR

Given the independence of ε_t and η_s for all t and s, the spectrum of the MN process at some frequency ω is

$$\begin{aligned} f(\omega) &= f_\varepsilon(\omega) + f_\mu(\omega) \\ &= f_\varepsilon(\omega) + \frac{p}{p^2 + (1-p)(2 - 2\cos(\omega))} f_\eta(\omega), \end{aligned} \tag{7}$$

where $f_\varepsilon(\omega), f_\mu(\omega)$, and $f_\eta(\omega)$ denote the spectra of ε_t, μ_t, and η_t. The order of integration, d, can be computed from the elasticity of the spectrum at frequencies arbitrarily close to 0, that is,

$$\begin{aligned} d &= -.5 \lim_{\omega \to 0} \frac{\partial \log f}{\partial \log \omega} \\ &= -.5 \lim_{\omega \to 0} \frac{\omega}{f(\omega)} \Biggl(f_\varepsilon'(\omega) - \frac{2p(1-p)\sin(\omega) f_\eta(\omega)}{(p^2 + (1-p)(2 - 2\cos(\omega)))^2} \\ &\qquad + \frac{p f_\eta'(\omega)}{p^2 + (1-p)(2 - 2\cos(\omega))} \Biggr) \\ &= 0. \end{aligned}$$

Thus, as for any short-memory process, the correct value of d equals 0.

The GPH estimate of d equals the least squares coefficient from a regression of the log periodogram on $X_j \equiv -\log(2 - 2\cos(\omega_j)) \approx -\log \omega_j^2$ for $j = 1, 2, \ldots, J$, where $\omega_j = 2\pi j/T$ and $J < T$. For this estimator to be consistent, it must be that $J \to \infty$ as $T \to \infty$. However, because long memory reveals itself in the properties of the spectrum at low frequencies, J must be small relative to T; that is, a necessary condition for consistency is that $J/T \to 0$ as $T \to \infty$. A popular rule of thumb is $J = T^{1/2}$, as recommended originally by Geweke and Porter-Hudak (1983).

The GPH estimate is

$$\hat{d} = d_* + \frac{\sum_{j=1}^{J} (X_j - \overline{X}) \log(\hat{f}_j / f_j)}{\sum_{j=1}^{J} (X_j - \overline{X})^2},$$

where $\hat{f}_j$ denotes the periodogram evaluated at ω_j, f_j denotes the spectrum evaluated at ω_j, and

$$d_* \equiv \frac{\sum_{j=1}^{J} (X_j - \overline{X}) \log f_j}{\sum_{j=1}^{J} (X_j - \overline{X})^2}. \tag{8}$$

For the MN process, the true value of d equals 0, and the bias of the GPH estimator is

$$bias(\hat{d}) = d_* + \frac{\sum_{j=1}^{J} (X_j - \overline{X}) E(\log(\hat{f}_j / f_j))}{\sum_{j=1}^{J} (X_j - \overline{X})^2}. \tag{9}$$

The first term, d_*, represents the bias induced by the short-memory components of the time series. The second term arises because the log periodogram is a biased estimator of the log spectrum.

Hurvich et al. (1998) proved that, for Gaussian long-memory processes, the second term in (9) is $O(\log^3 J/J)$ and thus is negligible. Deo and Hurvich (2001) obtained similar asymptotic results for the GPH estimator in a partially non-Gaussian stochastic volatility model. However, they required that the long-memory component of the model be Gaussian. For fully non-Gaussian processes, such as the RLS process in (3), existing theoretical results require that the periodogram ordinates be pooled across frequencies before the log periodogram regression is run (Velasco 2000). This pooling enables the second term to be proven negligible. Thus for non-Gaussian processes like RLS or Markov switching, current theoretical results do not allow formal treatment of the second component in (9). Nonetheless, the simulations in Section 3.1 suggest that this component is unimportant.

3.1 Illustrating the GPH Bias

To demonstrate the GPH bias, I simulate data from the stationary RLS process for various parameter settings and apply the GPH estimator using the rule-of-thumb value $J = T^{1/2}$. I present the results from GPH estimation in Table 1. The rows labeled "GPH" contain the mean values of $\hat{d}$ over 1,000 Monte Carlo trials, and the rows labeled "Exact d" contain the values d_* computed from the population spectrum as in (8). The sample size, T, ranges from 1,000 to 10,000. This range corresponds to the samples sizes that typically arise in economics and finance with data measured at weekly or daily frequencies.

For $p < .05$, the GPH estimator is substantially biased. In almost all cases, however, the exact value d_* closely corresponds

Table 1. Mean Values of $\hat{d}$ for a RLS Process

p		$\sigma_\xi^2 = 1$, $T = 1{,}000$	$\sigma_\xi^2 = 1$, $T = 5{,}000$	$\sigma_\xi^2 = 1$, $T = 10{,}000$	$\sigma_\xi^2 = 3$, $T = 1{,}000$	$\sigma_\xi^2 = 3$, $T = 5{,}000$	$\sigma_\xi^2 = 3$, $T = 10{,}000$
.25	GPH($\hat{d}$)	.053	.012	.005	.059	.012	.006
	SE	(.137)	(.083)	(.072)	(.136)	(.082)	(.070)
	Exact(d_*)	.050	.010	.005	.055	.011	.006
.05	GPH($\hat{d}$)	.420	.196	.127	.448	.202	.130
	SE	(.156)	(.090)	(.071)	(.159)	(.090)	(.072)
	Exact(d_*)	.434	.198	.124	.463	.204	.128
.01	GPH($\hat{d}$)	.684	.626	.548	.790	.656	.563
	SE	(.152)	(.097)	(.085)	(.152)	(.099)	(.085)
	Exact(d_*)	.717	.635	.550	.814	.663	.565
.005	GPH($\hat{d}$)	.634	.735	.697	.769	.788	.726
	SE	(.205)	(.095)	(.080)	(.199)	(.097)	(.081)
	Exact(d_*)	.695	.747	.702	.830	.796	.730

NOTE: The rows labeled "GPH" give the average of the GPH estimate across 1,000 realizations of size T from the RLS process $y_t = \mu_t + \varepsilon_t$, $\mu_t = (1 - s_t)\mu_{t-1} + s_t\xi_t$, $s_t \sim$ iid Bernoulli(p), $\xi_t \sim$ iid(0, σ_ξ^2), and $\varepsilon_t \sim$ N(0, 1). The GPH statistic is computed with $J = T^{1/2}$. The rows labeled SE give the standard deviation of the GPH estimates across the 1,000 realizations. The asymptotic standard errors for Gaussian processes are .114, .076, and .064 for samples of size 1,000, 5,000, and 10,000 (see Hurvich et al. 1998). The rows labeled Exact(d_*) give the GPH estimate computed using the log spectrum in place of the log periodogram.

to the average GPH estimate. This proximity shows that d_* dominates the GPH bias, and indicates that we can ignore the contribution of the second term in (9). Agiakloglou, Newbold, and Wohar (1993) demonstrated the same phenomenon for stationary AR(1) processes.

The only case in Table 1 where the average GPH estimate deviates from d_* is when $T = 1{,}000$ and p is very close to 0. In this case some of the Monte Carlo realizations contain no level shifts, causing $\hat{d}$ to be close to 0 for those realizations. However, conditional on there being at least one break in a sample, the average GPH estimate is close to d_*. This bisection leads to a bimodal distribution for $\hat{d}$, a feature that Diebold and Inoue (2001) also documented. This bimodal property results from the discontinuity in the spectrum of the process at the point where the probability of a break equals 0.

There are several other notable features in Table 1. First, as T increases, the average GPH estimate approaches 0. This convergence is not surprising, given that the dominant term in the bias, d_*, is $O(J^2/T^2)$ and thus converges to 0 as $T \to \infty$ (see Hurvich et al. 1998, lemma 1). Second, the standard errors monotonically decrease in T in all cases. Third, the average estimate of d increases in σ_ξ^2, the size of the level shifts. This association arises because larger shifts increase the importance of the persistent μ_t term relative to the iid ε_t term. However, the effect of shift size on the GPH bias diminishes as T increases. This diminution is consistent with Theorem 1, given in the next section, which shows that shift size does not matter asymptotically.

3.2 Asymptotic GPH Bias

Hurvich et al. (1998, lemma 1) showed that d_* converges pointwise to 0 as T increases, that is, $d_* \to 0$ as we move from left to right along the rows in Table 1. For large p, this convergence occurs quickly. However, d_* can be far away from 0 when p is small, even for large T. When $p = 0$, the value of d_* is identically 0 for all T. Thus the pointwise limit of d_* provides a satisfactory approximation when $p = 0$ and when p is large, but we need a better approximation when p lies in a local neighborhood of 0.

This problem parallels that of estimating the largest root in an autoregression when that root is near unity. Both cases involve an estimator that exhibits substantial bias in the neighborhood of a point of discontinuity. Influential work by Phillips (1988) and Cavanagh, Elliott, and Stock (1995) showed that by specifying the autoregressive parameter as lying in a local neighborhood of unity, a better large-sample approximation to the distribution of the least squares estimator can be obtained.

Diebold and Inoue (2001) used a similar technique to analyze a Markov-switching process with rare shifts. They specified the switching probability within a local neighborhood of 0 and showed that the variance of partial sums is of the same order of magnitude as the variance of partial sums of a fractionally integrated process. However, this result does not admit a particular order of magnitude for the local neighborhood of 0 that contains p. Thus it implies that a Markov-switching process can be approximated by an integrated process of any nonnegative order, depending the chosen neighborhood. Breidt and Hsu (2002) provided similar results for the RLS process.

In Theorem 1 I show that the appropriate choice of neighborhood size depends critically on J, the number of terms in the GPH regression. This dependence on J emanates from the denominator of the second term in the spectrum (7), which is $p^2 + (1-p)(2 - 2\cos(\omega)) \approx p^2 + \omega^2$ for small ω and p. By setting p to the same order of magnitude as ω, I isolate the dominant component of the spectrum. This isolation produces a good approximation to d_* for small values of p. Specifically, it is appropriate to set $p_T = cJ/T$, where c is a positive constant. Because $\mathrm{var}(T^{-1}\sum_{t=1}^T \mu_t) = \sigma_\eta^2/p_T T$, this condition implies that $\mathrm{var}(T^{-1}\sum_{t=1}^T \mu_t) = O(J^{-1})$. The following theorem formalizes these arguments and obtains the asymptotic bias of the GPH estimator.

Theorem 1. Consider the MN process in (1) and (2) and suppose that $f_\varepsilon(\omega) < \widetilde{B}_0 < \infty, f_\eta(\omega) < \widetilde{B}_0 < \infty$, and $|f'_\eta(\omega)| <$

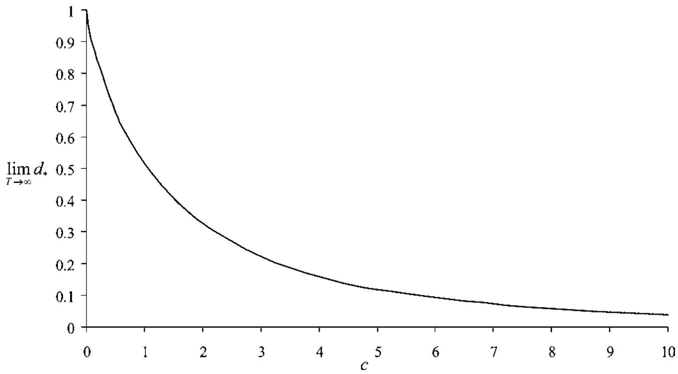

Figure 2. Asymptotic Bias of the GPH Estimator. This curve applies to data generated by $y_t = \mu_t + \varepsilon_t$, $\mu_t = (1 - p_T)\mu_{t-1} + \sqrt{p_T}\eta_t$, $\eta_t \sim$ short memory, $\varepsilon_t \sim$ short memory, and $t = 1, 2, \ldots, T$. The bandwidth (number of frequencies) in the GPH regression is $J = p_T T/c$.

$\widetilde{B}_1 < \infty$ for all ω in a neighborhood of 0. Assume that $JT^{-1}\log(J) \to 0$ and let $p_T = cJ/T$, where $c > 0$. Then

$$\lim_{J,T\to\infty} d_* = 1 - .25\Phi\big(-(2\pi/c)^2, 2, .5\big),$$

where $\Phi(x, s, a)$ denotes the Lerch transcendent function.

Theorem 1 provides an approximation to the bias in the GPH estimator when a general MN process generates the data. This asymptotic bias is a function of the Lerch transcendent function, also known as the Lerch zeta function, which is a generalization of the Riemann zeta function (Gradshteyn and Ryzhik 1980, p. 1072). The only parameter in the asymptotic bias is $c = p_T T/J$. This result indicates that the reduction in bias from decreasing J asymptotically equals the decrease in bias from a proportionate increase in p or T. Figure 2 plots the asymptotic bias and shows that it ranges between 0 and 1 and decreases monotonically in c. This graph emphasizes the fact that the GPH bias depends both on the properties of the data and on the specification of the estimator. One cannot say that a level-shift process appears to be fractionally integrated of some order d without reference to the bandwidth J.

Theorem 1 implies that d_* does not converge to 0 uniformly in $p \in [0, 1]$; that is, $\sup_{p_0,p_1\in[0,1]} d_*$ does not converge to 0 as $T \to \infty$. In other words, for every T there exists a value of p such that $d_* > 0$, despite the fact that d_* converges pointwise to 0. Thus, traveling from left to right along the rows of Table 1, d_* decreases toward 0, but there exists a path through the table in a southeast direction for which d_* does not converge to 0.

4. MODIFIED GPH REGRESSION

As shown in Section 3, the GPH statistic may erroneously indicate long memory when the MN process generates the data. In this section I use the result in Theorem 1 to suggest a simple modification to the GPH estimator that reduces bias when the data-generating process contains level shifts. An important advantage of the modified GPH estimator is its simplicity. It can be implemented easily by adding an extra regressor to the GPH regression. This straightforwardness makes it a useful diagnostic tool for signalling whether a fully specified model with level shifts could outperform a long-memory model.

The asymptotic bias in Theorem 1 derives from the fact that when p is small, the dominant component of the spectrum at low frequencies is $-\log(p^2 + \omega^2)$ plus a constant. This dominant term is nonlinear in $\log(\omega)$, so adding $-\log(p^2 + \omega^2)$ as an extra regressor in the GPH regression would reduce the bias caused by level shifts. However, this strategy is infeasible, because p is unknown. I create a feasible estimator by setting $p_T = kJ/T$ for some constant $k > 0$ and running the regression

$$\log\hat{f}_j = \alpha + dX_j + \beta Z_{kj} + \hat{u}_j,$$

where

$$Z_{kj} = -\log\left(\frac{(kJ)^2}{T^2} + \omega_j^2\right)$$

and $X_j = -\log(2 - 2\cos(\omega_j))$ as before.

The modified GPH estimator is

$$\hat{d}^k = d_*^k + (\widetilde{X}'M_Z\widetilde{X})^{-1}\widetilde{X}'M_Z\log(\hat{f}/f),$$

where $\widetilde{X} \equiv X - \overline{X}$, $M_Z = I - \widetilde{Z}_k(\widetilde{Z}_k'\widetilde{Z}_k)^{-1}\widetilde{Z}_k'$, $\widetilde{Z}_k \equiv Z_k - \overline{Z}_k$, $\overline{X} = J^{-1}\sum_{j=1}^J X_j$, $\overline{Z}_k = J^{-1}\sum_{j=1}^J Z_{kj}$, and d_*^k denotes the estimator computed from the spectrum rather than from the periodogram. Next I derive the asymptotic properties of the modified GPH estimator.

4.1 Asymptotic Properties of the Modified GPH Estimator

The MN process includes Gaussian processes as a special case. However many useful models, including RLS or Markov switching, are non-Gaussian. As discussed in Section 3, current theoretical results do not allow formal treatment of the second component of the bias for log periodogram regression with non-Gaussian data. Thus, as in Theorem 1, I focus on the dominant component of the bias, d_*^k. I derive an approximation to d_*^k for the potentially non-Gaussian MN process.

Theorem 2. Consider the MN process in (1) and (2) and suppose that $f_\varepsilon(\omega) < \widetilde{B}_0 < \infty, f_\eta(\omega) < \widetilde{B}_0 < \infty$, and $|f_\eta'(\omega)| < \widetilde{B}_1 < \infty$ for all ω in a neighborhood of 0. Assume that $JT^{-1}\log(J) \to 0$ and let $p_T = cJ/T$, where $c > 0$. Then

$$\lim_{J,T\to\infty} d_*^k = \frac{1}{v_k}\big(1 - .25\Phi\big(-(2\pi/c)^2, 2, .5\big) - r_k h_k\big),$$

where

$$\begin{aligned} h_k \equiv 1 &- \frac{ck}{4\pi^2}\tan^{-1}\left(\frac{2\pi}{c}\right)\tan^{-1}\left(\frac{2\pi}{k}\right) \\ &- \frac{k+c}{4\pi}\left(\mathrm{Im}\left(\mathrm{Li}_2\left(\frac{k+2\pi i}{c+k}\right)\right)\right. \\ &\left.+ \frac{1}{2}\tan^{-1}\left(\frac{2\pi}{k}\right)\log\left(\frac{c^2+4\pi^2}{(k+c)^2}\right)\right) \\ &+ \frac{|k-c|}{4\pi}\left(\mathrm{Im}\left(\mathrm{Li}_2\left(\frac{\min(k,c)+2\pi i}{-|k-c|}\right)\right)\right. \\ &\left.+ \frac{1}{2}\tan^{-1}\left(\frac{2\pi}{\min(k,c)}\right)\log\left(\frac{\max(k^2,c^2)+4\pi^2}{(k+c)^2}\right)\right), \end{aligned}$$

$$r_k \equiv \big(1 - .25\Phi\big(-(2\pi/k)^2, 2, .5\big)\big) \times \left(1 - \left(\frac{k}{2\pi}\tan^{-1}\left(\frac{2\pi}{k}\right)\right)^2 - \frac{k}{2\pi}\left(\mathrm{Im}\left(\mathrm{Li}_2\left(\frac{1}{2} + \frac{\pi i}{k}\right)\right) + \frac{1}{2}\tan^{-1}\left(\frac{2\pi}{k}\right)\log\left(\frac{1}{4} + \frac{\pi^2}{k^2}\right)\right)\right)^{-1},$$

$$v_k \equiv 1 - r_k\big(1 - .25\Phi\big(-(2\pi/k)^2, 2, .5\big)\big),$$

$\Phi(x, s, a)$ is the Lerch transcendent function, $\mathrm{Im}(x + iy) \equiv y$, and Li_2 is the dilogarithm.

The asymptotic bias of the modified GPH estimator is a function of c and k. Figure 3 plots this asymptotic bias for various k, along with the asymptotic bias of the GPH estimator for comparison. The asymptotic bias equals 0 when $k = c$, so for every c there exists a value of k that completely eliminates bias. For $k > c$, the bias is positive, and for $k < c$, the bias is negative. There are some other notable features of the asymptotic bias. First, the absolute bias of the modified GPH estimator is less than the GPH bias for all k. Second, as level shifts become more frequent (i.e., as $c \to \infty$), the asymptotic bias goes to 0 for all k. Third, the asymptotic bias increases in k, and it converges to the GPH bias in Theorem 1 as $k \to \infty$.

The curves in Figure 3 indicate that the modified GPH estimator can markedly reduce the bias in the GPH estimator due to occasional level shifts. However, such bias reduction becomes only useful if the requisite loss in precision is acceptable. To address this issue, I derive the asymptotic properties of the modified GPH estimator under the alternative model of Gaussian long memory.

Theorem 3. Consider the fractionally integrated process $y_t = (1 - L)^{-d}u_t$, where $\{u_t\}$ is a stationary short-memory process and $d \in (-.5, .5)$. Suppose that $f_u'(0) = 0$, $|f_u''(\omega)| < \widetilde{B}_2 < \infty$, and $|f_u'''(\omega)| < \widetilde{B}_3 < \infty$ for all ω in a neighborhood of 0. Assume that y_t is Gaussian and that $JT^{-1}\log(J) \to 0$. Then

$$E(\hat{d}^k - d) = b_k \frac{-2\pi^2}{9}\frac{f_u''(0)}{f_u(0)}\frac{J^2}{T^2} + o(J^2/T^2) + O(\log^3 J/J)$$

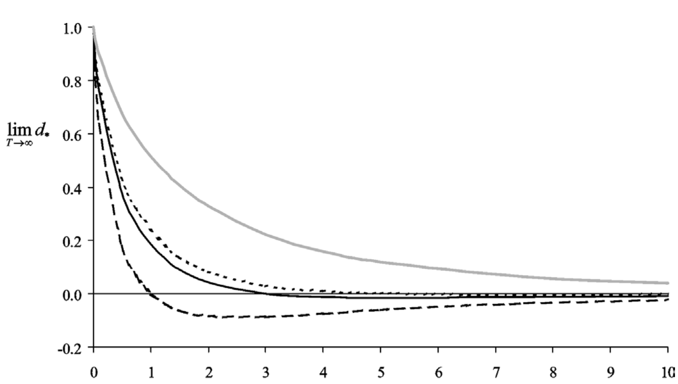

Figure 3. Asymptotic Bias of the Modified GPH Estimator. These curves apply to data generated by $y_t = \mu_t + \varepsilon_t$, $\mu_t = (1 - p_T)\mu_{t-1} + \sqrt{p_T}\eta_t$, $\eta_t \sim$ short memory, $\varepsilon_t \sim$ short memory, and $t = 1, 2, \ldots, T$. The bandwidth (number of frequencies) in the GPH regression is $J = p_T T/c$. (– – $k = 1$; —— $k = 3$; · · · · $k = 5$; —— GPH.)

and

$$\mathrm{var}(\hat{d}^k) = \frac{\pi^2}{24Jv_k} + o(J^{-1}),$$

where $b_k = \frac{1}{v_k}(1 - r_k((\frac{3k^2}{8\pi^2} + 1) - \frac{3k}{4\pi}(\frac{k^2}{4\pi^2} + 1)\tan^{-1}(\frac{2\pi}{k})))$, and r_k and v_k are as defined in Theorem 2.

Corollary 1. Consider the process in Theorem 3 and assume also that $J = o(T^{4/5})$ and $\log^2 T = o(J)$. Then $J^{1/2}(\hat{d}^k - d) \xrightarrow{d} \mathrm{N}(0, \pi^2/24v_k)$.

Except for the scale factors b_k and v_k, the asymptotic bias and variance expressions in Theorem 3 are the same as those of Hurvich et al. (1998) for the GPH estimator. The bias factor b_k takes the values $-.65$, $-.41$, $-.28$, $-.21$, and $-.15$ for $k = 1, 2, 3, 4$, and 5. Thus the bias of the modified GPH estimator is much smaller than the GPH bias. This bias reduction arises because the extra term in the modified GPH regression picks up some of the curvature in the log spectrum that causes bias in the GPH estimator. The variance factor v_k takes the values .17, .26, .31, .35, and .37 for $k = 1, 2, 3, 4$, and 5. Thus, for a given J, the variance of the modified GPH estimator is larger than the GPH variance. However, a suitable choice of J mitigates this efficiency loss.

I simulate the performance of the modified GPH estimator in two settings, the RLS process and a fractionally integrated process. I illustrate the performance of the estimator across different values of J for one set of parameter values. In Section 4.2 I give results for a range of parameter values, sample sizes, and methods for choosing J.

Figure 4 shows the performance of the modified GPH estimator as a function of J when applied to a RLS process with $T = 5{,}000$ and $p = .02$. Figures 4(a) and 4(b) shows that the asymptotic bias from Theorem 2 closely corresponds to the actual bias. For all values of k, this bias is markedly lower than for the GPH estimator. Figure 4(e) also reveals the lower bias of the modified GPH estimator. It shows that a standard t-test based on the modified GPH estimate erroneously rejects the null hypothesis that $d = 0$ less often than the GPH estimator. The size of this test increases in k because the estimator bias increases in k. Figure 4(c) shows that the values of J that minimize the root mean squared error (RMSE) of the modified GPH estimator exceed the values of J that minimize the RMSE of the GPH estimator. The minimum RMSE values are similar across the values of k and are less than those of the GPH estimator.

Figures 4(d), 4(c), and 4(f) illustrate that the variance and asymptotic normality results in Theorem 3 and its corollary also apply to the RLS process. Figure 4(d) shows that the estimated standard error of $\hat{d}^k$ closely corresponds to the actual standard error. I measure the actual standard error as the standard deviation of $\hat{d}^k$ across the Monte Carlo draws. The estimated standard error is computed as $(\pi/\sqrt{6})(\widetilde{X}'M_Z\widetilde{X})^{-1/2}$, which has a limiting value of $\pi/\sqrt{24Jv_k}$, from Theorem 3. As J increases toward 200, the estimated standard error becomes slightly biased downward, and the ratio of the estimated standard error to actual standard error decreases toward .85. However, for $J < 60$, the ratio exceeds .95.

Figures 4(e) and 4(f) show that a standard t-test rejects the null hypothesis that $d = 0$ with a similar frequency to a hypothetical test that assumes normality. The rejection frequency of

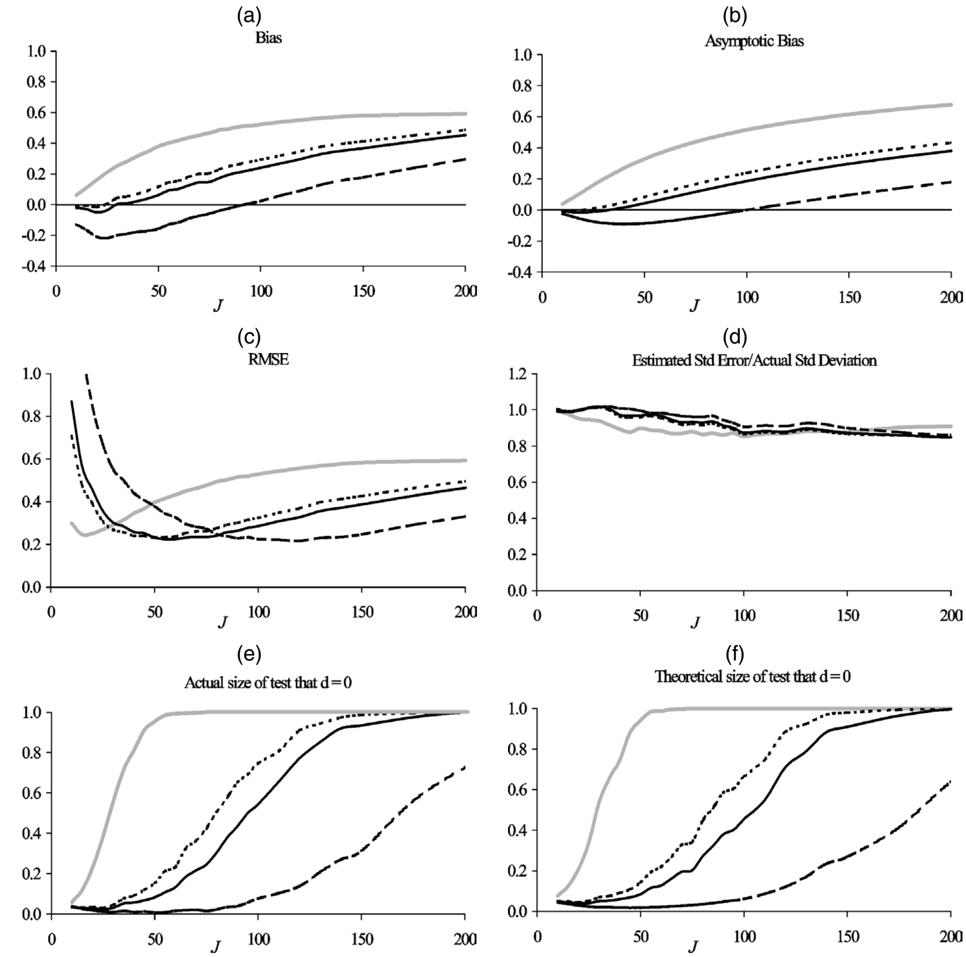

Figure 4. Performance of the Modified GPH Estimator for an RLS Process (— GPH; – – k = 1; — k = 3; · · · · k = 5). The data-generating process was $y_t = \mu_t + \varepsilon_t$, $\mu_t = (1 - s_t)\mu_{t-1} + s_t\xi_t$, $s_t \sim$ iid Bernoulli(p), $\xi_t \sim$ iid N(0, 1), $\varepsilon_t \sim$ iid N(0, 1), p = .02 and T = 5,000. The curves are generated from 1,000 Monte Carlo draws as follows: (a) average estimate of d across draws; (b) from Theorem 2; (c) computed from average and variance of estimates of d across draws; (d) estimated standard error equals $(\pi/\sqrt{6})(\widetilde{X}' M_Z \widetilde{X})^{-1/2}$, actual standard deviation equals standard deviation of estimate of d across draws; (e) proportion of rejections of null hypothesis that d = 0 against d > 0, nominal size = 5%; (f) size computed assuming that estimate of d is normally distributed with mean and variance given by average and variance of estimates of d across draws; nominal size = 5%.

this hypothetical test equals the probability that the estimate exceeds the one-sided 5% critical value, assuming that the estimator is normally distributed. The normal approximation appears to be adequate, especially for small values of J.

To assess the efficiency loss from the modified GPH estimator, I simulate from a fractionally integrated process with $d = .3$, where the innovations follow an AR(1) process with autoregressive parameter .4. Figure 5 presents the results for various J. Excluding the long-memory component, this process exhibits less dependence than the RLS process in Figure 4, so the RMSE in Figure 5 is minimized for greater values of J than in Figure 4. Figures 5(d), 5(e), and 5(f) corroborate the theoretical variance and asymptotic normality results in Theorem 3 and its corollary.

Figures 5(a) and 5(b) show that the bias of the GPH estimator exceeds the modified GPH bias for all k, as predicted by Theorem 3. However, for large J, the asymptotic bias overestimates the actual bias, because the second-order terms in the bias expression become nonnegligible. Furthermore, the asymptotic bias overestimates the actual bias by more for the modified GPH estimator than for the GPH estimator, which implies that the asymptotic results overstate the finite-sample RMSE efficiency loss of the modified GPH estimator. I study this point further in Section 4.2.

4.2 Choosing J and k

From Theorem 3, the MSE of $\hat{d}^k$ for Gaussian long memory is

$$\mathrm{MSE}(\hat{d}^k) = b_k^2 \frac{4\pi^4}{81}\left(\frac{f_u''(0)}{f_u(0)}\right)^2 \frac{J^4}{T^4} + \frac{\pi^2}{24 J v_k} + o(J^4 T^{-4}) + O(J T^{-2} \log^3 J) + o(J^{-1}).$$

The value of J that minimizes MSE is

$$J = (v_k b_k^2)^{-1/5} \left(\frac{27}{128\pi^4}\right)^{1/5} \left(\frac{f_u''(0)}{f_u(0)}\right)^{-2/5} T^{4/5}, \qquad (10)$$

ignoring the remainder terms and assuming that $f_u''(0) \neq 0$. The MSE-optimal value of J in (10) equals that for the GPH estimator (see Hurvich et al. 1998), except for the scale factor $(v_k b_k^2)^{-1/5}$. This scale factor takes the values 1.69, 1.88, 2.09, 2.33, and 2.58 for $k = 1, 2, 3, 4$, and 5. Thus for exam-

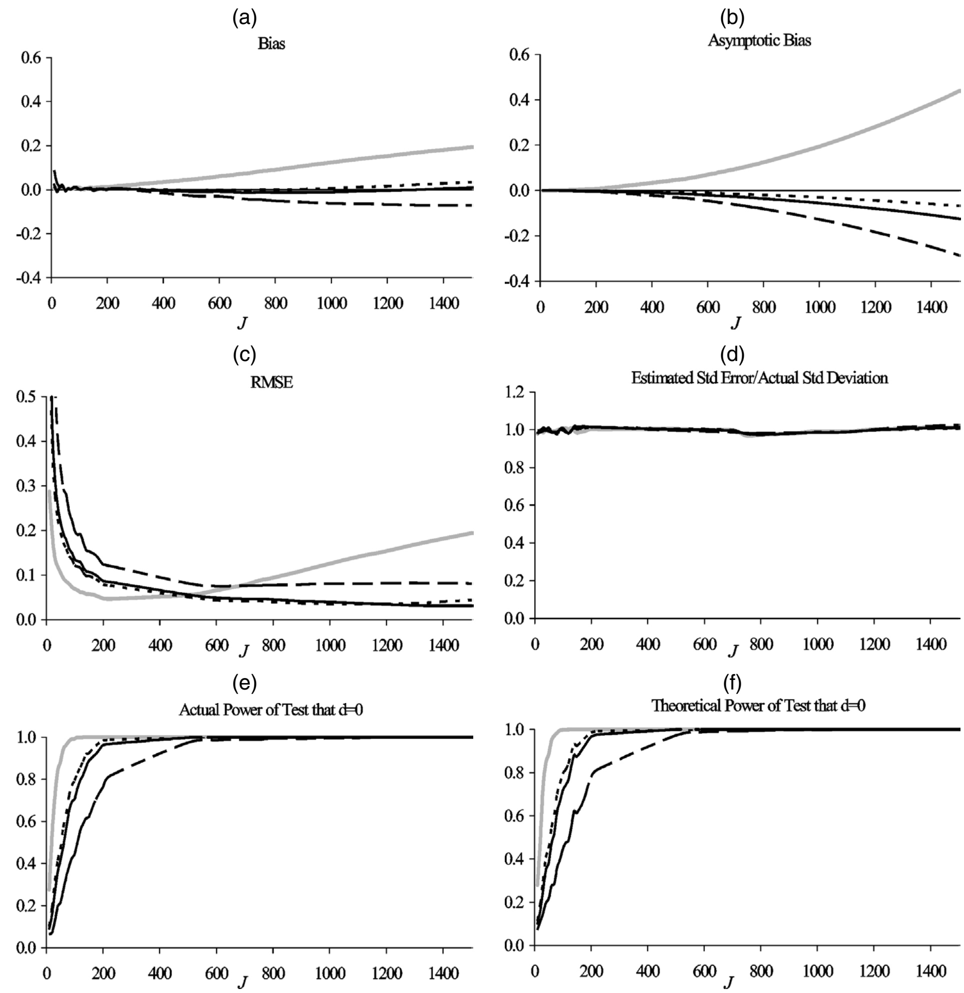

Figure 5. Performance of the Modified GPH Estimator for a FI Process (—— GPH; – – k = 1; —— k = 3; · · · · k = 5). The data-generating process was $y_t = (1 - L)^{-d} u_t$, where $d = .3$, $u_t = .4u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim$ iid N(0, 1), and T = 5,000. The curves are generated from 1,000 Monte Carlo draws as follows: (a) average estimate of d across draws; (b) from Theorem 3; (c) computed from average and variance of estimates of d across draws; (d) estimated standard error equals $(\pi/\sqrt{6})(\widetilde{X}' M_Z \widetilde{X})^{-1/2}$; actual standard deviation equals standard deviation of estimate of d across draws; (e) proportion of rejections of null hypothesis that d = 0 against d > 0, nominal size = 5%; (f) power computed assuming that estimate of d is normally distributed with mean and variance given by average and variance of estimates of d across draws, nominal size = 5%.

ple, if $k = 3$, then the MSE-optimal choice of J is approximately double the MSE-optimal choice for the GPH estimator. Hurvich and Deo (1999) proposed a consistent estimator for the ratio $f_u''(0)/f_u(0)$ that enables plug-in selection of the MSE-optimal J. Their estimator is the coefficient on $.5\omega_j^2$ in a regression of the log periodogram on X_j and $.5\omega_j^2$.

If J equals its MSE-optimal value, then, for Gaussian long memory,

$$\mathrm{MSE}(\hat{d}^k) = \left(\frac{|b_k|}{v_k^2}\right)^{2/5} \mathrm{MSE}(\hat{d}), \qquad (11)$$

excluding the remainder terms. Thus the RMSE of the modified GPH estimator is $(|b_k|/v_k^2)^{1/5}$ times the RMSE of the GPH estimator when J is chosen to be MSE-optimal. This scale factor takes the values 1.86, 1.44, 1.24, 1.11, and 1.02 for $k = 1, 2, 3, 4$, and 5. Thus, there is negligible asymptotic efficiency loss when $k = 5$.

The asymptotic MSE in (11) does not apply to the MN process for small p. In this case, as shown in Theorem 2, the bias is $O(1)$ and dominates the variance for all J. The modified GPH estimator often has smaller RMSE than the GPH estimator in this case, because the extra term in the log periodogram regression mitigates bias. Suppose that we choose J as in (10), implying that J increases in k. Because the local-to-0 parameter c equals pT/J, a larger value of J implies a smaller value of c. Figure 6 presents the asymptotic bias of the modified GPH estimator from Theorem 2, assuming that J is chosen as in (10). Recall that Figure 3 shows the asymptotic bias for the case when J is fixed across values of k. Figure 6 is the same as Figure 3, but with the curves stretched horizontally to reflect decreasing values of c as k increases. Figure 6 reveals negligible bias reduction for $k = 5$, but substantial bias reduction when $k < 5$.

Given a value of J, choosing $k = c$ implies that the asymptotic bias of the modified GPH estimator equals 0. However, c cannot be efficiently estimated, because it defines a shrinking neighborhood around 0, and thus larger samples bring little information about it. If one were ignorant about the value of c, then k could be chosen to minimize average bias over all possible values of c. To this end, I numerically integrate under the absolute value of the asymptotic bias curves in Theorem 2, and

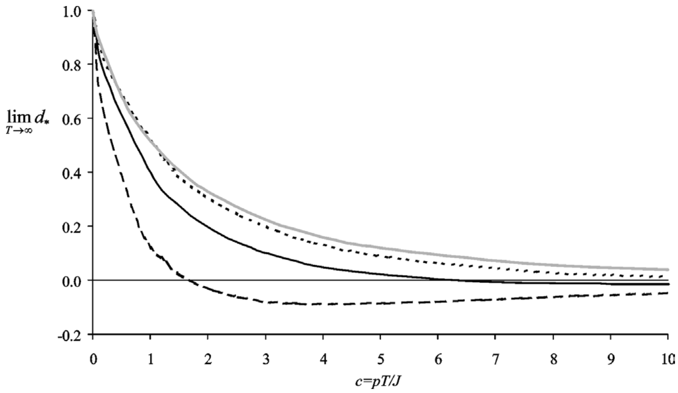

Figure 6. Asymptotic Bias of Modified GPH Estimator for MSE Optimal J (– – k = 1; —— k = 3; · · · · k = 5; —— GPH). These curves apply to data generated by $y_t = \mu_t + \varepsilon_t$, $\mu_t = (1 - p_T)\mu_{t-1} + \sqrt{p_T}\eta_t$, $\eta_t \sim$ short memory, $\varepsilon_t \sim$ short memory, and t = 1, 2, ..., T. The bandwidth (number of frequencies) in the GPH regression is $J = p_T T/c$.

find that average bias is minimized at $k = 3.16$; it is decreasing in k for $0 < k < 3.16$ and increasing in k for $3.16 < k < \infty$. Because the asymptotic bias is almost identical for $k = 3$ and $k = 3.16$, I recommend rounding to the nearest integer and setting $k = 3$.

For a given choice of J, choosing $k = 3$ minimizes average bias if the process contains rare level shifts. If we choose MSE-optimal J, then $k = 3$ implies a 24% higher asymptotic RMSE than for the GPH estimator if the true process contains long memory [see (11)]. However, in simulations that follow I show that the efficiency loss can be much less than 24% in finite samples. In fact, modified GPH regression has a lower RMSE than GPH regression in some long-memory cases.

I simulate the performance of the modified GPH estimator for various parameter settings. I use both the rule-of-thumb value of $J = T^{1/2}$ and the plug-in method of Hurvich and Deo (1999) to select J. Results for the RLS process are presented in Table 2, and results for a fractionally integrated process are contained in Table 3. For plug-in selection of J, the results for the modified GPH estimator with $k = 5$ closely match those for the GPH estimator. This correspondence is consistent with the asymptotic bias curves in Figure 6 and the similarity between the asymptotic RMSEs of each estimator. The modified GPH estimator with $k = 1$ can have substantially negative bias, which leads to high RMSE values in many cases.

For the RLS process, setting $k = 3$ results in the lowest RMSE when $p > .02$ and J is chosen using the plug-in method. For example, if $T = 10{,}000$ and $p = .05$, then the RMSE when $k = 3$ improves by 35% over the GPH estimator. The RMSE improves by 22% over GPH when $p = .02$ and by 20% when $p = .1$ for this same sample size. Size distortion also reduces substantially relative to the GPH estimator in these cases.

In the plug-in method, J increases with k according to the relationship in (10). This feature results in reduced RMSE for the modified GPH estimator over the GPH estimator in many cases, reinforcing the results in Figure 4. The only cases where RMSE for plug-in selection of J exceeds that for rule-of-thumb selection occur for RLS when p is small, which corroborates the findings of Hurvich and Deo (1999), who stated that the plug-in method works well unless the spectrum is too peaked near zero frequency.

Table 2. Properties of the Modified GPH Estimator for an RLS Process

		Plug-in selection of J				$J = T^{1/2}$			
T	p	GPH	k = 1	k = 3	k = 5	GPH	k = 1	k = 3	k = 5
Bias									
1,000	.01	.49	.66	.56	.51	.72	.69	.74	.74
	.02	.52	.48	.54	.55	.64	.29	.49	.53
	.05	.34	−.02	.24	.35	.42	−.13	.12	.18
	.10	.17	−.16	.05	.16	.22	−.19	−.02	.02
10,000	.01	.59	.47	.58	.61	.55	.01	.24	.30
	.02	.40	.05	.30	.40	.36	−.12	.06	.11
	.05	.15	−.10	.05	.12	.12	−.09	−.02	.00
	.10	.06	−.07	−.01	.03	.04	−.04	−.01	−.01
RMSE									
1,000	.01	.51	.74	.58	.54	.71	.91	.81	.80
	.02	.54	.66	.58	.57	.66	.63	.60	.61
	.05	.40	.50	.38	.41	.45	.53	.34	.33
	.10	.28	.52	.31	.29	.26	.56	.31	.27
10,000	.01	.60	.52	.59	.62	.56	.22	.29	.33
	.02	.41	.18	.32	.41	.36	.24	.16	.17
	.05	.17	.17	.11	.15	.14	.22	.14	.12
	.10	.10	.14	.08	.09	.08	.20	.14	.12
Rejection frequency									
1,000	.01	.98	.83	.94	.98	1.00	.42	.76	.84
	.02	.96	.61	.86	.95	.99	.17	.50	.62
	.05	.67	.22	.44	.66	.90	.27	.10	.17
	.10	.41	.14	.27	.39	.51	.01	.03	.04
10,000	.01	1.00	.88	.99	.99	1.00	.06	.56	.76
	.02	.99	.20	.89	.99	1.00	.01	.12	.23
	.05	.67	.00	.20	.55	.58	.01	.03	.04
	.10	.34	.01	.07	.19	.14	.02	.04	.04

NOTE: The data-generating process was $y_t = \mu_t + \varepsilon_t$, $\mu_t = (1 - s_t)\mu_{t-1} + s_t\xi_t$, $s_t \sim$ iid Bernoulli(p), $\xi_t \sim$ iid N(0, 1), $\varepsilon_t \sim$ iid N(0, 1). The elements in the table are averages across 1,000 Monte Carlo realizations. The plug-in method was used with $L = .1T^{6/7}$ frequencies in first-stage regression.

Table 3. Properties of the Modified GPH Estimator for a Fractionally Integrated Process

		Plug-in selection of J				$J = T^{1/2}$			
T	p	GPH	$k = 1$	$k = 3$	$k = 5$	GPH	$k = 1$	$k = 3$	$k = 5$
Bias									
1,000	.0	−.03	−.10	−.05	−.02	.01	.01	.01	.01
	.4	.01	−.14	−.05	.00	.02	.00	.00	.00
	.8	.14	−.19	.01	.11	.10	−.09	−.02	.00
10,000	.0	−.01	−.02	−.01	−.01	.00	.00	.00	.00
	.4	.01	−.04	−.02	.00	.01	.01	.01	.01
	.8	.05	−.08	−.02	.03	.01	.00	.00	.00
RMSE									
1,000	.0	.13	.34	.19	.13	.14	.52	.30	.27
	.4	.14	.32	.18	.14	.14	.50	.30	.27
	.8	.26	.40	.26	.26	.17	.53	.31	.27
10,000	.0	.04	.08	.05	.04	.07	.21	.14	.13
	.4	.04	.09	.05	.04	.07	.20	.14	.12
	.8	.11	.11	.08	.10	.07	.21	.14	.13
Rejection frequency									
1,000	.0	.79	.48	.69	.80	.75	.14	.27	.33
	.4	.79	.37	.69	.80	.75	.13	.27	.31
	.8	.78	.27	.58	.73	.89	.10	.23	.29
10,000	.0	.99	.93	.99	1.00	.99	.46	.74	.81
	.4	1.00	.93	.99	1.00	.99	.47	.75	.82
	.8	1.00	.79	.97	.99	.99	.48	.71	.80

NOTE: The data-generating process was $y_t = (1 - L)^{-d} u_t$, where $d = .3$, $u_t = pu_{t-1} + \varepsilon_t$, and $\varepsilon_t \sim$ iid N(0,1). The elements in the table are averages across 1,000 Monte Carlo realizations. The plug-in method was used with $L = .1T^{6/7}$ frequencies in first-stage regression.

For the long-memory process with $p = .8$, the modified GPH estimator with $k = 3$ outperforms the GPH estimator. It has a smaller bias and RMSE. Thus the modified GPH estimator has the power to correct bias caused by pure autoregressive processes. When the short-memory component is less persistent ($p = 0$ and $p = .4$), the RMSE of the modified GPH estimator slightly exceeds that for the GPH estimator. In summary, the modified GPH estimator with $k = 3$ and J selected using the plug-in method performs well in most settings.

4.3 Applications

To illustrate the modified GPH estimator, I apply it to the weekly relative price of soybeans to soybean oil and to daily volatility in the S&P 500. The soybean price data span January 1, 1953 to June 30, 2001 and contain the average weekly soybean price in central Illinois and the average weekly soybean oil price in Decatur, Illinois. There are a total of 2,455 observations. Given that soybean oil derives from soybeans, the prices of these two commodities should have a common trend, which implies that the ratio of their prices should be mean-reverting. Figure 7(a) plots the log relative price series and indicates that it is mean-reverting with strong positive dependence. This structure suggests that potential candidate models for the relative price include long memory and short memory with level shifts.

Table 4 presents the estimated values of d from the GPH and modified GPH estimators. The estimated value of d for

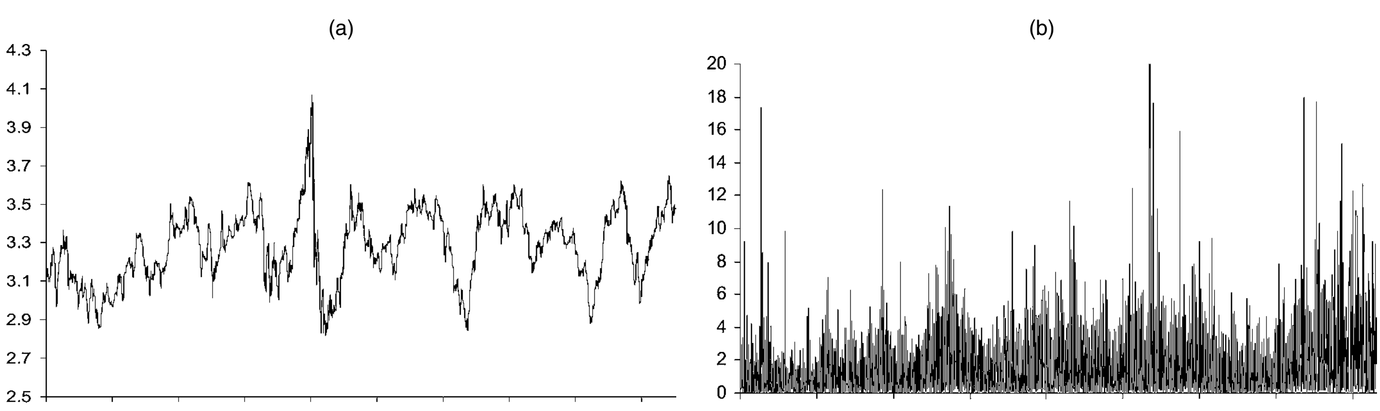

Figure 7. Soybean and S&P 500 Time Series. (a) Log relative price of soybeans and soybean oil; (b) absolute daily returns on the S&P 500. The soybean data span January 1, 1953 to June 30, 2001 and contain the average weekly soybean price in central Illinois and the average weekly soybean oil price in Decatur, Illinois. There are a total of 2,455 observations, and the units of measurement are cents per bushel for soybeans and cents per pound for soybean oil. The S&P 500 data span January 1, 1961 to July 31, 2002 and comprise the absolute daily returns on the S&P 500 stock index. There are a total of 10,463 observations.

Table 4. Estimates of the Long-Memory Parameter

	GPH	Modified GPH $k = 1$	Modified GPH $k = 3$	Modified GPH $k = 5$
Relative price of soybeans to soybean oil				
Plug-in	.79∗	−.13	.24	.34∗
	(.09)	(.30)	(.20)	(.19)
$J = T^{1/2} = 49$	.84∗	.03	.58∗	.76∗
	(.10)	(.31)	(.17)	(.14)
Absolute daily returns on S&P 500				
Plug-in	.33∗	.50∗	.48∗	.48∗
	(.03)	(.12)	(.11)	(.11)
$J = T^{1/2} = 102$	.38∗	.45∗	.42∗	.44∗
	(.07)	(.18)	(.11)	(.09)

NOTE: The cells contain estimates of d, the long-memory parameter, with standard errors below each estimate in parentheses. ∗ indicates significance at 5%, using standard normal critical values and a one-sided alternative. The soybean data span January 1, 1953 to June 30, 2001 and contain the average weekly soybean price in central Illinois and the average weekly soybean oil price in Decatur, Illinois. There are a total of 2,455 observations. The stock market data span January 1, 1961 to July 31, 2002 and contain the absolute daily returns on the S&P 500 stock index. There are a total of 10,463 observations. The plug-in method was used with $L = .1T^{6/7}$ frequencies in first-stage regression. The estimated plug-in value of J for the GPH estimator are 67 for soybeans and 657 for the S&P 500. The plug-in values of J for the modified GPH estimator equal the scale factors in (10) multiplied by 67 (for soybeans) and 657 (for the S&P 500).

the soybean data equals .79 when the plug-in method is used to select J and .84 when $J = T^{1/2}$. These values are significantly different from 0. The modified GPH estimates are substantially smaller than the GPH estimates for both methods of bandwidth selection. When the plug-in selection method is used with $k = 3$, the estimate of d equals .24 and is insignificantly different from zero. When $k = 1$, the estimated value of d is also insignificant. Thus a short-memory model with level shifts is a viable alternative to long memory for these data.

Liu (2000), Granger and Hyung (1999), Lobato and Savin (1998), and others have cited financial market volatility as one setting in which long memory and level shifts provide competing model specifications. I apply the modified GPH estimator to absolute daily returns on the S&P 500. The data are plotted in Figure 7(b). The sample period is January 1, 1961 to July 31, 2002, and returns are measured as the log price change. Table 4 indicates that a short-memory model with level shifts is not a viable alternative to long memory for this series. In fact, the modified GPH estimates exceed the GPH estimates for all values of J and k. The GPH estimates are .33 and .38 for the two methods of choosing J, whereas the modified GPH estimates range from .42 to .50. This result is not sensitive to the measure of volatility; using squared returns and the log of absolute returns leads to same conclusion. Thus long memory in volatility of S&P 500 returns appears to not be illusory.

5. CONCLUSION

This article has addressed the illusion of fractional integration, or long memory, in time series containing level shifts. I have focused on the GPH estimator, which is used liberally in empirical work. When applied to a short-memory MN process, the GPH estimator is biased and often erroneously indicates the presence of long memory. I have derived a large-sample approximation to this bias and used it to formulate a new estimator that has markedly smaller bias. I illustrated the modified GPH estimator with applications to the relative price of soybeans to soybean oil and to stock market volatility.

The modified GPH estimator requires choosing a value for a nuisance parameter k. This parameter proxies for a local-to-0 parameter that cannot be well estimated from the data. I recommend setting $k = 3$, which minimizes average absolute bias across all possible values of the true parameter c. For a given bandwidth J, this recommendation leads to positive bias in the modified GPH estimator if $c < 3$ and negative bias if $c > 3$. Despite this trade-off, the modified GPH estimator with $k = 3$ exhibits less absolute bias than the original GPH estimator for all values of c (see Figs. 3 and 6). Thus, although it does not completely eliminate bias due to level shifts, the modified GPH estimator with $k = 3$ significantly reduces bias relative to the GPH estimator.

The modified GPH estimator suggests whether a short-memory model with level shifts should be considered as an alternative to long memory. It is based on the spectrum, which represents the linear dependence properties of a time series. However, a process with discrete level shifts has a nonlinear dependence structure, because the innovations that define break points are much more persistent than other innovations. Models that capture this nonlinearity will generate more accurate inference about the features of the data than can be achieved with estimators such as modified GPH.

Specifying models that identify the persistent innovations in a time series is nontrivial, especially given that each of these shocks may have a different origin. They may arise from a political event, a weather event, a war, a new technology, an earnings announcement, or a government policy change, to name a few possibilities. Most Markov-switching models, and the particular STOPBREAK model of Engle and Smith (1999), take an agnostic approach and focus only on the time series characteristics of the data when identifying break points. However, Filardo (1994) and Filardo and Gordon (1998) estimated Markov-switching models that use observed data to help identify break points. Further research in this vein will improve model performance and enable better discrimination between models with occasional persistent shocks and linear long-memory models.

ACKNOWLEDGMENTS

The author is a member of the Giannini Foundation of Agricultural Economics. He thanks Oscar Jorda, Prasad Naik, Doug Steigerwald, the editor, an associate editor, and an anonymous referee for suggestions and comments that improved the quality of this article.

APPENDIX: PROOFS

Proof of Theorem 1

Recall that $p_T = cJ/T$. We can decompose $\log(f_j)$ as

$$\log f_j = \log\left(f_\varepsilon(\omega_j) + \frac{p_T}{p_T^2 + (1-p_T)(2 - 2\cos(\omega_j))} f_\eta(\omega_j)\right)$$

$$= -\log(1 + p_T^{-2}\omega_j^2)$$

$$
\begin{aligned}
&+\log\biggl(f_\varepsilon(\omega_j)(1+p_T^{-2}\omega_j^2)\\
&\qquad+\frac{p_T(1+p_T^{-2}\omega_j^2)}{p_T^2+(1-p_T)(2-2\cos(\omega_j))}f_\eta(\omega_j)\biggr)\\
&=-\log\bigl(1+(2\pi/c)^2(j/J)^2\bigr)\\
&\quad+\log(1+D_{jT})+\log(f_\eta(\tilde{\omega}_J)/p_T),
\end{aligned}
$$

where $\tilde{\omega}_J=\arg\min_{1\le j\le J} f_\eta(\omega_j)$ and

$$
\begin{aligned}
D_{jT}&=\frac{p_T}{f_\eta(\tilde{\omega}_J)}f_\varepsilon(\omega_j)(1+p_T^{-2}\omega_j^2)\\
&\quad+\frac{(1+p_T^{-2}\omega_j^2)f_\eta(\omega_j)}{f_\eta(\tilde{\omega}_J)(1+p_T^{-2}(1-p_T)(2-2\cos(\omega_j)))}-1.
\end{aligned}
$$

Using the fact that for all j, $2-2\cos(\omega_j)=\omega_j^2-\omega_j^3\sin(\bar{\omega}_j)/3\le\omega_j^2$ for some $0\le\bar{\omega}_j\le\omega_j$, we have

$$
\begin{aligned}
D_{jT}&=\frac{p_T}{f_\eta(\tilde{\omega}_J)}f_\varepsilon(\omega_j)(1+p_T^{-2}\omega_j^2)\\
&\quad+\bigl(f_\eta(\omega_j)(1+p_T^{-2}\omega_j^2)\\
&\qquad-f_\eta(\tilde{\omega}_J)\bigl(1+p_T^{-2}(1-p_T)\bigl(2-2\cos(\omega_j)\bigr)\bigr)\bigr)\\
&\quad\times\bigl(f_\eta(\tilde{\omega}_J)\bigl(1+p_T^{-2}(1-p_T)\bigl(2-2\cos(\omega_j)\bigr)\bigr)\bigr)^{-1}\\
&=\bigl(1+(2\pi/c)^2(j/J)^2\bigr)\frac{p_Tf_\varepsilon(\omega_j)}{f_\eta(\tilde{\omega}_J)}\\
&\quad+\frac{(1+(2\pi/c)^2(j/J)^2)(f_\eta(\omega_j)-f_\eta(\tilde{\omega}_J))}{f_\eta(\tilde{\omega}_J)(1+p_T^{-2}(1-p_T)(2-2\cos(\omega_j)))}\\
&\quad+\frac{p_T^{-1}(2-2\cos(\omega_j))+\frac{1}{3}p_T^{-2}\omega_j^3\sin(\bar{\omega}_j)}{1+p_T^{-2}(1-p_T)(2-2\cos(\omega_j))}.
\end{aligned}
$$

Note that $D_{jT}\ge 0$, and because $\omega_j<\pi$, we have $0\le\sin(\bar{\omega}_j)\le 1$, which implies that

$$
\begin{aligned}
D_{jT}&\le\bigl(1+(2\pi/c)^2(j/J)^2\bigr)\frac{p_Tf_\varepsilon(\omega_j)+f_\eta(\omega_j)-f_\eta(\tilde{\omega}_J)}{f_\eta(\tilde{\omega}_J)}\\
&\quad+p_T^{-1}\omega_j^2+\frac{1}{3}p_T^{-2}\omega_j^3\\
&=\bigl(1+(2\pi/c)^2(j/J)^2\bigr)\frac{cJT^{-1}f_\varepsilon(\omega_j)+f_\eta(\omega_j)-f_\eta(\tilde{\omega}_J)}{f_\eta(\tilde{\omega}_J)}\\
&\quad+\frac{(2\pi j)^2}{cJT}+\frac{(2\pi j)^3}{3(cJT)^2}.
\end{aligned}
$$

Now

$$
\begin{aligned}
d_*&=\frac{\sum_{j=1}^J(X_j-\overline{X})\log f_j}{\sum_{j=1}^J(X_j-\overline{X})^2}\\
&=\frac{-\sum_{j=1}^J(X_j-\overline{X})\log(1+(2\pi/c)^2(j/J)^2)}{\sum_{j=1}^J(X_j-\overline{X})^2}\\
&\quad+\frac{\sum_{j=1}^J(X_j-\overline{X})\log(1+D_{jT})}{\sum_{j=1}^J(X_j-\overline{X})^2}. \qquad \text{(A.1)}
\end{aligned}
$$

The results from Hurvich and Beltrao (1994),

$$
|X_j-\overline{X}|=O(\log J)=O(\log T),
$$
$$
\sum_{j=1}^J(X_j-\overline{X})^2=4J\bigl(1+o(1)\bigr),
$$

and

$$
X_j-\overline{X}=-2\log j+2J^{-1}\sum_{k=1}^J\log k+O(J^2/T^2),
$$

and the formula $J^{-1}\sum_{k=1}^J\log k=\log J-1+o(1)$ imply that

$$
\begin{aligned}
&\frac{-\sum_{j=1}^J(X_j-\overline{X})\log(1+(2\pi/c)^2(j/J)^2)}{\sum_{j=1}^J(X_j-\overline{X})^2}\\
&\quad=\frac{\sum_{j=1}^J(1+\log(j/J)+o(1))\log(1+(2\pi/c)^2(j/J)^2)}{2J(1+o(1))}\\
&\quad\to .5\int_0^1\bigl(1+\log(x)\bigr)\log\bigl(1+(2\pi/c)^2x^2\bigr)\,dx\\
&\quad=1-.25\Phi\bigl(-(2\pi/c)^2,2,.5\bigr),
\end{aligned}
$$

where $\Phi(x,s,a)$ denotes the Lerch transcendent function (Gradshteyn and Ryzhik 1980, p. 1072).

For the second term in (A.1), note that $|f_\eta(\omega_j)-f_\eta(\tilde{\omega}_J)|\le\widetilde{B}_1|\omega_j-\tilde{\omega}_J|=O(JT^{-1})$ for all $j=1,2,\dots,J$. Then, using $|\log(1+x)|\le x$ for $x\ge 0$ yields

$$
\begin{aligned}
&\left|\frac{\sum_{j=1}^J(X_j-\overline{X})\log(1+D_{jT})}{\sum_{j=1}^J(X_j-\overline{X})^2}\right|\\
&\le\sum_{j=1}^J|X_j-\overline{X}|\Biggl(\biggl(1+\Bigl(\frac{2\pi j}{cJ}\Bigr)^2\biggr)\\
&\qquad\times\frac{cJT^{-1}f_\varepsilon(\omega_j)+f_\eta(\omega_j)-f_\eta(\tilde{\omega}_J)}{f_\eta(\tilde{\omega}_J)}+\frac{(2\pi j)^2}{cJT}+\frac{(2\pi j)^3}{3(cJT)^2}\Biggr)\\
&\qquad\times\Biggl(\sum_{j=1}^J(X_j-\overline{X})^2\Biggr)^{-1}\\
&=O\bigl(J^{-1}\log(J)\bigr)\sum_{j=1}^J\Biggl(\biggl(1+\Bigl(\frac{2\pi j}{cJ}\Bigr)^2\biggr)\frac{cJT^{-1}f_\varepsilon(\omega_j)}{f_\eta(\tilde{\omega}_J)}\\
&\qquad+\biggl(1+\Bigl(\frac{2\pi j}{cJ}\Bigr)^2\biggr)O\Bigl(\frac{J}{T}\Bigr)+\frac{(2\pi)^2}{c}\frac{j^2}{JT}+\frac{(2\pi)^3}{3c^2}\frac{j^3}{J^2T^2}\Biggr)\\
&=O\bigl(JT^{-1}\log(J)\bigr)+O\bigl(JT^{-1}\log(J)\bigr)\\
&\quad+O\bigl(JT^{-1}\log(J)\bigr)+O\bigl(JT^{-2}\log(J)\bigr)\\
&=O\bigl(JT^{-1}\log(J)\bigr).
\end{aligned}
$$

Thus, under the assumption that $JT^{-1}\log(J)\to 0$, we have $d_*\to 1-.25\Phi(-(2\pi/c)^2,2,.5)$ as $T\to\infty$.

Proof of Theorem 2

We can write

$$d_*^k = \left(\sum_{j=1}^{J} (X_j - \overline{X}) \log f_j - \frac{(\sum_{j=1}^J (X_j - \overline{X})(Z_{kj} - \overline{Z}_k))(\sum_{j=1}^J (Z_{kj} - \overline{Z}_k) \log f_j)}{\sum_{j=1}^J (Z_{kj} - \overline{Z}_k)^2} \right) \times \left(\sum_{j=1}^{J} (X_j - \overline{X})^2 - \frac{(\sum_{j=1}^J (X_j - \overline{X})(Z_{kj} - \overline{Z}_k))^2}{\sum_{j=1}^J (Z_{kj} - \overline{Z}_k)^2} \right)^{-1}.$$

From the proof of Theorem 1, we have

$$\sum_{j=1}^{J} (X_j - \overline{X})(Z_{kj} - \overline{Z}_k) = 4\big(1 - .25\Phi\big(-(2\pi/k)^2, 2, .5\big)\big) J\big(1 + o(1)\big),$$

$$\sum_{j=1}^{J} (X_j - \overline{X}) \log f_j = 4\big(1 - .25\Phi\big(-(2\pi/c)^2, 2, .5\big)\big) J\big(1 + o(1)\big),$$

and

$$\sum_{j=1}^{J} (X_j - \overline{X})^2 = 4J\big(1 + o(1)\big),$$

where $\Phi(x, s, a)$ denotes the Lerch transcendent function (Gradshteyn and Ryzhik 1980, p. 1072).

Note that

$$\overline{Z}_k = -J^{-1} \sum_{j=1}^{J} \log\big(1 + (2\pi/k)^2 (j/J)^2\big) - 2 \log(kJ/T) = -\big(1 + o(1)\big) \int_0^1 \log\big(1 + (2\pi/k)^2 x^2\big)\, dx - 2 \log(kJ/T) = -\big(1 + o(1)\big)\big(\log\big(1 + (2\pi/k)^2\big) + (k/\pi) \tan^{-1}(2\pi/k) - 1\big) - 2 \log(kJ/T).$$

Thus

$$J^{-1} \sum_{j=1}^{J} (Z_{kj} - \overline{Z}_k)^2 \to \int_0^1 \left(\log\left(\frac{1 + (2\pi/k)^2 x^2}{1 + (2\pi/k)^2} \right) - (k/\pi) \tan^{-1}(2\pi/k) + 1 \right)^2 dx = 4 - \left(\frac{k}{\pi} \tan^{-1}\left(\frac{2\pi}{k} \right) \right)^2 - \frac{2k}{\pi} \left(\mathrm{Im}\left(\mathrm{Li}_2\left(\frac{1}{2} + \frac{\pi i}{k} \right) \right) + \frac{1}{2} \tan^{-1}\left(\frac{2\pi}{k} \right) \log\left(\frac{1}{4} + \frac{\pi^2}{k^2} \right) \right),$$

where $\mathrm{Im}(x + iy) \equiv y$, and Li_2 signifies the dilogarithm.

Similarly, for $\sum_{j=1}^J (Z_{kj} - \overline{Z}_k) \log f_j$, we have

$$J^{-1} \sum_{j=1}^{J} (Z_{kj} - \overline{Z}_k) \log f_j \to \int_0^1 \left(\log\left(\frac{1 + (2\pi/k)^2 x^2}{1 + (2\pi/k)^2} \right) - (k/\pi) \tan^{-1}(2\pi/k) + 1 \right) \times \log\big(1 + (2\pi/c)^2 x^2\big)\, dx$$
$$= 4 - \frac{ck}{\pi^2} \tan^{-1}\left(\frac{2\pi}{c} \right) \tan^{-1}\left(\frac{2\pi}{k} \right) - \frac{k + c}{\pi} \left(\mathrm{Im}\left(\mathrm{Li}_2\left(\frac{k + 2\pi i}{c + k} \right) \right) + \frac{1}{2} \tan^{-1}\left(\frac{2\pi}{k} \right) \log\left(\frac{c^2 + 4\pi^2}{(k + c)^2} \right) \right) + \frac{|k - c|}{\pi} \left(\mathrm{Im}\left(\mathrm{Li}_2\left(\frac{\min(k, c) + 2\pi i}{-|k - c|} \right) \right) + \frac{1}{2} \tan^{-1}\left(\frac{2\pi}{\min(k, c)} \right) \log\left(\frac{\max(k^2, c^2) + 4\pi^2}{(k + c)^2} \right) \right)$$
$$\equiv 4h_k.$$

Define

$$r_k \equiv \big(1 - .25\Phi\big(-(2\pi/k)^2, 2, .5\big)\big) \times \left(1 - \left(\frac{k}{2\pi} \tan^{-1}\left(\frac{2\pi}{k} \right) \right)^2 - \frac{k}{2\pi} \left(\mathrm{Im}\left(\mathrm{Li}_2\left(\frac{1}{2} + \frac{\pi i}{k} \right) \right) + \frac{1}{2} \tan^{-1}\left(\frac{2\pi}{k} \right) \log\left(\frac{1}{4} + \frac{\pi^2}{k^2} \right) \right) \right)^{-1},$$

and

$$v_k \equiv 1 - r_k \big(1 - .25\Phi\big(-(2\pi/k)^2, 2, .5\big)\big);$$

then the result follows.

Proof of Theorem 3

The log spectrum of the fractionally integrated process is $\log f_j = dX_j + \log f_{uj}$. Thus the modified GPH estimator is

$$\hat{d}^k = d + (\widetilde{X}' M_Z \widetilde{X})^{-1} \widetilde{X}' M_Z \log f_u + (\widetilde{X}' M_Z \widetilde{X})^{-1} \widetilde{X}' M_Z \log(\hat{f}/f), \quad \text{(A.2)}$$

where $\widetilde{X} = X - \overline{X}$, $M_Z = I - \widetilde{Z}_k (\widetilde{Z}_k' \widetilde{Z}_k)^{-1} \widetilde{Z}_k'$, and $\widetilde{Z}_k = Z_k - \overline{Z}_k$.

Consider the second term on the right side of (A.2). We have

$$\widetilde{X}' M_Z \log f_u = \sum_{j=1}^{J} (X_j - \overline{X}) \log f_{uj} - r_k \big(1 + o(1)\big) \sum_{j=1}^{J} (Z_{kj} - \overline{Z}_k) \log f_{uj},$$

where r_k is as defined in the proof of Theorem 2. From Hurvich et al. (1998), a second-order expansion of f_u around $\omega = 0$ yields

$$\log f_{uj} = \log f_u(0) + \frac{\omega_j^2}{2}\frac{f_u''(0)}{f_u(0)} + \frac{\omega_j^3}{6}K_j,$$

where K_j is bounded uniformly in j for sufficiently large T. Given this, Hurvich et al. (1998) showed that

$$\sum_{j=1}^{J}(X_j - \overline{X})\log f_{uj} = \frac{-8\pi^2}{9}\frac{f_u''(0)}{f_u(0)}\frac{J^3}{T^2} + o(J^3/T^2),$$

where f_u'' denotes the second derivative of f_u. Similarly,

$$\begin{aligned}\sum_{j=1}^{J}&(Z_{kj} - \overline{Z}_k)\log f_{uj}\\ &= \sum_{j=1}^{J}(Z_{kj} - \overline{Z}_k)\left(\frac{\omega_j^2}{2}\frac{f_u''(0)}{f_u(0)} + \frac{\omega_j^3}{6}K_j\right)\\ &= \frac{2\pi^2 f_u''(0)}{f_u(0)}\frac{J^3}{T^2}\left(J^{-1}\sum_{j=1}^{J}(Z_{kj} - \overline{Z}_k)(j/J)^2\right)\\ &\quad + O\left(T^{-3}\sum_{j=1}^{J} j^3\right),\end{aligned}$$

where I used the fact that $|Z_{kj} - \overline{Z}_k| = O(1)$ uniformly in j.

Now, using arguments from the proof of Theorem 2, we have

$$\begin{aligned}J^{-1}&\sum_{j=1}^{J}(Z_{kj} - \overline{Z}_k)(j/J)^2\\ &= \int_0^1\left(\log\left(\frac{1+(2\pi/k)^2x^2}{1+(2\pi/k)^2}\right)\right.\\ &\qquad \left. - 2\big((k/2\pi)\tan^{-1}(2\pi/k) - 1\big)\right)x^2\,dx + o(1)\\ &= \left(-\frac{4}{9}\left(\frac{3k^2}{8\pi^2}+1\right) + \frac{k}{3\pi}\left(\frac{k^2}{4\pi^2}+2\right)\tan^{-1}\left(\frac{2\pi}{k}\right)\right)\\ &\quad + o(1).\end{aligned}$$

Thus

$$\begin{aligned}\sum_{j=1}^{J}&(Z_{kj} - \overline{Z}_k)\log f_{uj}\\ &= \frac{-8\pi^2}{9}\frac{f_u''(0)}{f_u(0)}\frac{J^3}{T^2}\\ &\quad \times\left(\left(\frac{3k^2}{8\pi^2}+1\right) - \frac{3k}{4\pi}\left(\frac{k^2}{4\pi^2}+2\right)\tan^{-1}\left(\frac{2\pi}{k}\right)\right)\\ &\quad + o(J^3/T^2).\end{aligned}$$

Then the second term on the right side of (A.2) is

$$(\widetilde{X}'M_Z\widetilde{X})^{-1}\widetilde{X}'M_Z\log f_u = -b_k\frac{2\pi^2}{9}\frac{f_u''(0)}{f_u(0)}\frac{J^2}{T^2} + o(J^2/T^2),$$

where $b_k = \frac{1}{v_k}(1 - r_k((\frac{3k^2}{8\pi^2}+1) - \frac{3k}{4\pi}(\frac{k^2}{4\pi^2}+1)\tan^{-1}(\frac{2\pi}{k})))$, and v_k is as defined in the proof of Theorem 2.

For the last term in (A.2), I use the proof of lemma 8 of Hurvich et al. (1998). Their proof goes through if their $a_j = X_j - \overline{X}$ is replaced by $((X_j - \overline{X}) - r_k(1+o(1))(Z_{kj} - \overline{Z}_k))$ and their $2S_{xx}$ is replaced by $4Jv_k(1+o(1))$. These replacements are valid, because the substituted terms are of the same order of magnitude as their replacements. It follows that $(\widetilde{X}'M_Z\widetilde{X})^{-1}\widetilde{X}'M_Z E(\log(\hat{f}/f)) = O(\log^3 J/J)$. Thus the bias is

$$E(\hat{d}^m - d) = -b_k\frac{2\pi^2}{9}\frac{f_u''(0)}{f_u(0)}\frac{J^2}{T^2} + o(J^2/T^2) + O(\log^3 J/J).$$

For the variance, I use the proof of theorem 1 of Hurvich et al. (1998). Replacing their a_j by $((X_j - \overline{X}) - r_k(1+o(1))(Z_{kj} - \overline{Z}_k))$ and their $2S_{xx}$ by $4Jv_k(1+o(1))$ leads to

$$\begin{aligned}\operatorname{var}(\hat{d}^k) &= (\widetilde{X}'M_Z\widetilde{X})^{-2}\operatorname{var}\big(\widetilde{X}'M_Z\log(\hat{f}/f)\big)\\ &= \frac{\pi^2}{24Jv_k} + o(J^{-1}).\end{aligned}$$

Proof of Corollary 1

This result follows directly from Theorems 2 and 3 herein and theorem 2 of Hurvich et al. (1998), with their a_j replaced by $((X_j - \overline{X}) - r_k(1+o(1))(Z_{kj} - \overline{Z}_k))$ and their $2S_{xx}$ replaced by $4Jv_k(1+o(1))$.

[Received January 2003. Revised May 2004.]

REFERENCES

Agiakloglou, C., Newbold, P., and Wohar, M. (1993), "Bias in an Estimator of the Fractional Difference Parameter," *Journal of Time Series Analysis*, 14, 235–246.

Bai, J., and Perron, P. (1998), "Testing for and Estimation of Multiple Structural Changes," *Econometrica*, 66, 47–79.

Breidt, F. J., and Hsu, N. J. (2002), "A Class of Nearly Long-Memory Time Series Models," *International Journal of Forecasting*, 18, 265–281.

Cavanagh, C. L., Elliott, G., and Stock, J. H. (1995), "Inference in Models With Nearly Integrated Regressors," *Econometric Theory*, 11, 1131–1147.

Chen, C., and Tiao, G. C. (1990), "Random Level-Shift Time Series Models, ARIMA Approximations, and Level-Shift Detection," *Journal of Business & Economic Statistics*, 8, 83–97.

Chib, S. (1998), "Estimation and Comparison of Multiple Change-Point Models," *Journal of Econometrics*, 86, 221–241.

Deo, R. S., and Hurvich, C. M. (2001), "On the Log Periodogram Regression Estimator of the Memory Parameter in Long-Memory Stochastic Volatility Models," *Econometric Theory*, 17, 686–710.

Diebold, F. X., and Inoue, A. (2001), "Long Memory and Regime Switching," *Journal of Econometrics*, 101, 131–159.

Engle, R. F., and Smith, A. D. (1999), "Stochastic Permanent Breaks," *Review of Economics and Statistics*, 81, 553–574.

Filardo, A. J. (1994), "Business Cycle Phases and Their Transitional Dynamics," *Journal of Business & Economic Statistics*, 12, 299–308.

Filardo, A. J., and Gordon, S. (1998), "Business Cycle Durations," *Journal of Econometrics*, 85, 99–113.

Geweke, J., and Porter-Hudak, S. (1983), "The Estimation and Application of Long-Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221–238.

Gradshteyn, I. S., and Ryzhik, A. D. (1980), *Table of Integrals, Series, and Products: Corrected and Enlarged Edition*, New York: Academic Press.

Granger, C. W. J. (1981), "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*, 16, 121–130.

Granger, C. W. J., and Ding, Z. (1996), "Varieties of Long-Memory Models," *Journal of Econometrics*, 73, 61–77.

Granger, C. W. J., and Hyung, N. (1999), "Occasional Structural Breaks and Long Memory," Discussion Paper 99-14, University of California, Dept. of Economics.

Granger, C. W. J., and Joyeux, N. (1980), "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15–39.

Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

——— (1994), *Time Series Analysis*, Princeton, NJ: Princeton University Press.

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

Hurvich, C. M., and Beltrao, K. I. (1994), "Automatic Semiparametric Estimation of the Memory Parameter of a Long-Memory Time Series," *Journal of Time Series Analysis*, 15, 285–302.

Hurwich, C. M., and Deo, R. (1999), "Plug-in Selection of the Number of Frequencies in Regression Estimates of the Memory Parameter of a Long-Memory Time Series," *Journal of Time Series Analysis*, 20, 331–341.

Hurvich, C. M., Deo, R., and Brodsky, J. (1998), "The Mean Squared Error of Geweke and Porter-Hudak's Estimator of the Memory Parameter of a Long-Memory Time Series," *Journal of Time Series Analysis*, 19, 19–46.

Hurst, H. E. (1951), "Long-Term Storage Capacity of Reservoirs," *Transactions of the American Society of Civil Engineers*, 116, 770–799.

Liu, M. (2000), "Modeling Long Memory in Stock Market Volatility," *Journal of Econometrics*, 99, 139–171.

Lobato, I. N., and Savin, N. E. (1998), "Real and Spurious Long-Memory Properties of Stock Market Data," *Journal of Business & Economic Statistics*, 16, 261–268.

McCulloch, R. E., and Tsay, R. S. (1993), "Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series," *Journal of the American Statistical Association*, 88, 968–978.

Phillips, P. C. B. (1988), "Regression Theory for Near-Integrated Time Series," *Econometrica*, 56, 1021–1043.

Timmermann, A. (2001), "Structural Breaks, Incomplete Information, and Stock Prices," *Journal of Business & Economic Statistics*, 19, 299–314.

Velasco, C. (2000), "Non-Gaussian Log Periodogram Regression," *Econometric Theory*, 16, 44–79.