Long Memory or Structural Change: Testing Method and Empirical Examination

Chih-Chiang Hsu
(presenter)

Department of Economics
National Central University
Chung-Li 32054, Taiwan
Tel: +886 3 422 7151 /6308
Fax: +886 3 422 2876
Email: cchsu@mgt.ncu.edu.tw

Chung-Ming Kuan

Institute of Economics
Academia Sinica
Taipei 115, Taiwan
Tel: +886 2 2782 2791 /645
Fax: +886 2 2785 3946
Email: ckuan@ieas.econ.sinica.edu.tw

This version: January 15, 2000
(Preliminary; do not cite or quote without permission)

† Author for correspondence: C.-C. Hsu

Abstract

In this paper, we focus on how to test for long-range dependence when the process may have a one-time mean change and how to estimate the change point when data may be long-range dependent. We first analyzed why traditional long-memory tests have serious size distortions when data have short memory with breaks. In order to overcome this problem, a local Whittle method is proposed. Simulation results confirm that our change-point estimator is well behaved even when data are long-range dependent, and that our test for long memory maintains proper size when a change is present. These results indicate that our method is practically useful and has a much wider applicability. In order to assess the empirical relevance of our procedure, we applied it to analyze monthly G7 inflation rates.

Keywords: change point, local Whittle estimation, long memory, R/S statistic, spurious change

1 Introduction

In recent years, increasing interest has been devoted to the research of long-range dependence (long memory) in economic and financial variables. To detect long-range dependence, Hurst (1951) suggested the normalized *rescaled range* (R/S) test. Mandelbrot and Wallis (1969), Mandelbrot (1975), and Mandelbrot and Taqqu (1979) further showed that the R/S test is robust to non-Gaussian distributions or distributions with infinite variance. Lo (1991) modified the R/S statistic to accommodate short-range dependence; see also Beran (1994) for a more detailed survey. In addition to the R/S test, Geweke and Porter-Hudak (1983) and Robinson (1995a) suggested a frequency-domain approach which is based on regressing the logarithm of the periodogram on trigonometric function at low frequencies. The estimated slope coefficient is an estimate of the long-memory parameter d. As the normalized slope coefficient is asymptotically normally distributed, it can be used to test for long-range dependence.

On the other hand, numerous empirical studies suggest that many economic and financial data exhibit changing mean or changing variance. In Kuan and Hsu (1998) we find that least-squares estimation of the change point may suggest a spurious change when data have long-range dependence. When data have structural changes, the aforementioned tests are unable to detect long-range dependence; see e.g., Klemeš (1974) and Teverovsky and Taqqu (1997). In this paper, we first show that, when data are weakly dependent with a mean change, existing tests have serious size distortions in finite samples so that the null hypothesis of short memory is rejected far too often even when it is true. These results suggest that to estimate a change point, one should know whether data are long-range dependent, and that to test for long-range dependence, one must know if a change exists. This creates a dilemma in practice.

In this paper, we propose a semi-parametric method for jointly estimating the change point and long-memory parameter d. Our simulation results indicate that the normalized estimate of d is asymptotically normally distributed, from which a test for long-range dependence can be constructed. As this method takes both structural change and long-range dependence into account, the resulting change-point estimator does not have the "spurious change" problem observed in Kuan and Hsu (1998), and the proposed test does not suffer size distortions when a change is present. This method thus enables us to distinguish between long-memory series and short-memory series with a mean change. Our procedure is also applied to analyze monthly inflation rates in G7 countries.

This paper is organized as the following. In Section 2, we show by simulations and analytically that the existing tests are inappropriate to detect long-range dependence when there is a structural change. The proposed method is discussed in Section 3. Monte Carlo simulation results are reported in Section 4. An empirical study of the G7 inflation rates is included in Section 5. Section 6 concludes.

2 Tests for Long-Range Dependence

In this section, we discuss traditional tests for long-range dependence and analyze their behavior when data are weakly dependent with a mean change. Consider a time series

$$y_t = \mu + \varepsilon_t, \qquad t = 1, 2 \ldots, \tag{1}$$

where $\mu = \mathrm{E}(y_t)$, and ε_t are random variables with mean zero. We impose the following condition on ε_t.

[A] $\{\varepsilon_t\}$ is a strong-mixing sequence with mixing coefficients α_j such that for some $\delta > 2$, $\sum_{j=1}^{\infty} \alpha_j^{1-(2/\delta)} < \infty$ and $\sup_t E|\varepsilon_t|^\delta < \infty$.

Let $Y(t) = \sum_{i=1}^{t} y_i$ denote the partial sum of t observations. Then the classical rescaled range (R/S) statistic is proposed by Hurst (1951)

$$Q(T) = \frac{1}{s_T}\left[\max_{1\le t\le T}\left(Y(t) - \frac{t}{T}Y(T)\right) - \min_{1\le t\le T}\left(Y(t) - \frac{t}{T}Y(T)\right)\right], \tag{2}$$

where

$$s_T = \left[\frac{1}{T}\sum_{t=1}^{T} y_t^2 - \frac{1}{T^2}Y(T)^2\right]^{1/2}, \tag{3}$$

is the sample standard deviation.

Lo (1991) pointed out that the classical R/S statistic is sensitive to the presence of short-range dependence. He modified the R/S statistic (2) by incorporating short-range dependence into the statistic. Specifically, let

$$\sigma^2 = \lim_{T\to\infty} \frac{1}{T}\mathrm{E}\left(\sum_{t=1}^{T} \varepsilon_t\right)^2.$$

When ε_t are serially uncorrelated (i.e., no short-range dependence),

$$\sigma^2 = \lim_{T\to\infty} \frac{1}{T}\mathrm{E}\left(\sum_{t=1}^{T} \varepsilon_t^2\right) \equiv \sigma_\varepsilon^2,$$

but they are not the same otherwise. Lo (1991) suggested using a serial-correlation-consistent estimator for σ^2 to replace the standard variance estimator s_T^2. For a given truncation lag (q), the estimator for σ is

$$\tilde{s}_T(q) = \left[\sum_{j=-q}^{q} \omega_q(j) \hat{\gamma}(j) \right]^{1/2}, \tag{4}$$

where $\hat{\gamma}(j)$ represent the usual estimators for autocovariances and the weights $\omega_q(j)$ are determined by some kernels; see e.g., Newey and West (1987) and Andrews (1991). Let $Q(T,q)$ denote the modified R/S statistic. When $q = 0$, $Q(T,0) = Q(T)$, the classical R/S statistic. Under the null hypothesis of short memory, Lo (1991) showed that

$$V(T,q) = \frac{1}{\sqrt{T}} Q(T,q) \Rightarrow \text{range}(B^0),$$

where $\Rightarrow$ denotes weak convergence (of associated probability measures), and B^0 is a standard Brownian bridge on the unit interval. In what follows, we also denote $\to^p$ as convergence in probability, and $\to^d$ as convergence in distribution.

When $\{y_t\}$ is a long-range dependent process, the spectral density of y_t at zero frequency is unbounded:

$$f_y(0) = \sum_{j=0}^{k} \gamma(j) = O(k^{2d}),$$

for $0 < d < 0.5$.[1] The parameter d is known as the fractionally differencing parameter, which measures the intensity of memory. Mandelbrot (1975) also proved that if y_t are long-range dependent, then

$$\mathrm{E}[Q(T)] \sim C_d\, T^{0.5+d},$$

where C_d is a positive and finite constant. In practice, d may be estimated as follows. First partition data into m blocks, each starts at $m_i = (iT/m) + 1$, $i = 1, 2, \ldots$, and has n observations, where $m_i + n \leq T$. For each i, we use the observations from m_i through n to compute the R/S statistic $Q_i(n)$, where $n = 1, 2, \ldots$. By regressing $\log Q_i(n)$ on $\log n$, the estimated slope is an estimate of $d + 1/2$, which should tend to $1/2$ for short-memory data.

[1]For $-0.5 < d < 0$, $f_y(0) = 0$ so that the sum of all covariances is zero. For $d = 0$, y_t is a stationary process with bounded spectral density.

Geweke and Porter-Hudak (1983) (henceforth GPH) introduced a frequency-domain method to estimate d. Let $I(v_j)$ be the periodogram of y_t evaluated at different spectral ordinates $v_j = 2\pi j/T$, $j = 1, \ldots, [T^c] < [T/2]$, where c is a constant determining the truncation of higher ordinates. The parameter d is then estimated from the following regression:

$$\log\{I(v_j)\} = a - d \log\{4 \sin^2(v_j/2)\} + \epsilon_j.$$

Consistency and asymptotic normality of this estimator were established in Robinson (1995a). In practice, the GPH method uses only lower ordinates of the periodogram, for example, $c = 0.5$, in the regression above. Robinson (1995a) also suggested to truncate the beginning ordinates, so as to prevent biases.

To understand the performance of the R/S and GPH tests, we simulate these tests based on six different data generating processes (DGPs). Let u_t be a Gaussian white noise and L denote the back-shift operator. The DGPs are:

DGP (1): $y_t = 1 + \varepsilon_t, \qquad \varepsilon_t = u_t.$

DGP (2): $y_t = 1 + \varepsilon_t, \qquad \varepsilon_t = 0.5\varepsilon_{t-1} + u_t.$

DGP (3): $y_t = \begin{cases} 1 + \varepsilon_t, & t = 1, \ldots, [T/2], \\ 2 + \varepsilon_t, & t = [T/2] + 1, \ldots, T, \end{cases} \quad \varepsilon_t = u_t.$

DGP (4): $y_t = \begin{cases} 1 + \varepsilon_t, & t = 1, \ldots, [T/2], \\ 2 + \varepsilon_t, & t = [T/2] + 1, \ldots, T, \end{cases} \quad \varepsilon_t = 0.5\varepsilon_{t-1} + u_t.$

DGP (5): $y_t = 1 + \varepsilon_t, \qquad (1 - L)^{0.3}\varepsilon_t = u_t.$

DGP (6): $y_t = 1 + \varepsilon_t, \qquad (1 - 0.5L)(1 - L)^{0.3}\varepsilon_t = u_t.$

Here, (1) and (2) are data with short memory; (3) and (4) are data with short memory and an one-time change in the middle of sample; (5) and (6) are data with long memory. We follow the methods suggested by McLeod and Hipel (1978), Hosking (1984), and Chung (1994) to generate long memory data.

The modified R/S statistic are computed with $q = 0, 10, 20$, and q^*, where q^* is chosen by the data-dependent formula of Andrews (1991). For the GPH approach, low frequencies used in the regression start from $[T^{0.1}]$, and different upper truncations are considered at $[T^c]$ with $c = 0.4, 0.5, 0.6$. For each experiment, the number of replications is $5,000$, $T = 512$, and the nominal level is 5%. The results are summarized in Table 1.

Table 1: Finite sample performance of long memory tests

DGPs	R/S Test				GPH Test		
	$V(T,0)$	$V(T,10)$	$V(T,20)$	$V(T,q^*)$	$c=0.4$	$c=0.5$	$c=0.6$
(1)	3.1	2.3	1.4	3.1	5.4	4.6	4.3
(2)	66.5	3.6	1.4	3.4	6.0	5.9	13.2
(3)	100	100	100	100	76.9	84.0	84.4
(4)	100	99.3	96.3	98.8	55.9	59.3	69.9
(5)	96.2	44.2	20.1	51.2	38.8	59.0	83.6
(6)	99.9	45.5	18.5	11.4	41.0	66.6	93.5

From Table 1, we can see that the modified R/S test does not suffer much size distortion when data are weakly dependent (DGP (2)), but it is not very powerful against long-memory data (DGP (5) and (6)). The GPH test, on the other hand, is more robust to weak dependence in data and relatively more powerful than the modified R/S test. These two tests, however, reject the null hypothesis of short memory with high probabilities when data are short memory with a mean change (DGP (3) and (4)). These results suggest that the R/S and GPH tests may reject the null hypothesis because of either long-range dependence or a structural change. As such, these tests could yield very misleading inferences.

To understand why such size distortions occur, we analyze the asymptotic behavior of the R/S statistic when y_t are weakly dependent with mean changes. Consider the following time-varying behavior of μ_t:

$$\mu_t = \mu_0 + T^{-\nu}\psi(t/T), \tag{5}$$

where $\psi(\tau)$, $\tau = t/T$, is a real-valued function of bounded variation on $\tau \in [0,1]$ and $\nu = 0$ or $1/2$. Note that $\psi(\tau)$ can represent a wide variety of structural changes, e.g., multiple structural changes or continuous parameter changes. When $\nu = 0$, (5) is a global alternative to the constant mean; when $\nu = 1/2$, (5) is a sequence of local alternatives.

Theorem 2.1 *Given (1) and (5), suppose that condition [A] is satisfied.*

1. *If* $\nu = 1/2$, *then*

$$V(T,q) \xrightarrow{D} \text{range}_{0\leq\tau\leq 1}\left(B^0(\tau) + \frac{1}{\sigma}\Psi(\tau)\right),$$

where for a process f, $\text{range}(f) = \max(f) - \min(f)$, B^0 *is the Brownian bridge, and* $\Psi(\tau) = \int_0^\tau \psi(z)dz - \tau\int_0^1 \psi(z)dz$.

2. *If* $\nu = 0$, *and* $\tilde{s}_T^2(q)$ *is* $O_p(1)$, *then*

$$V(T,q) \to^p \infty.$$

In Theorem 2.1, the first result indicates that the modified R/S statistic has non-trivial local power and the second result shows that for the global alternative $\mu_t = \mu_0 + \psi(t/T)$, the modified R/S statistic diverges in probability so that it will reject the null hypothesis with probability approaching one. This explains why the R/S statistic rejects the null hypothesis of short memory far too often when data are generated according to DGP (3) and (4). Although this is a result for the R/S statistic, one can see from Table 1 that the GPH test must have a similar problem.

3 The Local Whittle Method

Kuan and Hsu (1998) shows that the estimated change point could also be a spurious one when data are long-range dependent. On the other hand, the preceding section shows that traditional tests for long-range dependence could be misleading when a structural change is present. These results suggest that the change-point estimation must take potential long-range dependence into account and that tests for long-range dependence should also consider potential structural changes. In this section, we propose a method for jointly estimating the change point and long-memory parameter, so that aforementioned difficulties are avoided. This method will also be referred to as the local Whittle method.

When y_t is an ARFIMA(p,d,q) process, Sowell (1992) suggested to estimate the parameters by the method of maximum likelihood. The log-likelihood function is

$$L_T(\boldsymbol{y};\mu,\boldsymbol{\beta}) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\log|\Omega(\boldsymbol{\beta})| - \frac{1}{2}(\boldsymbol{y}-\mu)'\Omega^{-1}(\boldsymbol{\beta})(\boldsymbol{y}-\mu),$$

where $\boldsymbol{y} = (y_1, \ldots, y_T)'$, $\{\Omega\}_{ij} = \gamma_{|i-j|}$, $\mu = \mathrm{E}(y_t)$, $\boldsymbol{\beta}$ is a vector of parameters including d, ARMA coefficients, and unconditional variance. As the autocovariances of an ARFIMA process are complex functions of $\boldsymbol{\beta}$, calculating maximum the exact likelihood estimators (MLEs) are computationally quite demanding.

Based on the approximation proposed in Whittle (1953), maximizing the log-likelihood function is equivalent to minimizing the spectral likelihood function:

$$L_T^W(\boldsymbol{\beta}) = \sum_{j=1}^{[T/2]} \left\{ \frac{I(v_j)}{f(v_j;\boldsymbol{\beta})} + \log f(v_j;\boldsymbol{\beta}) \right\}, \tag{6}$$

where $f(v_j;\boldsymbol{\beta})$ is the spectral density of y_t and

$$I(v_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} (y_t - \mu) e^{itv_j} \right|^2,$$

see e.g., Beran (1994). The resulting minimizer is therefore an approximate to the MLE and is known as the Whittle estimator. A novel feature of $L_T^W(\boldsymbol{\beta})$ is that it does not depend on the unknown mean μ. Moreover, $L_T^W(\boldsymbol{\beta})$ is a simpler function of $\boldsymbol{\beta}$, and the resulting Whittle estimator is therefore easier to compute. The simulation results of Cheung and Diebold (1994) also demonstrate that the Whittle estimation is efficient relative to the exact MLE. A disadvantage of the Whittle estimator is that a parametric form of $f(v;\boldsymbol{\beta})$ must be specified a priori.

As far as the estimation of the fractionally differencing parameter d is concerned, Künsch (1987) and Robinson (1995b) suggested a local Whittle estimator. Robinson (1995b) showed that this method has many advantages. Importantly, it does not impose the Gaussian assumption nor does it require a correct specification of the spectral density, $f(v_j;\boldsymbol{\beta})$. This method is also asymptotically more efficient than the Whittle estimator. Observe that the spectral density of a long memory process is essentially

$$f(v) \sim G|v|^{-2d}, \tag{7}$$

as $v \to 0$, where $G \in (0, \infty)$ and $d \in (-0.5, 0.5)$. Note that this characterization focuses only on the spectral density in the neighborhood of frequency zero and ignores short-run dependence. Replacing $f(v;\boldsymbol{\beta})$ with Gv^{-2d}, we can obtain a "local" version of $L_T^W(\boldsymbol{\beta})$ which depends only on G and d but not on ARMA coefficients.

The local analogue of L_T^W in (6) is then

$$\mathcal{R}(G,d) = \frac{1}{m} \sum_{j=1}^{m} \left(\frac{I(v_j) v_j^{2d}}{G} + \log G v_j^{-2d} \right).$$

As G can be consistently estimated by

$$\hat{G}(d) = \frac{1}{m} \sum_{j=1}^{m} I(\upsilon_j)\upsilon_j^{2d}.$$

The local Whittle estimator for d is then obtained by minimizing

$$\begin{aligned} \mathcal{LW}(d) &= \mathcal{R}(\hat{G}, d) - 1 \\ &= \log\left(\frac{1}{m}\sum_{j=1}^{m} I(\upsilon_j)\upsilon_j^{2d}\right) - \frac{2d}{m}\sum_{j=1}^{m}\log \upsilon_j. \end{aligned} \tag{8}$$

The bandwidth parameter m is an integer less than $[T/2]$, but it should tend to infinity at a rate slower than T. Robinson (1995b) assumed that the bandwidth satisfies

$$\frac{1}{m} + \frac{m}{T} \to 0,$$

as $T \to \infty$. Henry and Robinson (1996) also discussed how to choose the optimal bandwidth for this semiparametric analysis of long-range dependence.

When y_t have a mean change:

$$y_t = \begin{cases} \mu_1 + \eta_t & t = 1, \ldots, k_0, \\ \mu_2 + \eta_t & t = k_0 + 1, \ldots, T, \end{cases} \tag{9}$$

where k_0 is the unknown change point. The relative location of k_0 is $\tau_0 = k_0/T$. For each hypothetical change point τ, we can estimate μ_1 and μ_2 use, respectively, the pre- and post-change observations:

$$\hat{\mu}_1([T\tau]) = \frac{1}{[T\tau]} \sum_{t=1}^{[T\tau]} y_t, \qquad \hat{\mu}_2([T\tau]) = \frac{1}{T - [T\tau]} \sum_{t=[T\tau]+1}^{T} y_t.$$

Given $[T\tau]$, the residuals are

$$\hat{\eta}_t(\tau) = \begin{cases} y_t - \hat{\mu}_1([T\tau]), & t \le [T\tau], \\ y_t - \hat{\mu}_2([T\tau]), & t > [T\tau]. \end{cases}$$

The corresponding periodogram of $\hat{\eta}_t$ is then

$$I(\upsilon_j, \tau) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} \hat{\eta}_t(\tau) e^{it\upsilon_j} \right|^2.$$

It follows from (8) that for each hypothetical change point τ, we can estimate d by minimizing

$$\mathcal{LW}(d,\tau) = \log\left(\frac{1}{m}\sum_{j=1}^{m} I(v_j,\tau)v_j^{2d}\right) - \frac{2d}{m}\sum_{j=1}^{m}\log v_j,$$

and obtain the local Whittle estimator $\tilde{d}(\tau)$. The change-point estimator $\tilde{\tau}$ is then

$$\tilde{\tau} = \operatorname{argmin}_{\tau\in[\underline{\tau},\overline{\tau}]} \mathcal{LW}(\tilde{d}(\tau),\tau),$$

where $[\underline{\tau},\overline{\tau}] \subseteq (0,1)$, and the estimator for d is now $\tilde{d}(\tilde{\tau})$. A novel feature of the proposed estimation method is that structural change and long-range dependent are both taken into account.

In Kuan and Hsu (1998), we have shown that the quasi-maximum likelihood estimator of the change point, $\hat{\tau}$, is consistent for fractionally integrated data, i.e.,

$$\Pr(|\hat{\tau} - \tau_0| > \epsilon) \to 0.$$

Recall that the spectral likelihood function is invariant with respect to the mean and that the (local) Whittle method is the frequency-domain counterpart of the maximum likelihood method. It is then reasonable to believe that the consistency of $\hat{\tau}$ carries over to the present case. Robinson (1995b) proved that, for the normalized local Whittle estimator $\hat{d}$,

$$\hat{\mathcal{H}} = 2m^{1/2}(\hat{d} - d_0) \to^d N(0,1).$$

Our simulation also confirm that asymptotic normality of $\hat{d}$ carries over to $\tilde{d}(\tilde{\tau})$. Thus,

$$\tilde{\mathcal{H}} = 2m^{1/2}(\tilde{d}(\tilde{\tau}) - d_0) \tag{10}$$

can be used as a test statistic for long-range dependence with the asymptotic standard normal distribution. This method is also readily generalized to allow for multiple breaks.

In contrast with the graphical method proposed by Teverovsky and Taqqu (1997), our method can test long-range dependence and estimate the possible change point. When the change point is properly estimated, the model can be modified to fit the data and generate more accurate forecasts. Although Teverovsky and Taqqu (1997) considered the case of multiple jumps, their graphical method needs very long data series to plot the variances on a log-log plot and obtain the estimate of d. Our method are applicable in relatively small samples.

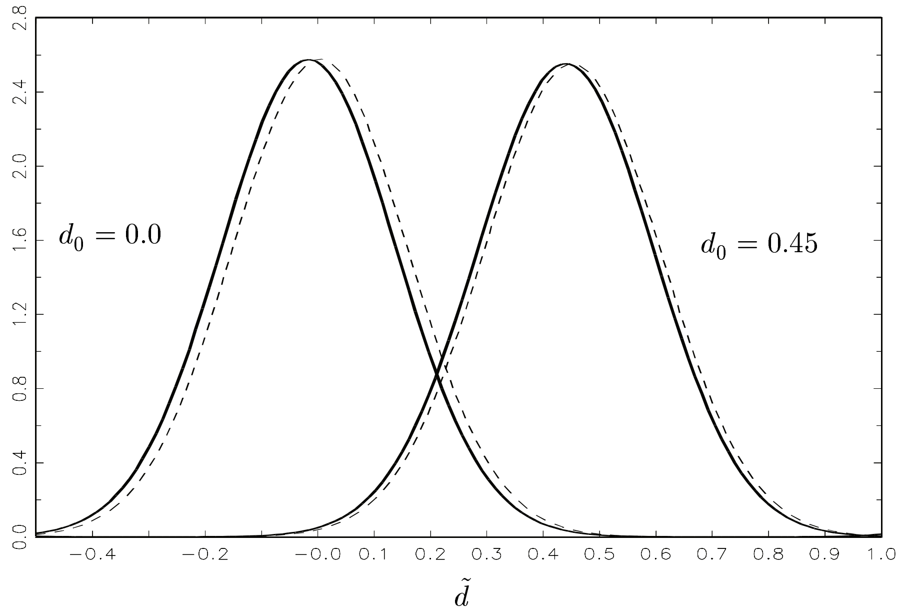

Figure 1: Compare empirical distributions of $\tilde{d}$ with normal distributions.

4 Simulations

In this section we investigate the finite sample performance of the proposed estimators for d and τ. We first use the density smoothing technique to demonstrate asymptotic normality of $\tilde{d}(\tilde{\tau})$. A changing mean DGP is specified according to (9) with $\mu_1 = 1$, $\mu_2 = 2$, and $\tau_0 = 0.5$. Figure 1 graphs the densities of $\tilde{d}(\tilde{\tau})$ for short memory ($d_0 = 0$) and long memory ($d_0 = 0.45$) data with $T = 1,028$, $m = T/4$, and 100 replications. For short-memory data, $\{\eta_t\}$ is a standard normal white noise; for long-memory data, $\{\eta_t\}$ is an ARFIMA$(0, d, 0)$ process. The smoothed densities were calculated using a Gaussian kernel with the bandwidth $h = 0.15$; see Silverman (1986, pp. 61–65). We also graph the corresponding normal densities (dashed lines) with mean equal to the true value of d and variance equal to the simulated variance of $\tilde{d}(\tilde{\tau})$.[2] Note that the finite-sample distributions of $\tilde{d}(\tilde{\tau})$ is very close to the normal distributions for both short- and long-memory data.

Table 2 presents the bias[3] of $\tilde{d}(\tilde{\tau})$ and empirical sizes of the $\tilde{\mathcal{H}}$ test under the null

[2]The sample standard deviation (0.033) for the local Whittle estimator is quite close to the theoretical value of $\sqrt{1/(4m)} = 0.031$.

[3]Here we define bias $= \frac{1}{J}\sum_{j=1}^{J} \tilde{d}_j(\tilde{\tau}) - d_0$, where $\tilde{d}_j(\tilde{\tau})$ is the estimator $\tilde{d}(\tilde{\tau})$ in the jth replication,

Table 2: Empirical size of the modified local Whittle method

$\tau_0 = 0.5$						
Significance Level	$T = 128$			$T = 256$		
	$m = T/8$	$m = T/4$	$m = T/3$	$m = T/8$	$m = T/4$	$m = T/3$
Size 1%-Test	0.9	0.5	0.7	0.5	0.5	0.6
Size 2.5%-Test	1.3	1.3	1.0	1.4	0.9	1.3
Size 5%-Test	2.1	2.0	2.1	2.5	2.2	2.1
Size 10%-Test	3.0	3.9	3.4	4.1	4.2	4.4
Bias	−0.330	−0.142	−0.107	−0.138	−0.067	−0.050
$\tau_0 = 0.3$						
Size 1%-Test	1.0	0.4	0.5	0.6	0.8	0.3
Size 2.5%-Test	1.8	0.9	0.8	1.1	1.2	0.7
Size 5%-Test	2.3	1.4	2.0	1.8	2.3	1.8
Size 10%-Test	3.3	3.0	3.4	3.3	4.6	4.2
Bias	−0.330	−0.136	−0.104	−0.138	−0.064	−0.049

hypothesis of short memory ($d = 0$) with an one-time break, i.e., (9) with $\mu_1 = 1$, $\mu_2 = 2$, and $\tau_0 = 0.5$. In this simulation we chose two different sample sizes, $T = 128$ and $T = 256$, and for each sample, three values of m were considered: $T/8$, $T/4$, and $T/3$. The number of replications for each experiment is $1,000$. Unlike the R/S and GPH tests, Table 2 shows that the empirical sizes of the $\tilde{\mathcal{H}}$ test are lower than the nominal size in all cases considered. This low size may be explained by the negative bias which may be resulted from the bandwidth chosen or the change-point estimate. If we increase the number of observations and choose an appropriate bandwidth, the negative bias becomes smaller. For $\tau_0 = 0.3$, $\tilde{\mathcal{H}}$ has similar empirical sizes and biases, as shown in Table 2.

Table 3 reports the power of the local Whittle $\tilde{\mathcal{H}}$ test against the alternative of long-range dependence with a change point at $\tau_0 = 0.5$. We follow McLeod and Hipel (1978) and Hosking (1984) to generate η_t as $I(d)$ series in (9). For sample size $T = 256$ and $m = T/8$, the power of the $\tilde{\mathcal{H}}$ test increases from 22% for $d = 0.15$ to about 91% for $d = 0.45$. If we change the bandwidth to $m = T/4$, the power improves from

and J is the number of replications.

Table 3: Empirical power of the modified local Whittle method

Power 5%-Test ($T = 256$)	$d = 0.15$	$d = 0.25$	$d = 0.35$	$d = 0.45$
$m = T/8$	22.0	48.5	73.1	90.5
Bias	−0.118	−0.114	−0.112	−0.106
$m = T/4$	45.5	83.8	98.2	99.8
Bias	−0.062	−0.065	−0.055	−0.054

approximately 46% for $d = 0.15$ to about 100% for $d = 0.45$. These simulation results show that the power of the $\tilde{\mathcal{H}}$ test increases with d. As an additional result, we consider the case that the DGP has long memory ($d = 0.35$) and contains no change, we find that the power of our approach is about 97.3% under $T = 256$ and $m = T/4$. This shows that the $\tilde{\mathcal{H}}$ test does have good power against long-range dependence.

For change-point estimation, we consider the DGP (9) with $\mu_1 = 1$, $\mu_2 = 2$, $\tau_0 = 0.5$, and $\{\eta_t\}$ is an $I(d)$ series with $d = 0.35$. In our simulations, $T = 128$, the bandwidth $m = T/4$, and the number of replications is $5,000$. Figure 2(a) shows the empirical distribution of $\tilde{k}$ ($\tilde{\tau} = \tilde{k}/T$) on the interval $[\,[0.1 \times T], [0.9 \times T]\,]$. It is clear that $\tilde{k}$ is concentrated at the true change point. If we increase the sample size (e.g., $T = 256$), the precision of $\tilde{k}$ also improves. When there is no change so that $\mu_1 = \mu_2 = 1$, the empirical distribution of $\tilde{k}$ is presented in Figure 2(b). It can be seen that $\tilde{k}$ are more concentrated at two end points ($[0.1 \times T]$ and $[0.9 \times T]$). In fact, this can also be interpreted as a consistency result. In contrast with Kuan and Hsu (1998), our estimator can locate the change point correctly even when data have long memory. Unlike the least-squares estimator of the change point, there is no “spurious change” problem with the local Whittle estimator.

Figure 2(c) and (d) contain the histograms of $\tilde{k}$ for $d = 0$ when there is an one-time change or none, respectively. The proposed estimator still works in the short memory case. The relative frequency at $\tilde{k} = k_0$ is close to that for the least-squares estimator in Kuan and Hsu (1998). To summarize, our method is quite successful in handling long-range dependence and structural change.

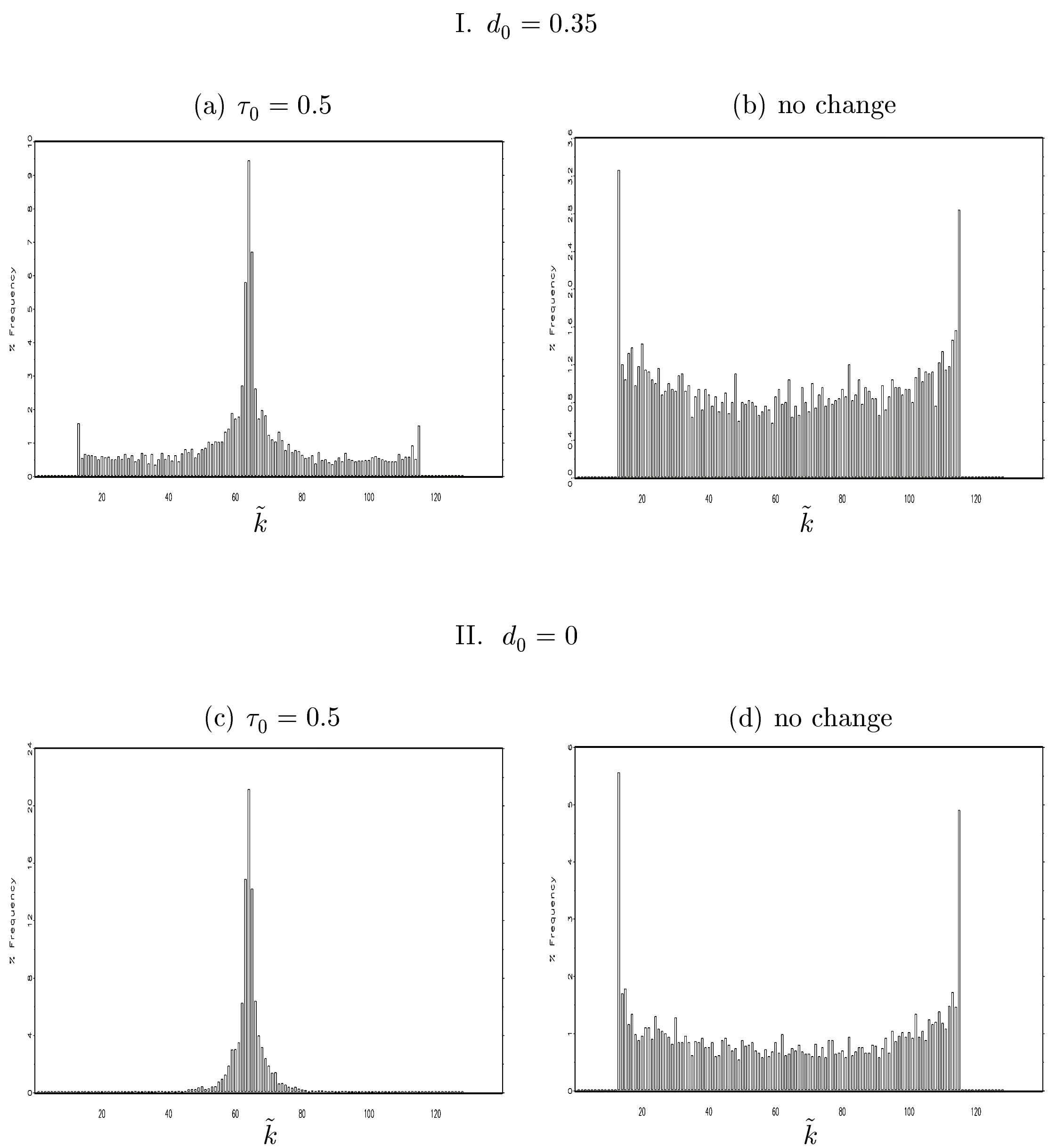

Figure 2: Empirical distributions of $\tilde{k}$.

5 Empirical Studies of G7 Inflation Rates

An important issue in macroeconomics is how aggregate prices respond to monetary and energy shocks. Hassler and Wolters (1995) and Baillie, Chung, and Tieslau (1996) suggested that inflation rates are better described by ARFIMA processes rather than unit-root processes for most G7 countries. However, in Section 2 we have shown that most long memory tests may reject the null hypothesis of short memory too often when a structural change is present. This is closely related to the problem that a trend-stationary series with a break may be characterized as a unit-root process, as noted in Perron (1989). Hence, one should be careful to explain the persistent behavior of inflation rates. Bos, Franses, and Ooms (1999) examined the impact of mean shifts on the estimates of ARFIMA parameters. However, they assumed the break dates are fixed exogenously at two oil shocks. The purpose of our empirical studies is to investigate the sensitivity of long memory tests to exogeneity assumption concerning break dates.

In this section, we apply the local Whittle method to study monthly inflation rates in G7 countries from January 1957 through December 1998, which contain a total of 504 observations. The dataset of the Consumer Price Indices (CPI, p_t) of G7 countries are taken from the International Financial Statistics (IFS). Inflation rates are computed from the price indices by taking $\pi_t = 100 \cdot \log(p_t/p_{t-1})$. For each series, we eliminate the seasonal effect by dummy variables. Figure 3 shows the raw data with seasonal adjustment. As the inflation rates exhibit rather erratic behavior in the first years of the sample, we only use the data starting in 1958. Most sample autocorrelations of π_t exhibit the clear pattern of slow decay and are quite large even after long lags (not reported). We first consider the traditional long memory tests as a benchmark to compare with the proposed test. Table 4 gives the results of R/S and GPH tests. The R/S tests with different q suggest that the inflation rates in all G7 countries have long memory at conventional 5% level. Similarly, the GPH test also shows that $\hat{d}$ is significantly different from zero. Hence, these results suggest that the inflation rates may have long memory and strong persistence.

We also use the conditional sum of squares (CSS) method, proposed by Baillie, Chung, and Tieslau (1996), to estimate an array of ARFIMA(p, d, q) models with different p and q;

$$\phi(B)(1-B)^d(\pi_t - \mu) = \theta(B)v_t.$$

The ARFIMA(1,d,0) model appears to capture the correlation behavior of inflation

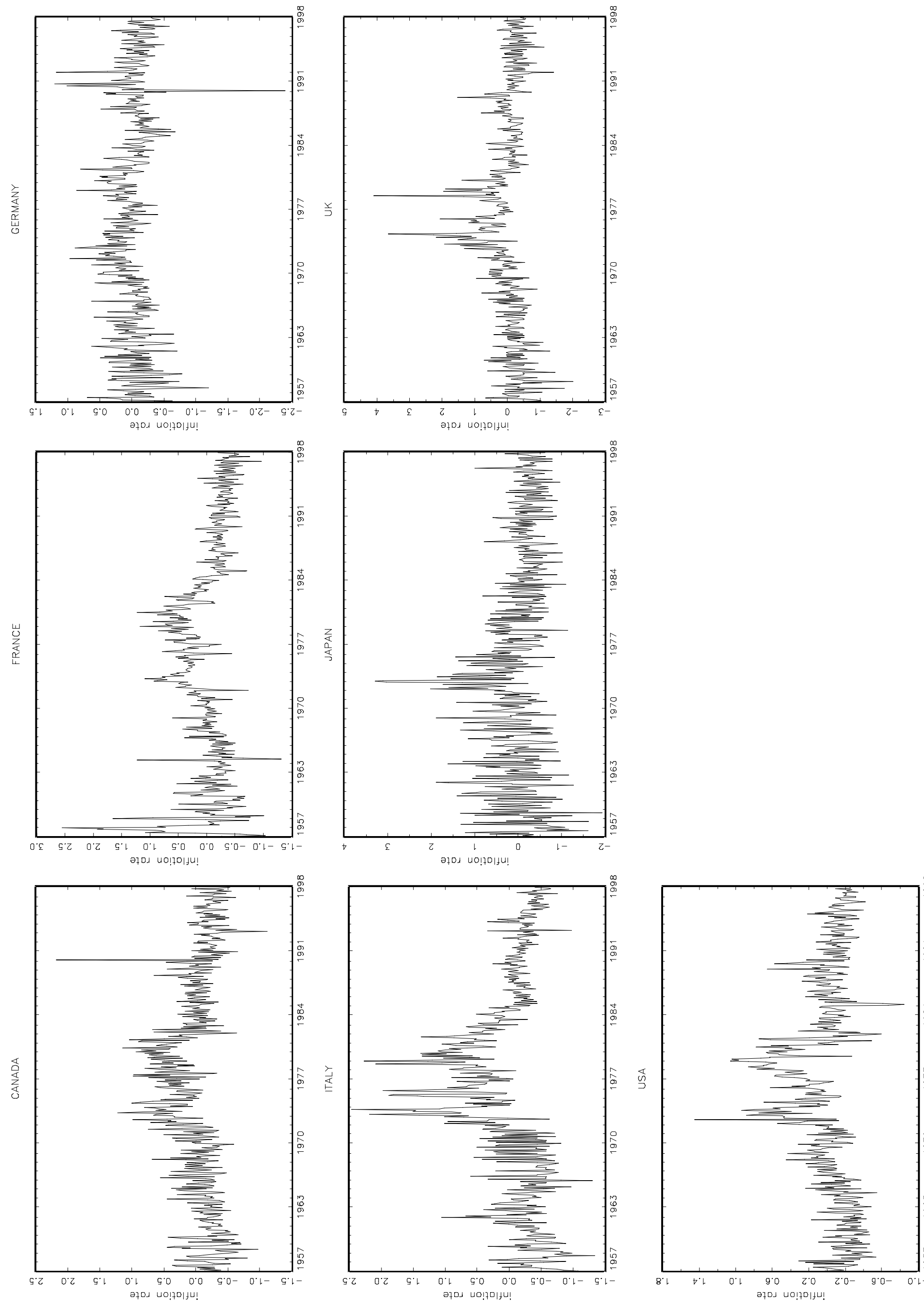

Figure 3: G7 inflation rates: 1957–1998 (seasonally adjusted).

Table 4: Long-memory tests for monthly G7 inflation rates: 1958–1998.

R/S test							
$V(T,q)$	Canada	France	Germany	Italy	Japan	UK	USA
$q=0$	6.040	6.404	3.824	7.415	4.114	5.600	6.317
$q=10$	2.629	2.627	2.317	2.761	2.328	2.429	2.420
$q=20$	1.980	1.981	1.829	2.088	1.828	1.874	1.839
q^*	2.738	2.289	2.612	2.231	2.864	2.347	2.005
	(9)	(14)	(6)	(17)	(5)	(11)	(16)
GPH test							
$\hat{d}$	Canada	France	Germany	Italy	Japan	UK	USA
$c=0.55$	0.731	0.637	0.497	0.665	0.687	0.513	0.884
	(5.701)	(3.260)	(4.135)	(5.083)	(5.783)	(3.320)	(6.147)
$c=0.65$	0.443	0.343	0.222	0.572	0.406	0.552	0.533
	(4.633)	(2.656)	(2.389)	(6.149)	(4.286)	(5.113)	(4.841)
$c=0.75$	0.336	0.332	0.281	0.479	0.275	0.381	0.421
	(4.892)	(4.005)	(3.749)	(7.284)	(3.749)	(4.993)	(5.900)
$m=[T/2]$	0.227	0.316	0.225	0.412	0.192	0.335	0.446
	(4.651)	(6.161)	(4.477)	(8.198)	(3.890)	(6.216)	(8.559)

Note: 1%, 5%, and 10% critical value for the R/S test is 2.001, 1.747, and 1.620, respectively. t-values are in parentheses for the GPH test.

Table 5: Estimated ARFIMA model for inflation rates in G7 countries.

	ARFIMA(1,d,0)						
Parameter	Canada	France	Germany	Italy	Japan	UK	USA
$\hat{\mu}$	-0.052	-0.221	-0.014	-0.175	0.009	-0.106	-0.086
	(-0.448)	(-1.366)	(-0.265)	(-0.622)	(0.077)	(-0.540)	(-0.524)
$\hat{d}$	0.391	0.415	0.245	0.495	0.285	0.391	0.497
	(10.05)	(9.151)	(4.735)	(11.41)	(6.711)	(8.828)	(10.53)
$\hat{\phi}_1$	-0.221	-0.022	-0.026	-0.150	-0.181	-0.107	-2.186
	(-4.016)	(-0.343)	(-0.388)	(-2.544)	(-3.115)	(-1.767)	(15.68)
$\hat{\sigma}_v^2$	0.093	0.082	0.083	0.135	0.342	0.218	0.044
	(26.71)	(32.96)	(35.59)	(21.61)	(19.82)	(29.15)	(27.54)
$\log \mathcal{L}$	-113.4	-81.82	-85.66	-205.2	-434.1	-322.9	70.43

Note: t-values are in parentheses.

rates quite well. In Table 5 the estimated fractionally differencing parameters are in the range of $0 < d < 0.5$ for all countries. Like the estimates obtained by GPH method, the parametric approach also provide strong evidence of long memory in inflation. Although these findings are in accordance with the results of Hassler and Wolters (1995) and Baillie, Chung, and Tieslau (1996), the traditional tests do not take into account the influence of regime switch on inflation. Therefore, we must apply the proposed $\tilde{\mathcal{H}}$ test instead of traditional tests for long memory.

In Table 6, the results of the local Whittle estimation are presented. We first consider the case that there is only a one-time change. For the proposed method, the estimates of $\tilde{d}$ with $m = T/4$ are still significant at the 5% level. That is, the long-memory effect on inflation is robust to a one-time change. Comparing with the results of the GPH and ARFIMA estimation, the estimates of $\tilde{d}$ are quite small in Germany and Japan after a level shift is allowed. On the other hand, our method suggests that the break date is in the period of oil crisis for Germany, Italy, Japan, and USA.

The second panel of Table 6 concerns the case of two breaks. It can be seen that allowing for two breaks has a huge effect on the estimates of fractional integration. For Germany and Japan, the long-range dependence on inflation is no longer significant. The

Table 6: The proposed test for monthly G7 inflation rates: 1958–1998.

One Break							
	Canada	France	Germany	Italy	Japan	UK	USA
$[T\tilde{\tau}]$	1991:02	1959:02	1982:07	1973:10	1981:05	1967:09	1982:07
$\tilde{d}$	0.348	0.416	0.191	0.473	0.092	0.308	0.462
$\tilde{\mathcal{H}}$	7.716*	9.230*	4.227*	10.49*	2.092*	6.828*	10.25*
Two Breaks							
	Canada	France	Germany	Italy	Japan	UK	USA
$[T\tilde{\tau}_1]$	1972:11	1973:03	1969:10	1973:10	1973:01	1973:08	1973:01
$[T\tilde{\tau}_2]$	1982:06	1985:05	1982:07	1984:03	1974:11	1981:05	1981:09
$\tilde{d}$	0.180	0.179	0.086	0.313	0.015	0.167	0.316
$\tilde{\mathcal{H}}$	3.985*	3.966*	1.914	6.940*	0.337	3.700*	6.998*

Note: $*$ indicates significant at the 5% level.

estimates of $\tilde{d}$ for most G7 countries are less than 0.2 except for Italy and USA. Hence, the persistence on inflation may be overestimated if we do not account for structural changes. In the meantime, change-point estimates can correctly locate on dates of two oil shocks for most G7 countries. In contrast with Bos, Franses, and Ooms (1999), our break dates are estimated by data and the estimates of $\tilde{d}$ are smaller than theirs.

6 Conclusions

In this paper we focus on how to test for long-range dependence when the process has an one-time mean change and how to estimate the change point when data are long-range dependent. We first pointed out that traditional long-memory tests have serious size distortions when the data are short memory with changes. In order to overcome this shortcoming and the spurious change problem discussed in Kuan and Hsu (1998), a local Whittle method is discussed. Simulation results confirm that our change-point estimator is well behaved even when data are long-range dependent, and that our test for long memory maintains proper size when a change is present. These results indicate that our method is practically useful and has a much wider applicability.

There are two directions for future research. First, our method may also be appropriate for multiple breaks by estimating break dates simultaneously; cf. Bai and Perron (1997). Second, it is interesting to develop a procedure for testing the structural change in a long-memory environment. Intuitively, we can modify the method proposed by Hidalgo and Robinson (1996), using the change-point estimate instead of a known change. These issues are currently under study.

Appendix

Proof of Theorem 2.1: When $\nu = 1/2$, substitute (1) and (5) into $V(T,q)$, we have

$$\begin{aligned} & \frac{1}{\tilde{s}_T(q)\sqrt{T}}\left(Y(t) - \frac{t}{T}Y(T)\right) \\ = \quad & \frac{1}{\tilde{s}_T(q)\sqrt{T}}\left(\sum_{i=1}^{t}\varepsilon_i - \frac{t}{T}\sum_{i=1}^{T}\varepsilon_i\right) + \frac{1}{\tilde{s}_T(q)T}\left(\sum_{i=1}^{t}\psi(\frac{i}{T}) - \frac{t}{T}\sum_{i=1}^{T}\psi(\frac{i}{T})\right). \quad (11) \end{aligned}$$

Given the condition [A], the functional central limit theorem ensures that

$$\left(\frac{1}{\tilde{s}_T(q)\sqrt{T}}\sum_{i=1}^{[T\tau]}\varepsilon_i, \quad 0 \le \tau \le 1\right) \Rightarrow \left(B_0(\tau), \quad 0 \le \tau \le 1\right),$$

where $[T\tau]$ denotes the integer part of $T\tau$ and B_0 is the standard Brownian motion. Hence the first term weakly converges to the Brownian bridge,

$$\left(\frac{1}{\tilde{s}_T(q)\sqrt{T}}\left(\sum_{i=1}^{[T\tau]}\varepsilon_i - \frac{[T\tau]}{T}\sum_{i=1}^{T}\varepsilon_i\right), \quad 0 \le \tau \le 1\right) \Rightarrow \left(B^0(\tau), \quad 0 \le \tau \le 1\right).$$

Also,

$$\frac{1}{T}\sum_{i=1}^{[T\tau]}\psi(\frac{i}{T}) \to \int_0^{\tau}\psi(z)dz,$$

uniformly in τ. It follows from the continuous mapping theorem that

$$\frac{1}{\tilde{s}_T(q)\sqrt{T}}\max_{1\le t\le T}\left(Y(t) - \frac{t}{T}Y(T)\right) \Rightarrow \max_{0\le\tau\le 1}\left(B^0(\tau) + \frac{1}{\sigma}\Psi(\tau)\right).$$

Similarly,

$$\frac{1}{\tilde{s}_T(q)\sqrt{T}}\min_{1\le t\le T}\left(Y(t) - \frac{t}{T}Y(T)\right) \Rightarrow \min_{0\le\tau\le 1}\left(B^0(\tau) + \frac{1}{\sigma}\Psi(\tau)\right).$$

Consequently,

$$\begin{aligned} V(T,q) \quad & \Rightarrow \quad \text{range}_{0\le\tau\le 1}\left\{B^0(\tau) + \frac{1}{\sigma}\left(\int_0^{\tau}\psi(z)dz - \tau\int_0^1\psi(z)dz\right)\right\} \\ & \equiv \quad \text{range}_{0\le\tau\le 1}\left(B^0(\tau) + \frac{1}{\sigma}\Psi(\tau)\right). \end{aligned}$$

For $\nu = 0$, $\tilde{s}_T^2(q)$ is not a consistent estimator of σ^2; the limit of $\tilde{s}_T^2(q)$ depends on the change function ψ. Using the fact that $\sum_{i=1}^{[T\tau]} \varepsilon_t = O_p(T^{1/2})$ and the assumption $\tilde{s}_T^2(q)$ is $O_p(1)$,

$$T^{-1/2}V(T,q) \quad = \quad \text{range}_{0\leq\tau\leq 1}\left\{\frac{1}{\tilde{s}_T(q)}\left(\int_0^\tau \psi(z)dz - \tau\int_0^1 \psi(z)dz\right) + o_p(1)\right\},$$

which is $O_p(1)$. This implies $V(T,q)$ diverges in probability. □

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, **59**, 817–858.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes, *Econometrica*, **66**, 47–78.

Baillie, R. T., C.-F. Chung, & M. A. Teslau (1996). Analysing inflation by the fractionally integrated ARFIMA-GARCH model, *Journal of Applied Econometrics*, **11**, 23–40.

Beran, J. (1994). *Statistics for Long-Memory Processes*, Monographs on Statistics and Applied Probability, **61**, Chapman & Hall, New York.

Bos, C. S., P. H. Franses, & M. Ooms (1999). Long memory and level shifts: Reanalyzing inflation rates, *Empirical Economics*, **24**, 427–449.

Cheung, Y.-W. & F. X. Diebold (1994). On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean, *Journal of Econometrics*, **62**, 301–316.

Chung, C.-F. (1994). A note on calculating the autocovariances of the fractionally integrated ARMA models, *Economics Letters*, **45**, 293–297.

Geweke, J. & S. Porter-Hudak (1983). The estimation and application of long memory time series models, *Journal of Time Series Analysis*, **4**, 221–237.

Hassler, U. & J. Wolters (1995). Long memory in inflation rates: International evidence, *Journal of Business and Economic Statistics*, **13**, 37–45.

Henry, M. & P. M. Robinson (1996). Bandwidth choice in Gaussian semiparametric estimation of long range dependence, in P. M. Robinson & M. Rosenblatt (eds.), *Proceedings of the Athens Conference on Applied Probability and Time Series*, Springer-Verlag, New York, 220–232.

Hidalgo, J. & P. M. Robinson (1996). Testing for structural change in a long-memory environment, *Journal of Econometrics*, **70**, 159–174.

Hosking, J. R. M. (1984). Modelling persistence in hydrological time series using fractional differencing, *Water Resources Research*, **20**, 1898–1908.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs, *Transactions of the American Society of Civil Engineers*, **116**, 770–799.

Klemeš, V. (1974). The Hurst phenomenon: a puzzle? *Water Resources Research*, **10**, 675–688.

Kuan, C.-M. and C.-C. Hsu (1998). Change-point estimation of fractionally integrated processes, *Journal of Time Series Analysis*, **19**, 693–708.

Künsch, H. R. (1987). Statistical aspects of self-similar processes, in Yu. Prohorov and V. V. Sazanov (eds.), *Proceedings of the First World Congress of the Bernoulli Society*, **1**, VNU Science, Utrecht, 67–74.

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

Mandelbrot, B. B. (1975). Limit theorems on the self-normalized range for weakly and strongly dependent processes, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **31**, 271–285.

Mandelbrot, B. B. & M. Taqqu (1979). Robust R/S analysis of long-run serial correlation, *Bulletin of the International Statistical Institute*, **48**, Book 2, 59–104.

Mandelbrot, B. B. & J. Wallis (1969). Computer experiments with fractional Gaussian noises, Part 1,2,3, *Water Resources Research*, **5**, 228–267.

McLeod, A. I. & K. W. Hipel (1978). Preservation of the rescaled adjusted range 1. A reassessment of the Hurst phenomenon, *Water Resources Research*, **14**, 491–508.

Newey, W. K. & K. D. West (1987). A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, **55**, 703–708.

Perron, P. (1989). The great crash, the oil price shock and the unit root hypothesis. *Econometrica*, **57**, 1361–1401.

Robinson, P. M. (1995a). Log-periodogram regression of times series with long range dependence, *Annals of Statistics*, **23**, 1048–1072.

Robinson, P. M. (1995b). Gaussian semiparametric estimation of long range dependence, *Annals of Statistics*, **23**, 1630–1661.

Silverman, B. W. (1986). *Density estimation for statistics and analysis*, Chapman & Hall, New York.

Sowell, F. B. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics*, **53**, 165–188.

Teverovsky, V. & M. Taqqu (1997). Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator, *Journal of Time Series Analysis*, **18**, 279–304.

Whittle, P. (1953). Estimation and information in stationary time series, *Arkiv for Matematik*, **2**, 423–434.