# A Test Against Spurious Long Memory

**Zhongjun Qu**[*]

**Boston University**

July 15, 2010

**Abstract**

This paper proposes a test statistic for the null hypothesis that a given time series is a stationary long memory process against the alternative hypothesis that it is affected by regime change or a smoothly varying trend. The proposed test is in the frequency domain and is based on the derivatives of the profiled local Whittle likelihood function in a degenerating neighborhood of the origin. The assumptions used are mild, allowing for non-Gaussianity or conditional heteroskedasticity. The resulting null limiting distribution is nuisance parameter free and can be easily simulated. Furthermore, the test is straightforward to implement. In particular, it does not require one to specify the form of the trend or the number of different regimes under the alternative hypothesis. Monte Carlo simulation shows that the test has decent size and power properties. The paper also considers three empirical applications to illustrate the usefulness of the test.

**JEL Classification Number**: C12, C22.

**Keywords**: fractional integration, frequency domain estimates, semiparametric, structural change, trend.

# 1 Introduction

A scalar process is said to have long memory if its spectral density at frequency $\lambda$ is proportional to $\lambda^{-2d}$ ($d \neq 0$) as $\lambda$ approaches zero. Long memory models provide a middle ground between short memory and unit root specifications, hence allowing for more flexibility in modelling the persistence of a time series. Estimation theory has been worked out. Among others, Fox and Taqqu (1986) studied parametric long memory models and Geweke and Porter-Hudak (1983) and Robinson (1995a,b) considered semiparametric models. The latter literature has been very influential because it does not require specific parametric assumptions on the process generating the difference of order $d$ of the series. Applications to macroeconomics and finance are numerous. For example, Ding et al. (1993) argued that stock returns volatility is well described by a long memory process. Also see Andersen et al. (2001), Andersen et al. (2003) and Deo et al. (2006) for other applications.

Two important features of a long memory process are that its spectral density at the origin is unbounded and that its autocorrelation function decays at a hyperbolic rate at long lags. However, these features can also be present for a short memory process affected by regime change or a smooth trend, leading to so-called spurious long memory. This has been widely documented, see Perron and Qu (2010) and references therein. Whether the observed long memory characteristics are actually genuine is of substantial empirical importance. For example, Taylor (2000) showed that the long memory assumption has a significant impact on the term structure of implied volatilities. Ohanissian et al. (2004) showed that when the true data generating process (DGP) is of spurious long memory, using either a short memory model or a pure long memory model leads to severe underpricing of call options. Unfortunately, it is rather difficult to distinguish between these two types of processes, partly because tests for structural change are often biased toward rejection (of the null hypothesis of no change) when the process is indeed fractionally integrated.

This paper proposes a test statistic to distinguish between true and spurious long memory. Its construction is in the frequency domain and it is based on the observation that the aforementioned

processes exhibit different properties over different frequency bands local to zero. The test is of the Kolmogorov-Smirnov type and is very simple to implement. In particular, it does not require specifying the form of the trend or the number or locations of the different regimes that occur under the alternative hypothesis. We derive its null limiting distribution under two alternative sets of conditions. The first is the same as Robinson (1995b), allowing for non-Gaussianity but not heteroskedasticity. The second is due to Shao and Wu (2007), which allows for both features but imposes stronger conditions on the spectral density of the short memory component and the bandwidth parameter for local Whittle estimation. We prove that the test is consistent against alternatives of interest. We also provide a simple pre-whitening procedure to control its size when significant short memory dynamics are present. The test is applied to three time series commonly studied in the long memory literature and the null hypothesis of stationary long memory is rejected at a 5% significance level for two of these series. The result, although limited, suggests that the evidence for stationary long memory may not be as strong as is often perceived.

This paper is closely related to Ohanissian et al. (2008), who explored the idea that temporal aggregation does not change the order of fractional integration. Their result requires Gaussianity. Also, the assumption they place on the bandwidth of the GPH estimates is very stringent and requires a very large sample size for the test to be useful. Another closely related work is Perron and Qu (2010), who studied spectral domain properties of a short memory process contaminated by occasional level shifts. Their results inspired the test statistic proposed in this paper. Our work is also related to Müller and Watson (2008), who discussed a general framework for testing for low frequency variability. Other related works include Sibbertsen and Venetis (2003), Berkes et al. (2006) and Giraitis et al. (2006).

From a methodological perspective, the idea of using integrated periodograms as the basis for specification testing can be traced back to Grenander and Rosenblatt (1953, 1957) and Bartlett (1955). The literature concerning long memory processes is relatively sparse with Ibragimov (1963) being a seminal contribution. Recently, Kokoszka and Mikosch (1997) derived a functional cen-

tral limit theorem for the integrated periodogram of a long memory process with finite or infi-
nite variance. They only considered parametric models and the parameters were assumed to be
known. Nielsen (2004) considered semiparametric models for multivariate long memory processes
and proved the weak convergence of the integrated periodogram. His results greatly facilitate our
subsequent analysis. However, he also assumed the memory parameter is known and his result only
applies to $d \in (0, 1/4)$. This paper obtains a weak convergence result for weighted periodograms
that involve estimated parameters in a semiparametric setting. This result is of independent interest
and can be used to construct other specification tests for semiparametric long memory models.

The remainder of this paper is organized as follows. Section 2 discusses the hypotheses of
interest. Section 3 introduces the test statistic. Section 4 discusses the assumptions, null limiting
distribution and consistency of the test. Section 5 introduces a simple procedure to control the
size. Section 6 includes simulation results to assess finite sample properties. Section 7 considers
three empirical applications and Section 8 concludes. The proofs are contained in two appendices,
main and supplementary, with the main appendix containing proofs of the main results. The
supplementary appendix is available from the author's web page.

The following notation is used. The subscript 0 indicates the true value of a parameter. $|z|$
denotes the modulus of $z$; the imaginary unit is denoted by $i$. $[x]$ denotes the integer part of a
real number $x$. For a real-valued random variable $\xi$, write $\xi \in \mathcal{L}^p$ if $\|\xi\|_p = (E |\xi|^p)^{1/p} < \infty$ and
$\|\xi\| = \|\xi\|_2$. "$\Rightarrow$" and "$\rightarrow^p$" signify weak convergence under the Skorohod topology and convergence
in probability. And $O_p(\cdot)$ and $o_p(\cdot)$ are the usual symbols for stochastic orders of magnitude.

## 2 The hypotheses of interest

Let $x_t$ $(t = 1, 2, ..., n)$ be a scalar process with $n$ the sample size. Let $f(\lambda)$ denote its spectral
density at frequency $\lambda$. Then the null hypothesis is:

• $\mathbf{H}_0 : x_t$ is stationary with

$$f(\lambda) \simeq G\lambda^{-2d} \text{ as } \lambda \to 0+ \text{ with } d \in (-1/2, 1/2) \text{ and } G \in (0, \infty), \qquad (1)$$

where "$\simeq$" means that the ratio of expressions on the left and right sides tends to unity.

A special case of a process satisfying (1) is the $ARFIMA(p, d, q)$ process, introduced by Granger and Joyeux (1980) and Hosking (1981):

$$A(L) (1 - L)^d x_t = B(L)\varepsilon_t, \tag{2}$$

where $A(L) = 1 - a_1 L - ... - a_p L^p$, $B(L) = 1 + b_1 L + ... + b_q L^q$ and $\varepsilon_t$ is a white noise process with $E(\varepsilon_t^2) = \sigma_\varepsilon^2$. The spectral density of (2) satisfies

$$f(\lambda) \simeq \frac{\sigma_\varepsilon^2}{2\pi} \frac{|B(1)|^2}{|A(1)|^2} \lambda^{-2d}$$

as $\lambda \to 0+$.

**Remark 1** *Under the null hypothesis, the behavior of the spectral density is specified only in a neighborhood of the zero frequency. This allows one to minimize possible complications arising from misspecifying the higher frequency dynamics.*

The periodogram of $x_t$ evaluated at frequency $\lambda_j = 2\pi j/n$ $(j = 1, 2, ..., [n/2])$ is given by $I_x(\lambda_j) = (2\pi n)^{-1} \left| \sum_{t=1}^n x_t \exp(i\lambda_j t) \right|^2$. Its statistical properties have been extensively studied, see Robinson (1995a, b). Theorem 2 in Robinson (1995a) is central to our analysis. It states that when $x_t$ satisfies (1), then under fairly mild conditions,

$$E\left(I_x(\lambda_j)/f(\lambda_j)\right) \to 1 \text{ for all } j \text{ such that } j \to \infty \text{ as } n \to \infty \text{ but } j/n \to 0.$$

This uniform behavior of $I_x(\lambda_j)$ is the very property that allows us to detect spurious long memory.

Under the alternative hypothesis, the process $x_t$ has short memory but is contaminated by level shifts or a smooth trend. This is often referred to as spurious long memory because the estimate of $d$ is biased away from zero and the autocovariance function exhibits a slow rate of decay, akin to a long memory process. In the following, we will compare spectral domain properties of processes with true and spurious long memory to motivate the construction of the test statistic.

4

First, consider short memory processes contaminated by level shifts. Perron and Qu (2010) considered the following model involving random level shifts for some series $x_t$:

$$x_t = z_t + \mu_t \text{ with } \mu_t = \mu_{t-1} + \pi_t \eta_t, \tag{3}$$

where $z_t$ is a stationary short memory process (e.g., an ARMA process), $\eta_t \sim i.i.d.$ $(0, \sigma_\eta^2)$ and $\pi_t$ is a Bernoulli random variable that takes value 1 with probability $p_n$ and 0 otherwise, i.e., $\pi_t \sim i.i.d.$ $B(1, p_n)$. They used $p_n = p/n$ with $0 < p < \infty$ to model rare shifts so that as $n$ increases, the expected number of shifts remains bounded. $\pi_t$, $\eta_t$ and $z_t$ are assumed to be mutually independent. Then, the periodogram of $x_t$ can be decomposed into the following three components:

$$
\begin{aligned}
I_x(\lambda_j) &= \frac{1}{2\pi n}\left|\sum_{t=1}^{n} z_t \exp(i\lambda_j t)\right|^2 + \frac{1}{2\pi n}\left|\sum_{t=1}^{n} \mu_t \exp(i\lambda_j t)\right|^2 + \frac{2}{2\pi n}\sum_{t=1}^{n}\sum_{s=1}^{n} z_t \mu_t \cos(\lambda_j(t-s)) \\
&= (\text{I}) + (\text{II}) + (\text{III}).
\end{aligned}
$$

They showed that for $\lambda_j = o(1)$, the orders of the three terms satisfy

$$(\text{I}) = O_p(1), \ (\text{II}) = O_p(n^{-1}\lambda_j^{-2}) \text{ and } (\text{III}) = O_p(n^{-1/2}\lambda_j^{-1}). \tag{4}$$

The orders are exact, see Proposition 3 in Perron and Qu (2010). Term (II) dominates (I) and (III) if $j = o(n^{1/2})$ and term (I) dominates (II) and (III) if $jn^{-1/2} \to \infty$. Hence, the level shift component affects the periodogram only up to $j = O(n^{1/2})$. Within this narrow range, for a fixed sample the slope of the log-periodogram is on average $-2$, implying $d = 1$. This is different from a long memory process defined generally by (1), or particularly by (2), where the slope of the log-periodogram is $-2d > -1$ for $j = o(n)$. Thus, if we use a local method to estimate the memory parameter of the level shift model (3), then the estimate will crucially depend on the number of frequencies used. It will tend to decrease as the number of frequencies used increases. More importantly, this non-uniform behavior in the neighborhood of $j = n^{1/2}$ suggests the possibility of distinguishing between the two processes from the spectral domain.

Now, consider a short memory process containing a smoothly varying trend:

$$x_t = h(t/n) + z_t, \tag{5}$$

5

where $h(s)$ is a Lipschitz continuous function on $[0,1]$ (i.e., there exists a $K < \infty$ such that $|h(s) - h(\tau)| \leq K|s - \tau|$ for all $s, \tau \in [0,1]$), $h(s) \neq h(\tau)$ for some $s \neq \tau$, and $z_t$ is defined as in (3).

**Lemma 1** *If $x_t$ is generated by (5), then $I_x(\lambda_j) = O_p(n^{-1}\lambda_j^{-2})$ for $j = O(n^{1/2})$ and $I_x(\lambda_j) = O_p(1)$ for $j$ satisfying $jn^{-1/2} \to \infty$.*

The Lemma extends Lemma 2 of Künsch (1986), who obtained the same result assuming $h(t/n)$ is monotonic but not necessarily smooth. Thus, as in the level shift model, the effect of the trend is visible only within a narrow range of frequencies and the threshold $j = O(n^{1/2})$ is of crucial importance. This result is weaker than the case with level shifts because the orders are not necessarily exact. However, it is sufficient to motivate the test statistic because, for the test to have power, what matters is not the order of $I_x(\lambda_j)$ *per se*, but the nonuniformity of its slope in a neighborhood of the zero frequency.

We now illustrate the asymptotic results using simulations. The sample size is $n = 1000$ and the results reported are empirical means based on 5000 replications. First, consider the level shift process (3), with $z_t, \eta_t \sim i.i.d.N(0,1)$ and $\pi_t \sim i.i.d.B(1, 5/n)$. Figure 1(a) depicts the log periodograms as a function of the frequency index. An extra line, $\log(0.0015\lambda_j^{-2})$, is superimposed to highlight the slope of the periodogram when $j = o(n^{1/2})$. The figure is very informative: for very low frequencies, the slope of the log periodogram is approximately $-2$, implying $d = 1$; for $j > n^{1/2}$, the slope is basically zero. Thus, nonuniformity is clearly present. We now impose the false restriction that the series are of true long memory and estimate $f(\lambda)$ using the local Whittle likelihood with bandwidth $m = n^{0.7}$. The fitted log spectral densities are plotted in Figure 1(a). The result is again quite informative, showing that for very low frequencies, $\hat{f}(\lambda)$ underestimates the slope of the log-periodogram and that for relatively higher frequencies, it does the opposite. The difference between $\hat{f}(\lambda)$ and $I_x(\lambda)$ provides an opportunity for testing the null hypothesis. Figure 1(b) reports the local Whittle estimates using different bandwidths $m$ ranging between $n^{1/3}$ and $n^{0.8}$. As predicted by the theory, the estimates are greater than 0.5 when $m$ is small and they

decrease substantially when $m$ is increased. Next, consider processes with smoothly varying trends of the form of (5). The series are generated using a simple polynomial trend function $h(x) = 2x - 4x^2$ with the errors $z_t$ being $i.i.d.N(0,1)$. The analogous results are reported in Figures 2(a) and 2(b). The finding here is similar and again confirms the asymptotic analysis. We also experimented with other low order polynomial trend functions and the results are qualitatively similar.

Thus, it is possible to distinguish between true and spurious long memory and when doing so, it is important to consider frequencies both below and above $n^{1/2}$. We need $m/n^{1/2} \to \infty$.

## 3 The test statistic

We now construct a test statistic for the null hypothesis that $x_t$ is a long memory process satisfying (1). It is based on the local Whittle likelihood function given by (see Künsch, 1987)

$$Q(G,d) = \frac{1}{m} \sum_{j=1}^{m} \left\{ \log G\lambda_j^{-2d} + \frac{I_j}{G\lambda_j^{-2d}} \right\}, \tag{6}$$

where $I_j = I_x(\lambda_j)$ and $m$ is some integer that is small relative to $n$. Minimizing $Q(G,d)$ with respect to $G$ leads to the profiled likelihood function: $R(d) = \log G(d) - 2m^{-1}d\sum_{j=1}^{m} \log \lambda_j$, where

$$G(d) = m^{-1} \sum_{j=1}^{m} \lambda_j^{2d} I_j.$$

The derivative of $R(d)$ is given by

$$\frac{\partial R(d)}{\partial d} = \frac{2G_0}{\sqrt{m}G(d)} \left\{ m^{-1/2} \sum_{j=1}^{m} v_j \left( \frac{I_j}{G_0\lambda_j^{-2d}} - 1 \right) \right\} \tag{7}$$

where

$$v_j = \log \lambda_j - (1/m) \sum_{j=1}^{m} \log \lambda_j$$

and $G_0$ is the true value of $G$. The term in the braces in (7) will be the main ingredient of our test statistic. It is also related to the LM statistic considered in Robinson (1994). Under the null hypothesis and evaluated at $d_0$, it equals to $m^{-1/2}\sum_{j=1}^{m} v_j(I_j\lambda_j^{2d_0}/G_0 - 1)$, which, as shown by Robinson (1995b), satisfies a central limit theorem. Building upon his result, we will prove later that

7

the quantity $m^{-1/2} \sum_{j=1}^{[mr]} v_j(I_j \lambda_j^{2d_0}/G_0) - 1)$, when treated as a process in $r$, satisfies a functional central limit theorem and hence is uniformly $O_p(1)$ under the null hypothesis. Meanwhile, under the alternative hypothesis, the summands in the preceding display do not have mean zero and the quantity diverges (c.f. the decomposition of $I_x(\lambda_j)$ in the previous section). This motivates the following test statistic:

$$W = \sup_{r \in [\varepsilon, 1]} \left( \sum_{j=1}^{m} v_j^2 \right)^{-1/2} \left| \sum_{j=1}^{[mr]} v_j \left( \frac{I_j}{G(\hat{d})\lambda_j^{-2\hat{d}}} - 1 \right) \right|, \tag{8}$$

where $\hat{d}$ is the local Whittle estimate of $d$ using $m$ frequency components and $\varepsilon$ is a small trimming parameter. Note that $m^{-1} \sum_{j=1}^{m} v_j^2 \to 1$ so that in principle $(\sum_{j=1}^{m} v_j^2)^{-1/2}$ could be replaced by $m^{-1/2}$. However, using the former brings the exact size of the test closer to its nominal level.

**Remark 2** *The number of frequency ordinates $m$ needs to satisfy $m/n^{1/2} \to \infty$ to achieve good power and be less than $n^{4/5}$ to avoid bias when estimating $d$. We will use simulations to examine the sensitivity of the results to such choices. It turns out that $m = n^{0.7}$ seems to achieve a good balance between size and power; thus it is suggested in practice.*

**Remark 3** *$v_j$ is unbounded for $j = o(m)$. The introduction of trimming permits asymptotic approximations that are adequate even in small samples. Note that the test statistic value is a decreasing function of $\varepsilon$ (smaller $\varepsilon$ implies more frequencies included in the statistic, thus a higher test statistic value). However, the critical value is also a decreasing function of $\varepsilon$ for the same reason. Thus, the relative power will depend on which one dominates. In practice, if the sample size is small, say $n < 500$, we suggest using $\varepsilon = 0.05$. For larger samples, smaller trimmings can be used. We will use simulations to evaluate the effect of different trimmings on the size and power of the test.*

The test is a score type test statistic. It does not require specifying the form of the trend or the number or locations of the different regimes that occur under the alternative hypothesis. It also

inherits two desirable properties of the local Whittle estimator, which allows for non-Gaussianity and is more efficient than the GPH estimator, see Robinson (1995b).

The $W$ test has features that are quite different from other tests in the literature. Dolado et al. (2005) proposed a test for the null hypothesis of fractional integration against the alternative hypothesis of short memory with a single structural break in the deterministic component (a constant or a linear trend). It is based on the t-statistic in an augmented Dickey-Fuller regression. Mayoral (2010) considered a similar problem and proposed a likelihood ratio-type test statistic. Their tests allow $d > 1/2$ and are useful for detecting a single structural change. In contrast, the $W$ test is designed to detect a more general class of alternatives without requiring specification of the form of the trend or the number of breaks. Shimotsu (2006) proposed a Wald test that compares local Whittle estimates obtained from $k$ equally sized subsamples, in line with Lobato and Savin (1998). By choosing $k$ greater than 2, the test can handle multiple breaks. He also proposed two tests based on the $d$th differenced series. We will compare the $W$ test with Shimotsu's tests using simulations.

## 4    The asymptotic properties of the test

The following assumptions are imposed, which are the same as in Robinson (1995b). The derivation below makes heavy use of his results.

**Assumption 1**. For some $\beta \in (0, 2]$, $f(\lambda) = G_0 \lambda^{-2d_0} (1 + O(\lambda^\beta))$ as $\lambda \to 0+$, where $G_0 \in (0, \infty)$ and $d_0 \in [\Delta_1, \Delta_2]$ with $-1/2 < \Delta_1 < \Delta_2 < 1/2$.

**Assumption 2**. Assume $x_t - E(x_0) = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}$ with $\sum_{j=0}^{\infty} \alpha_j^2 < \infty$, $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$, $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = 1$, $E(\varepsilon_t^3 | \mathcal{F}_{t-1}) = \mu_3$ and $E(\varepsilon_t^4 | \mathcal{F}_{t-1}) = \mu_4$ a.s. for $t = 0, \pm 1, ...$, where $\mathcal{F}_t$ is the $\sigma - field$ generated by $\varepsilon_s$, $s \leq t$. Also assume there exists a random variable $Z$ such that $E(Z^2) < \infty$ and $P(|\varepsilon_t| > \eta) \leq KP(|Z| > \eta)$ for all $\eta > 0$ and some $K > 0$.

**Assumption 3**. Let $\alpha(\lambda) = \sum_{l=0}^{\infty} \alpha_l \exp(i\lambda l)$. In a neighborhood $(0, \delta)$ of the origin, $\alpha(\lambda)$ is differentiable and $d\alpha(\lambda)/d\lambda = O(|\alpha(\lambda)|/\lambda)$ as $\lambda \to 0+$.

**Assumption 4**. As $n \to \infty$, $1/m + m^{1+2\beta} (\log m)^2 / n^{2\beta} \to 0$.

The assumptions allow for non-Gaussianity but preclude conditional heteroskedasticity. To allow for both features, we also consider an alternative set of conditions due to Shao and Wu (2007). Specifically, assume $x_t$ is generated by

$$(1 - L)^d(x_t - E(x_0)) = \xi_t, \ t = 0, \pm 1, \pm 2, ...,$$

where $\xi_t = F(..., e_{-1}, e_0, ..., e_t)$ with $e_t$ being i.i.d. random variables and $F$ being a measurable function such that $\xi_t$ is well defined. The next Assumption corresponds to Assumptions 2.3, 2.4 and 2.6 in Shao and Wu (2007).

**Assumption H**. Assume that Assumption 1 is satisfied with $\beta \geq 1$ and that the following hold: (**a**) $\sum_{k_1,k_2,k_3\in\mathbb{Z}} \left|\text{cum}(\xi_0, \xi_{k_1}, \xi_{k_2}, \xi_{k_3})\right| < \infty$, where cum(.) denotes the joint fourth cumulant and $\mathbb{Z}$ the set of integers. (**b**) Let $\xi_k' = F(..., e_{-1}, e_0', ..., e_k)$ with $e_0'$ being an independent copy of $e_0$. Assume $\xi_t \in \mathcal{L}^q$ and $\sum_{k=1}^{\infty} k \left\|\xi_k - \xi_k'\right\|_q < \infty$ with $q > 4$. (**c**) $m$ satisfies $(\log n)^3/m + m^3/n^2 \to 0$ if $\beta > 1$, and $(\log n)^3/m + m^3(\log m)^2/n^2 \to 0$ if $\beta = 1$.

The parameter $\beta$ concerns a particular property of the short memory component $\xi_t$. Specifically, the condition $\beta \geq 1$ implies that its spectral density is differentiable at the origin. The cumulant condition in H(a) is widely used in the time series literature. Andrews (1991, Lemma 1) shows that it is implied by an $\alpha$-mixing condition when $\xi_t \in \mathcal{L}^q$ with $q > 4$. H(b) quantifies the dependence of $\xi_t$ on $e_0$ by measuring the distance between $\xi_t$ and its coupled version $\xi_k'$ and can be verified for many nonlinear time series models. Shao and Wu (2007) show that Assumptions H(a) and (b) allow for ARFIMA$(p, d, q)$-GARCH$(r, s)$ and ARFIMA$(p, d, q)$-EGARCH$(r, s)$ processes provided $\xi_t \in \mathcal{L}^q$ with $q > 4$. These processes have $\beta = 2$. Note that H(b) implies the spectral density of $\xi_t$ is continuously differentiable over $[-\pi, \pi]$, a property not required in Assumptions 1 to 4. Also, H(c) imposes stronger conditions on $m$ than Assumption 4. These are the prices we pay for allowing for conditional heteroskedasticity. Nevertheless, simulation evidence presented in Section 6 indicates that the size of the test is insensitive to the choice of $m$, even when substantial conditional heteroskedasticity is present.

10

We first establish a uniform weak law of large numbers and a functional central limit theorem. They are needed for proving Theorem 1 and are also of independent interest.

**Lemma 2** *Let* $Q_m(r) = m^{-1} \sum_{j=1}^{[mr]} (I_j/(G_0\lambda_j^{-2d_0}))$. *Under Assumptions 1-4 or H,* $Q_m(r) \to^p r$ *uniformly on* $[0,1]$ *and* $m^{1/2}(Q_m(r) - r) \Rightarrow W(r)$ *with* $W(r)$ *being the Wiener process on* $[0,1]$.

**Theorem 1** *Under Assumptions 1-4 or Assumption H, as* $n \to \infty$,

$$W \Rightarrow \sup_{r\in[\varepsilon,1]} \left| \int_0^r (1+\log s)dW(s) - W(1)\int_0^r (1+\log s)ds - \Phi(r)\int_0^1 (1+\log s)dW(s) \right|,$$

*where* $\Phi(r) = \int_0^r (1+\log s)^2 ds$ *and* $W(r)$ *is a Wiener process defined on* $[0,1]$.

The limiting distribution in Theorem 1 is a result of the following first-order Taylor approximation to (8):

$$
\begin{aligned}
W = \sup_{r\in[\varepsilon,1]} &\left| m^{-1/2}\sum_{j=1}^{[mr]} v_j \left(I_j/(G_0\lambda_j^{-2d_0}) - 1\right) - \left(m^{-1/2}\sum_{j=1}^{m}\left(I_j/(G_0\lambda_j^{-2d_0}) - 1\right)\right)\left(m^{-1}\sum_{j=1}^{[mr]} v_j\right) \right. \\
&\left. + 2\Phi(r)m^{1/2}(\hat{d} - d_0) \right| + o_p(1).
\end{aligned}
$$

The three terms inside the absolute value signs correspond to the three terms in the limiting distribution. Weak convergence is proved by first showing finite-dimensional convergence and then verifying tightness. The additional difficulty presented by Assumption H is addressed using a martingale approximation technique developed in Wu and Shao (2007).

To simulate the limiting distribution, we approximate a Wiener process by $n^{-1/2}\sum_{i=1}^{[nr]} e_i$ with $e_i \sim i.i.d.N(0,1)$ and n=10,000 and record 10,000 realizations of $W$ to tabulate its critical values. The results are reported in Table 1, covering the following two cases $\varepsilon = 0.02$ and 0.05.

**Remark 4** *Lemma 2 and Theorem 1 are useful for studying the distribution of other related test statistics. For example, instead of (8), we can consider a Cramer-von Mises type statistic:*

$$\int_\varepsilon^1 \left(\sum_{j=1}^m v_j^2\right)^{-1}\left(\sum_{j=1}^{[mr]} v_j\left(\frac{I_j}{G(\hat{d})\lambda_j^{-2\hat{d}}} - 1\right)\right)^2 dr.$$

11

*Then, it follows immediately that its limiting distribution is given by*

$$\int_{\varepsilon}^{1} \left( \int_{0}^{r} (1 + \log s) dW(s) - W(1) \int_{0}^{r} (1 + \log s) ds - \Phi(r) \int_{0}^{1} (1 + \log s) dW(s) \right)^2 dr.$$

The following result implies that the $W$ test is consistent against the alternatives (3) and (5).

**Theorem 2** *Suppose $x_t$ is generated by (3) or (5). Assume $m/n^{1/2} \to \infty$, $P(\hat{d} > \epsilon) \to 1$ and $P(m^{-1} \sum_{j=1}^{m} I_j \lambda_j^{2\hat{d}} > \epsilon) \to 1$ as $n \to \infty$ with $\epsilon$ being some arbitrarily small constant. Then, $W \to^p \infty$ as $n \to \infty$.*

The assumption $m/n^{1/2} \to \infty$ is crucial for achieving consistency. $P(\hat{d} > \epsilon) \to 1$ ensures that $x_t$ exhibits spurious long memory asymptotically. This is needed because we do not impose specific parametric assumptions on the form of the trend or level shifts and thus are not directly ruling out the situation in which they are asymptotically negligible. To understand intuitively what leads to this condition, note that $\hat{d}$ solves (c.f. (7)) :

$$m^{-1} \sum_{j=1}^{m} \lambda_j^{2\hat{d}} I_j v_j = 0, \tag{9}$$

where $v_j$ satisfies $(1/m) \sum_{j=1}^{m} v_j = 0$, is monotonically increasing in $j$, and there exists $j^* = \kappa m$ with $\kappa > 0$ such that $v_j \leq 0$ ($v_j > 0$) for $j \leq j^*$ ($j > j^*$). Also, $j^*$ is of higher order than $n^{1/2}$ for large $n$ because $m/n^{1/2} \to \infty$. Now, if $x_t$ is a stationary short memory process, then $I_j$ is flat in $j$ for both $j \in [1, j^*]$ and $j \in [j^* + 1, m]$. Therefore, $\hat{d}$ will be close to zero in large samples. If $x_t$ is affected by level shifts or trends, then the $I_j$ in $[1, j^*]$ will tend to increase while those in $[j^* + 1, m]$ will stay flat, see (4) and Lemma 1. Therefore, $\hat{d}$ needs to be positive to satisfy (9). It will be strictly positive in large samples if the level shifts or trends are sufficiently pronounced. The condition $P(m^{-1} \sum_{j=1}^{m} I_j \lambda_j^{2\hat{d}} > \epsilon) \to 1$ is a variance condition. To see this, consider a true long memory process such as (2). Then, $m^{-1} \sum_{j=1}^{m} I_j \lambda_j^{2d_0}$ converges in probability to $\sigma_\varepsilon^2 |B(1)|^2 / (2\pi |A(1)|^2)$, the long run variance of $\{B(L)A(L)^{-1} \varepsilon_t\}$. Thus, these two conditions imply that $x_t$ mimics a long memory process with strictly positive variance.

12

There is a close connection between the above results and Müller and Watson (2008, Section 5). They considered a small number of frequencies and showed that it is very difficult to detect spurious long memory, which is consistent with the findings in (4) and Lemma 1. However, our result in Theorem 2 also shows that it is possible to construct useful tests if one is willing to consider frequencies beyond $j = O(n^{1/2})$, at the cost of making additional assumptions about these frequencies.

Now consider the situation where both long memory and level shifts or smooth trends are present. We say there is spurious long memory because the presence of the latter tends to bias the estimate of $d_0$ upward, thus spuriously increasing the apparent strength of long memory. We still consider the models (3) and (5) but with $z_t$ now being a true long memory process. The next Corollary shows the test is consistent against such alternatives.

**Corollary 1** *Suppose the process $x_t$ is generated by (3) or (5) but with $z_t$ being a true long memory process, with memory parameter $d_0$, satisfying Assumptions 1-4 or Assumption H. Also assume $m/n^{1/2} \to \infty$, $P(\hat{d} - d_0 > \epsilon) \to 1$, and $P(m^{-1} \sum_{j=1}^{m} I_j \lambda_j^{2\hat{d}} > \epsilon) \to 1$ as $n \to \infty$ with $\epsilon$ being some arbitrarily small constant. Then, $W \to^p \infty$ as $n \to \infty$.*

The condition $P(\hat{d} - d_0 > \epsilon) \to 1$ implies that the presence of level shifts or smooth trends spuriously increases the strength of long memory even asymptotically. The condition $P(m^{-1} \sum_{j=1}^{m} I_j \lambda_j^{2\hat{d}} > \epsilon) \to 1$ is again a variance condition, as in Theorem 2. These two conditions imply that $x_t$ mimics a long memory process with a memory parameter greater than $d_0$ and strictly positive variance.

**Remark 5** *The above approach corresponds to a "specific-to-general" modelling approach, where the simplest model (stationary long memory) is considered and tested, and additional features are seen to be present if the test rejects the null hypothesis. This approach has some advantages. It is simple because we do not need to specify the number of shifts or the form of the trend when conducting the W test. The result is easy to interpret because non-rejection implies that a long memory model is adequate and rejection suggests that the series contains a spurious long memory*

13

*component. It gives clear modelling suggestions, that is, to build a long memory model if the W*

*test does not reject and incorporate trends or level shifts into the model if it does.*

Now suppose the $W$ test rejects the null hypothesis. Then, to determine whether a true long memory component is also present, it is necessary to re-estimate the memory parameter after accounting for level shifts or a smooth trend.

Suppose a smooth trend is conjectured to be present, then the following procedures are useful. Beran and Feng (2002) proposed a class of models (SEMIFAR models) that allows for a fractional component and a deterministic trend component (I thank a referee for pointing this out). The fractional component is specified as an ARFIMA process and the trend component is modelled nonparametrically. The model can be estimated by combining parametric maximum likelihood estimation with kernel smoothing in an iterative fashion. Robinson (1997) and Hurvich et al. (2005) considered semiparametric estimation of the memory parameter in the presence of a smooth trend, with both the long memory and the trend components specified nonparametrically. Both papers apply the local Whittle estimator. In the latter, the estimator is applied to the tapered, differenced series with a number of low frequencies trimmed from the estimator's objective function. These studies provide confidence intervals for $\hat{d}$ using which the evidence for long memory can be assessed.

For the level shift case, we can consider the following three-step procedure. First, apply Lavielle and Moulines (2000) to estimate the number and locations of the level shifts. Suppose $k$ shifts are detected, resulting in $k+1$ exclusive subsamples. Second, apply the local Whittle estimator to each subsample to obtain $\hat{d}_j$ $(j = 1, ..., k+1)$. Third, construct an $F$ test for $d_1 = ... = d_{k+1} = 0$. Under the hypothesis that $d_j = 0$ $(j = 1, ..., k+1)$ and some additional assumptions, most importantly that the level shifts are not too small (formally, their magnitudes staying nonzero and fixed as $n \to \infty$), $\hat{d}_j$ have the usual asymptotic distribution as in Robinson (1995b) and are asymptotically independent for different $j$. Therefore, the $F$ test has a Chi-square limiting distribution with $k+1$

degrees of freedom. A rejection would then suggest that long memory is present after accounting for level shifts.

## 5 A finite sample correction to control the size

The short memory dynamics do not enter the asymptotic distribution. However, they may still have an important impact in small samples. In the following, we propose a "prewhitening" procedure that reduces the short memory component while maintaining the same limiting distribution for the test. It involves estimating a low order $ARFIMA$ model and filtering the series using the estimated autoregressive and moving average coefficients.

**Step 1**. Estimate an $ARFIMA(p, d, q)$ model for $p, q = 0, 1$. Determine the order using the Akaike Information Criterion (AIC). Let $\hat{\mu} = n^{-1} \sum_{i=1}^{n} x_t$ and $\hat{a}_1$ and $\hat{b}_1$ be the estimated autoregressive and moving average coefficients. Restrict the values of $\hat{a}_1$ and $\hat{b}_1$ to lie strictly within the unit interval to ensure stationarity and invertibility, i.e., $-1 + \delta \le \hat{a}_1, \hat{b}_1 \le 1 - \delta$ with $\delta$ being some small constant greater than zero. In simulations and empirical applications, we set $\delta = 0.01$.

**Step 2.** Compute $x_t^* = (1 - \hat{a}_1 L)(1 + \hat{b}_1 L)^{-1} (x_t - \hat{\mu})$. Explicitly, let $x_1^* = x_1 - \hat{\mu}$ and $x_t^* = (x_t - \hat{\mu}) - \sum_{k=1}^{t-1}(-\hat{b}_1)^{k-1}(\hat{a}_1 + \hat{b}_1)(x_{t-k} - \hat{\mu})$ for $t > 1$. Use $x_t^*$ instead of $x_t$ to construct $W$.

In Step 1, the time domain Gaussian likelihood is used for constructing the AIC. In simulations and empirical applications, it is computed using the *fracdiff* package in R that implements Haslett and Raftery (1989). The above procedure does not affect the null limiting distribution, provided Assumptions 2 and H are slightly strengthened.

**Assumption F**. (a) If Assumption 2 holds, then further assume $\alpha_j = O(j^{-1/2-c})$ with $c > 0$ as $j \to \infty$. (b) If Assumption H holds, then further assume $\xi_t$ satisfies a geometric moment contraction (GMC) condition with order $q > 4$. That is, let $\xi_k^* = F(..., e'_{-1}, e'_0, ..., e_{k-1}, e_k)$ with $e'_{-j}$ ($j = 0, 1, ...$) being independent copies of $e_{-j}$. Assume $\|\xi_k - \xi_k^*\|_q < C\rho^k$ for some $C < \infty$, $0 < \rho < 1$ and all positive integers $k$.

Assumption F(a) ensures the square summability condition in Assumption 2 is satisfied af-

ter filtering. It allows for stationary ARFIMA(p,d,q) models. F(b) ensures that Assumption H continues to hold after filtering. It allows for ARFIMA$(p, d, q)$-GARCH$(r, s)$ and ARFIMA$(p, d, q)$-EGARCH$(r, s)$ models. See Shao and Wu (2007) for other nonlinear time series satisfying GMC conditions.

**Corollary 2** *Assume $x_t$ satisfies the same conditions as in Theorem 1, with the additional requirement that Assumption F holds. Then, the W test constructed using $x_t^*$ has the same null limiting distribution as in Theorem 1.*

We do not assume the true DGP is an $ARFIMA$ process. It is just used as a reasonable approximation to the series so that some short memory dynamics can be removed. We have proposed using $ARFIMA(p, d, q)$ model with $p, q \leq 1$ as the basis for prewhitening. This is motivated by a host of research documenting that low order $ARFIMA$ models provide good approximations to many processes considered in economics and finance. For example, Andersen et al. (2003) studied the log realized volatility of the Deutschmark/Dollar and Yen/Dollar spot exchange rates and showed that an $AFRIMA(1, d, 0)$ approximation performs quite well when compared to other stationary time series models. Deo et al. (2006), Koopman et al. (2005) and Christensen and Nielsen (2007) studied other volatility series, arriving at similar conclusions. Note that $\hat{\mu}$ can be replaced by estimated seasonal dummies. The procedure remains asymptotically valid.

## 6 Simulations for finite sample properties

In this section, we examine the size and power properties of the $W$ test using simulations. We will also compare it to five test statistics in the existing literature.

The first test is due to Ohanissian et al. (2008), given by $ORT = (T\hat{D})'(T\Lambda T')^{-1}(T\hat{D})$, where $\hat{D} = (\hat{d}^{(m_1)}, ..., \hat{d}^{(m_M)})$, where $\hat{d}^{(m_j)}$ is the GPH estimate of $d$ at aggregation level $m_j$, $T$ is an $(M-1) \times M$ matrix of constants and $\Lambda$ is defined by equation (3) in their paper. As in Ohanissian et al. (2008), the aggregation levels are $2^{j-1}$ with $j = 1, ..., M$. Constrained by the sample size,

we set $M = 4$. The number of frequency ordinates used for each GPH estimate is the square root of the length of each (temporally aggregated) series. The second test is the $mean$-$t_d$ test of Perron and Qu (2010). Let $\hat{d}_{a,c}$ denote the log-periodogram estimate of $d$ when $c[n^a]$ frequencies are used and let $t_d(a, c_1; b, c_2) = \sqrt{24 c_1 [n^a]/\pi^2}(\hat{d}_{a,c_1} - \hat{d}_{b,c_2})$. Then, $mean$-$t_d$ is defined as the average of the $t_d(1/3, c_1; 1/2, 1)$ tests for $c_1 \in [1, 2]$. Its limiting distribution is not available. As in Perron and Qu (2010), a parametric bootstrap procedure is used to compute relevant critical values. Specifically, for a given series, an ARFIMA(1,d,1) model is estimated and then the null distribution of the test is simulated using this as the DGP. They also considered a sup-$t_d$ test, which performs very similar to the mean-$t_d$ test in our simulations and is omitted to save space. The remaining three test statistics are due to Shimotsu (2006). The $W_c$ test compares estimates of $d$ from $k$ equally sized subsamples. We set $k = 4$, following the suggestion in Shimotsu (2006). The $\eta_\mu$ and $Z_t$ tests apply the KPSS and Phillips-Perron test to the $d$th differenced series. For these three tests, the parameter $d$ is estimated using the local Whittle likelihood with $m = 4[(n^{0.7})/4]$. Results are obtained using the Matlab code provided by Shimotsu.

For all local Whittle-based tests ($W, \eta_\mu, W_c$ and $Z_t$), $\hat{d}$ is obtained by minimizing the objective function (6) over $d \in [-1/2, 1]$. Note that this interval is greater than that stated in Assumption 1. This does not affect the null limiting distribution of the test, however, and can deliver tests with better size properties when $d_0$ is close to 0.5 by avoiding the boundary issue. All tests are constructed by first applying the finite sample correction described in the previous section, except for the $mean$-$t_d$ test. Since the latter uses a bootstrap procedure to generate critical values, the finite sample correction is irrelevant. The data are generated using the *fracdiff* package designed for the R environment. All results are based on 10,000 replications.

To evaluate the size of the test, we consider models with parameter values that are typical in financial applications. The aforementioned research, along with others, suggests that $d$ typically takes on values between 0.30 and 0.45. We thus set $d = 0.4$ and consider the following specifications:

- **DGP 1**. $ARFIMA(0, d, 0) : (1 - L)^{0.4} x_t = e_t$.

- **DGP 2.** $ARFIMA(1, d, 0) : (1 - a_1 L)(1 - L)^{0.4} x_t = e_t$, where $a_1 = 0.4$ and $0.8$.

- **DGP 3.** $ARFIMA(0, d, 1) : (1 - L)^{0.4} x_t = (1 + b_1 L) e_t$, where $b_1 = 0.4$ and $0.8$.

- **DGP 4.** $ARFIMA(2, d, 0) : (1 - a_1 L)(1 - a_2 L)(1 - L)^{0.4} x_t = e_t$, where $a_1 = 0.5$ and $a_2 = 0.3$.

- **DGP 5.** $x_t = z_t + \eta_t$, where $(1 - L)^{0.4} z_t = e_t$ and $\eta_t \sim i.i.d. N(0, var(z_t))$.

- **DGP 6.** $(1 - L)^{0.4} x_t = u_t$ with $u_t = \sigma_t e_t$, $\sigma_t^2 = 1 + 0.10 u_{t-1}^2 + 0.85 \sigma_{t-1}^2$.

In all cases, $e_t \sim i.i.d. N(0, 1)$. DGP 1 corresponds to the ideal situation for local Whittle estimation. DGP 2 to DGP 4 contain substantial short memory components. Since DGP 4 is not a sub case of ARFIMA(1,d,1) models, it is used to illustrate the effectiveness of the finite sample correction. DGP 5 consists of a fractionally integrated process affected by measurement errors. Such a specification is relevant for applications to implied and realized volatility. The variance of the measurement error is set equal to the variance of the $z_t$ component, following the simulation design of Bandi and Perron (2006). DGP 6 exhibits strong conditional heteroskedasticity. For each DGP, we consider three sample sizes: $n$=500, 1000, 2000. They are similar to the sample sizes in the empirical applications in Section 7.

Table 2 presents empirical rejection frequencies at a 5% nominal level. The size of the $W$ test is fairly stable across different sample sizes, DGPs, and values of $m$ and $\varepsilon$. Overall, it appears to be conservative, except there is a slight tendency toward over-rejection when $m = n^{0.75}$ for DGP 4 and DGP 5. The ORT test exhibits the best size properties. The $W_c$ test of Shimotsu (2006) tends to overreject the null hypothesis, with the maximum rejection frequency being 0.108 for DGP 6. Other tests have decent size. We also repeated the analysis using unfiltered series to find that the finite sample correction has no effect on ORT, a small effect on $W_c$, and significantly improves the $W$, $\eta_\mu$ and $Z_t$ tests, especially for DGPs 2, 4 and 5. For example, for DGP 2 with $a_1 = 0.8$ and $n = 500$, the rejection frequencies would be 0.352 for the $W$ test ($\varepsilon$=0.02, $m = n^{0.70}$), 0.000 for the $\eta_\mu$ test and 0.593 for the $Z_t$ test if the finite sample correction is not used. The size does not improve when n is increased to 2000. Thus, the correction is indeed quite effective.

For power properties, we consider six alternative models. The first five are the same as in

Ohanissian et al. (2008). The last model contains a smooth but non-monotonic trend, for which parameter values are chosen to make $\hat{d}$ close to 0.4.

1. Nonstationary random level shift: $y_t = \mu_t + \varepsilon_t, \mu_t = \mu_{t-1} + \pi_t \eta_t, \pi_t \sim i.i.d.B(1, 6.10/n), \varepsilon_t \sim i.i.d.N(0, 5), \eta_t \sim i.i.d.N(0, 1)$.

2. Stationary random level shift: $y_t = \mu_t + \varepsilon_t, \mu_t = (1 - \pi_t)\mu_{t-1} + \pi_t\eta_t, \pi_t \sim i.i.d.B(1, 0.003), \varepsilon_t$ and $\eta_t \sim i.i.d.N(0, 1)$.

3. Markov switching with iid regimes: $y_t \sim i.i.d.N(1, 1)$ if $s_t = 0$ and $y_t \sim i.i.d.N(-1, 1)$ if $s_t = 1$, with state transition probabilities $p_{10} = p_{01} = 0.001$.

4. Markov switching with GARCH regimes: $r_t = \sqrt{h_t}\varepsilon_t$ and $h_t = 1 + 2s_t + 0.4r_{t-1}^2 + 0.3h_{t-1}$, where $\varepsilon_t \sim i.i.d.N(0, 1)$, $s_t = 0, 1$ and $p_{10} = p_{01} = 0.001$. $y_t = \log r_t^2$.

5. White noise with a monotonic deterministic trend: $y_t = 3t^{-0.1} + \varepsilon_t, \varepsilon_t \sim i.i.d.N(0, 1)$.

6. White noise with a non-monotonic deterministic trend: $y_t = \sin(4\pi t/n) + \varepsilon_t, \varepsilon_t \sim i.i.d.N(0, 3)$.

The studied sample sizes $n$ vary between 500 and 9000 and we set $m = n^{0.70}$. Other specifications are the same as before. Table 3 reports size-unadjusted power at a 5% nominal level, with bold numbers denoting the highest power among all tests. The results are encouraging. For models 1-4 and 6, the power of the $W$ test with $\varepsilon = 0.02$ is the highest among all tests once $n$ reaches 3000. The power difference is often substantial, particularly for models 2-4 and 6. The power of ORT, $mean\text{-}t_d$ and $Z_t$ is in general much lower. Note that Ohanissian et al. (2008) showed that the power of their test is 1 when the sample size is 610304. Thus, their test is more suitable for very large sample sizes, which may be available when analyzing high-frequency data[1]. The $\eta_\mu$ and $W_c$ tests of Shimotsu (2006) have merits, with the $\eta_\mu$ test being very powerful at detecting monotonic trends, a property inherited from the KPSS test. However, their power is quite low under DGPs 2, 3 and 4. Thus, the $W$ test performs the best overall, in the sense that it has decent power against a wide range of alternatives for sample sizes typical in financial applications.

---

[1] For $n = 9000$, we also tried to increase the number of aggregation levels to $M = 8$ for Ohanissian et al.'s (2008) test. The rejection frequencies are 0.46, 0.20, 0.11, 0.21, 0.24 and 0.23 for the six respective processes.

Finally, the results show that a large trimming ($\varepsilon = 5\%$) may lead to nonnegligible loss in power. Note that in financial applications we typically face samples of a few thousands observations. In such cases, based on these limited simulation results, $\varepsilon = 2\%$ seems to achieve a good balance in terms of size and power and is thus suggested in practice.

## 7 Applications

We apply the $W$ test to three time series for which empirical evidence of long memory has been documented. The series are: (1) monthly temperature for the northern hemisphere for the years 1854-1989, (2) monthly US inflation rates from January 1958 to December 2008, and (3) realized volatility for Japanese Yen/US dollar spot exchange rates from December 1, 1986 to June 30, 1999. The $W$ test uses $m = n^{0.70}$ unless stated otherwise. Other specifications are the same as above. The results from the tests are summarized in Table 4.

The temperature series was obtained from Beran (1994, pp.257-261). It contains 1632 observations and is shown in Figure 3(a). The local Whittle estimate is 0.33 using $m = n^{0.70}$. The $W$ test is significant at the 1% level for both $\varepsilon = 0.02$ and $\varepsilon = 0.05$. To obtain some further insight, we can consider local Whittle estimates using different numbers of frequencies: $n^{1/3} \leq m \leq n^{4/5}$. The estimates are shown in figure 3(b), as a function of $m$. The result is very informative. It shows that when a small number of frequencies is used, the estimate is above 0.5 and it decreases significantly as more frequencies are included. This finding is consistent with the presence of level shifts or smooth trends, but inconsistent with stationary fractional integration (see Figures 1(b) and 2(b)). Figure 3(c) provides more evidence. It reports memory parameter estimates using observations up to $n_b$, with $n_b$ ranging between 300 and 1632. The estimates vary substantially. This again suggests that nonstationarity is genuinely present in the sample. Among the other tests, $\eta_\mu$ and $W_c$ rejects at 1% level while the rest do not reject even at 5% level. This is consistent with the simulation evidence that $\eta_\mu$ and $W_c$ have relatively higher power among the remaining five tests.

The inflation series was constructed from the consumer price index for all urban consumers

and all items (the series SA0, seasonally unadjusted, available at Bureau of Labor Statistics). It contains 612 observations and is plotted in Figure 4(a). The local Whittle estimate is 0.33 using $m = n^{0.70}$. The $W$ test is significant at the 5% level when $\varepsilon$=0.02 and at the 1% level when $\varepsilon$=0.05. Figures 4(b) and 4(c) reveal qualitatively similar findings to Figure 3, presenting evidence against the null hypothesis of stationary long memory. Among the other tests, $W_c$ rejects at 1% level, $\eta_\mu$ barely rejects at 5% level and the rest do not reject.

The realized volatility series was constructed using 5-minute returns for the Yen/\$ spot rate, obtained from Olsen and Associates. The series was constructed following the same procedure as discussed in Andersen et al. (2001). More specifically, we first obtained the daily realized variances by summing the squared 5-minute returns. Then, we applied the logarithm transformation to obtain log realized volatility. The weekends and holidays were dropped following Andersen et al. (2001). This left a sample of 2960 observations. The local Whittle estimate is 0.47 using $m = n^{0.70}$. The $W$ test equals 0.41 when $\varepsilon$=0.02 and 0.37 when $\varepsilon$=0.05, which are well below the 10% critical values in both cases. Figures 5(b) and 5(c) show that the memory parameter estimates remain stable when changing the number of frequency ordinates or the sample size. Other tests lead to the same conclusion. We also repeated the analysis using 30-minute returns. The result remains qualitatively the same.

The above results suggest that the evidence for stationary long memory may not be as strong as is often perceived. They also suggest that more research is needed to understand the nonlinear and nonstationary aspects of the relevant processes.

## 7.1    Further simulation evidence

In the first two applications, the sample sizes are relatively small and there is pronounced heteroskedasticity. We now consider simulations calibrated to empirical estimates obtained from these two series to assess whether the $W$ test has decent size and power in such contexts. Specifically,

we estimate the model

$$x_t = z_t + \mu_t \text{ with } (1 - a_1 L)(1 - L)^d z_t = (1 + b_1 L)e_t \tag{10}$$

under both the null and the alternative hypotheses and then use the parameter estimates to generate simulated data. Note that (10) allows for long memory even under the alternative hypothesis.

For the temperature series, $\mu_t$ is assumed to be a smooth trend under the alternative hypothesis. It is estimated nonparametrically using the same kernel as in Robinson (1997): $\hat{\mu}_t = (nb)^{-1} \sum_{s=1}^{n} K\left((nt - s)/(nb)\right) x_t$ with $K(v) = (1/2)\{1 + \cos(\pi v)\}$ for $|v| \le 1$ and $b = 0.1$. Other parameters are estimated via a time domain Gaussian likelihood function applied to $(x_t - \hat{\mu}_t)$. To preserve the heteroskedasticity present in the data, we use the wild bootstrap (Wu, 1986). Specifically, the disturbances $e_t^*$ are generated from the estimated residuals by $e_t^* = s_t \hat{e}_t$ with $s_t$ being i.i.d. random variables taking value $-(\sqrt{5} - 1)/2$ with probability $(\sqrt{5} + 1)/(2\sqrt{5})$ and $(\sqrt{5} + 1)/2$ with probability $(\sqrt{5} - 1)/(2\sqrt{5})$. This leads to the following two DGPs:

- **Temp-H$_0$.** $(1 - 0.056L)(1 - L)^{0.354} z_t = (1 + 0.015L)e_t^*$ and $\hat{\mu}_t = -0.149$ for all t,

- **Temp-H$_1$.** $(1 - 0.560L)(1 - L)^{0.000} z_t = (1 - 0.174L)e_t^*$ and $\hat{\mu}_t$ is depicted in Figure 3(a).

For the inflation rate series, $\mu_t$ is assumed to contain two level shifts under the alternative hypothesis. The break dates are fixed exogenously as in Bos et al. (1999). They correspond to a break shortly before the first oil crisis (1973:07) and shortly after the second oil crisis (1982:07). $\mu_t$ is estimated using OLS and is depicted in figure 4(a). Other aspects are the same as before. This leads to the following two DGPs:

- **Infl-H$_0$.** $(1 - 0.896L)(1 - L)^{0.388} z_t = (1 - 0.892L)e_t^*$ and $\hat{\mu}_t = 0.039$ for all t,

- **Infl-H$_1$.** $(1 - 0.412L)(1 - L)^{0.118} z_t = (1 - 0.143L)e_t^*, \hat{\mu}_1 = 0.029, \hat{\mu}_{188} = 0.088$ and $\hat{\mu}_{296} = 0.029$.

The results are summarized in Table 5. The $W$ test exhibits good size properties, with rejection frequencies varying between 2.2% and 6.6%. It also has decent power given that the sample sizes are relatively small. Among the competing tests, the $\eta_\mu$ test performs very well under Temp-H$_1$, but has virtually no power under Infl-H$_1$. The $W_c$ test performs very well under Infl-H$_1$, but has very

low power under Temp-$H_1$. These findings are entirely consistent with the ones reported in Section 6. They are also broadly consistent with the pattern we observe in Table 4, except that there the $\eta_\mu$ test rejects the null hypothesis when applied to the inflation rate series. This is possibly due to its size distortions (c.f. Panel b in Table 5), or because the inflation series entails more complex dynamics than allowed in the model (10). For the temperature series, we varied $b$ between 0.05 and 0.20 while maintaining all other parameter values at the estimated values and the result remain the same qualitatively.

## 8   Conclusion

We have considered the issue of distinguishing between true and spurious long memory. We first compared the spectral domain properties of stationary long memory processes with short memory processes containing level shifts or smoothly varying trends. We then proposed a simple test statistic based on the derivatives of the profiled local Whittle likelihood function. The limiting distribution under the null hypothesis was derived using theory of empirical processes. Simulations showed the test has decent size and power properties. The test was applied to three time series for which empirical evidence for long memory has been documented. The result of this empirical exercise suggests that the evidence for stationary long memory may not be as strong as is often perceived.

## References

Andersen, T., T. Bollerslev, F.X. Diebold & H. Ebens (2001) The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43-76.

Anderson, T., T. Bollerslev, F.X. Diebold & P. Labys (2003) Modeling and forecasting realized volatility. *Econometrica* 71, 579-626.

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

Bandi, F.M & B. Perron (2006) Long memory and the relation between implied and realized volatility. *Journal of Financial Econometrics* 4, 636-670.

Bartlett, M.S. (1955) An Introduction to stochastic processes with special reference to methods and applications. Cambridge University Press, London.

Berkes, I., L. Horváth, P. Kokoszka & Q. Shao (2006) On discriminating between long-range dependence and changes in mean. *The Annals of Statistics* 34, 1140–1165.

Beran, J & Y. Feng (2002) SEMIFAR models-a semiparametric approach to modelling trends, long-range dependence and nonstationarity, *Computational Statistics & Data Analysis* 40, 393 – 419.

Billingsley, P. (1999) Convergence of probability measures. 2nd Edition. Wiley, New York.

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

Christensen, B.J. & M.Ø. Nielsen (2007) The effect of long memory in volatility on stock market fluctuations. *The Review of Economics and Statistics* 89, 684–700.

Deo, R., C. Hurvich & L. Yi (2006) Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment. *Journal of Econometrics* 131, 29–58.

Ding, Z., R.F. Engle, & C.W.J. Granger (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83-106.

Dolado, J.J., J. Gonzalo & L. Mayoral (2005) What is what?: a simple time-domain test of long-memory vs. structural breaks. Working paper, Universidad Carlos III de Madrid.

Fox, R. & M.S. Taqqu (1986) Large sample properties of parameter estimates for strongly dependent stationary time series. *The Annals of Statistics* 14, 517-532.

Geweke, J. & S. Porter-Hudak (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4, 221-238.

Giraitis, L., R. Leipus & A. Philippe (2006) A test for stationarity versus trends and unit roots for a wide class of dependent errors. *Econometric Theory* 22, 989-1029.

Granger, C.W.J. & R. Joyeux (1980) An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15-39.

Grenander, U. & M. Rosenblatt (1953) Statistical spectral analysis arising from stationary stochastic processes. *Annals of Mathematical Statistics* 24, 537-558.

——1957. Statistical analysis of stationary time series. Wiley, New York.

Koopman, S.J., Jungbacker, B & E. Hol (2002) Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* 12, 445-475.

Haslett, J. & A.E. Raftery (1989) Space-time modelling with long-memory dependence: assessing Ireland's wind power resource (with discussion). *Applied Statistics* 38, 1–50.

Hosking, J.R.M. (1981) Fractional differencing. *Biometrika* 68, 165-176.

Hurvich, C., Lang, G. & P. Soulier (2005) Estimation of long memory in the presence of a smooth nonparametric trend. *Journal of the American Statistical Association* 100, 853-871.

Ibragimov, I.A. (1963) On estimation of the spectral function of a stationary Gaussian process. *Theory of Probability and its Applications* VIII, 366-400.

Kokoszka, P. & T, Mikosch (1997) The integrated periodogram for long memory processes with finite or infinite variance. *Stochastic Processes and their Applications* 66, 55-78.

Künsch, H.R. (1986) Discriminating between monotonic trends and long-range dependence. *Journal of Applied Probability* 23, 1025-1030.

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

Lavielle, M. & E. Moulines (2000). Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis* 21, 33-59.

Lobato, I.N. & N.E. Savin (1998). Real and spurious long-memory properties of stock-market data. *Journal of Business and Economics Statistics* 16, 261-268.

Mayoral, L. (2010) Testing for fractional integration versus short memory with trends and structural breaks. Working paper, Universitat Pompeu Fabra.

Müller, U.K. & Watson, M. (2008). Testing models of low frequency variability. *Econometrica* 76, 979-1016.

Newey, W.K. (1991) Uniform convergence in probability and stochastic equicontinuity. *Econometrica* 59, 1161-1167.

Nielsen, M.Ø. (2004) Local empirical spectral measure of multivariate processes with long range dependence. *Stochastic Processes and their Applications* 109, 145–166.

Ohanissian, A., J.R. Russell & R.S. Tsay (2004). True or spurious long memory in volatility: does it matter for pricing options? Working paper, University of Chicago.

—— (2008). True or spurious long memory? A new test. *Journal of Business and Economics Statistics* 26, 161-175.

Perron, P. & Z. Qu (2010). Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business and Economic Statistics* 28, 275-290.

Robinson, P. M. (1994). Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* 89, 1420–1437.

——(1995a) Log periodogram regression of time series with long range dependence. *Annals of Statistics 23*, 1048-1072.

——(1995b) Gaussian semiparametric estimation of long range dependence. *Annals of statistics* 23, 1630-1661.

——(1997) Large-sample inference for nonparametric regression with dependent errors. *The Annals of Statistics* 25, 2054–2083.

Sibbertsen, P. & I. Venetis (2003) Distinguishing between long-range dependence and deterministic trends. Working paper, University of Dortmund.

Shao, X. & W.B. Wu (2007) Local Whittle estimation of fractional integration for nonlinear processes. *Econometric Theory* 23, 899-929.

Shimotsu, K. (2006) Simple (but effective) tests of long memory versus structural breaks. Working paper, Queen's University.

Smith, A. (2005) Level shifts and the illusion of long memory in economic time series. *Journal of Business and Economic Statistics* 23, 321-335.

Taylor, S.J. (2000) Consequences for option pricing of a long memory in volatility. Unpublished Manuscript. Department of Accounting and Finance, Lancaster University.

Wu, C.F.J. (1986) Jackknife bootstrap and other resampling methods in regression analysis. *Annals of Statistics* 14, 1261–1295.

Wu, W.B. & X. Shao (2007) A limit theorem for quadratic forms and its applications. *Econometric Theory* 23, 930-951.

## Appendix A. Proof of Main Results

This appendix contains proofs for Lemma 1, Theorem 1 and 2. Proofs for Lemma 2, Corollary 1 and 2 can be found in the supplementary appendix.

**Proof of Lemma 1.** The proof is a direct extension of Künsch (1986, p.1026). Consider

$$I_h(\lambda_j) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} h(t/n) \cos(\lambda_j t) \right|^2 + \frac{1}{2\pi n} \left| \sum_{t=1}^{n} h(t/n) \sin(\lambda_j t) \right|^2.$$

For $j \in \{1, 2, ..., [n/2]\}$,

$$\left| \sum_{t=1}^{n} h(t/n) \cos(\lambda_j t) \right|$$

$$\leq \sum_{t=1}^{n-1} |h((t+1)/n) - h(t/n)| \left| \sum_{s=1}^{t} \cos(\lambda_j s) \right| \leq \sum_{t=1}^{n-1} |h((t+1)/n) - h(t/n)| \max_{1 \leq t \leq n} \left| \sum_{s=1}^{t} \cos(\lambda_j s) \right|$$

$$\leq \frac{1}{2} \sum_{t=1}^{n-1} |h((t+1)/n) - h(t/n)| \left( \frac{\pi}{\lambda_j} + 1 \right) \leq \frac{(n-1)C}{2n} \left( \frac{\pi}{\lambda_j} + 1 \right) \quad \text{for some } 0 < C < \infty,$$

where the first inequality uses summation by parts and $\sum_{s=1}^{n} \cos(\lambda_j s) = 0$, the third is because $\sum_{s=1}^{t} \cos(\lambda_j s) = \sin((t+1/2)\lambda_j)/(2\sin(\lambda_j/2)) - 1/2$, and the fourth is due to Lipschitz continuity of $h(\cdot)$. Similarly,

$$\left| \sum_{t=1}^{n} h(t/n) \sin(\lambda_j t) \right| \leq \left( \frac{n-1}{n} C\pi \right) \lambda_j^{-1}.$$

Hence, $I_h(\lambda_j) = O(n^{-1} \lambda_j^{-2})$. ∎

**Proof of Theorem 1:** The proof consists of three steps.

**Step 1:** *Represent the statistic as a quantity linear in* $(\hat{d} - d_0)$. Since $m^{-1} \sum_{j=1}^{m} v_j^2 \to 1$, it suffices to consider

$$Y_m(r, \hat{d}) = m^{-1/2} \sum_{j=1}^{[mr]} v_j (I_j/(G(\hat{d})\lambda_j^{-2\hat{d}}) - 1).$$

Apply a first order Taylor expansion:

$$Y_m(r, \hat{d}) = m^{-1/2} \sum_{j=1}^{[mr]} v_j \left( \frac{I_j}{G(d_0)\lambda_j^{-2d_0}} - 1 \right) + m^{-1/2} \frac{\partial Y_m(r, \widetilde{d})}{\partial d} m^{1/2} (\hat{d} - d_0),$$

where

$$m^{-1/2} \frac{\partial Y_m(r, \widetilde{d})}{\partial d} = \frac{2}{mG(\widetilde{d})^2} \sum_{j=1}^{[mr]} v_j I_j \lambda_j^{2\widetilde{d}} \left\{ \frac{1}{m} \log \lambda_j \sum_{k=1}^{m} \lambda_k^{2\widetilde{d}} I_k - \frac{1}{m} \sum_{s=1}^{m} \lambda_s^{2\widetilde{d}} I_s \log \lambda_s \right\}, \qquad \text{(A.1)}$$

and $\widetilde{d}$ lies between $d_0$ and $\hat{d}$. The first term in the Taylor expansion can be rewritten as

$$m^{-1/2}\frac{G_0}{G(d_0)}\sum_{j=1}^{[mr]}v_j\left(\frac{I_j}{G_0\lambda_j^{-2d_0}}-1\right)-m^{-3/2}\frac{G_0}{G(d_0)}\left(\sum_{j=1}^{[mr]}v_j\right)\sum_{j=1}^{m}\left(\frac{I_j}{G_0\lambda_j^{-2d_0}}-1\right). \tag{A.2}$$

Next, consider (A.1). Adding and subtracting $(m^{-1}\sum_{j=1}^{m}\log\lambda_j)(m^{-1}\sum_{k=1}^{m}I_k\lambda_k^{2\widetilde{d}})$, the term in the curly brackets in (A.1) can be rewritten as $m^{-1}v_j\sum_{k=1}^{m}\lambda_k^{2\widetilde{d}}I_k-m^{-1}\sum_{s=1}^{m}v_s\lambda_s^{2\widetilde{d}}I_s$. Thus,

$$(A.1)=\frac{2}{G(\widetilde{d})^2}(\frac{1}{m}\sum_{k=1}^{m}I_k\lambda_k^{2\widetilde{d}})(\frac{1}{m}\sum_{j=1}^{[mr]}v_j^2I_j\lambda_j^{2\widetilde{d}})-\frac{2}{G(\widetilde{d})^2}(\frac{1}{m}\sum_{j=1}^{[mr]}v_jI_j\lambda_j^{2\widetilde{d}})(\frac{1}{m}\sum_{s=1}^{m}v_sI_s\lambda_s^{2\widetilde{d}}).$$

Lemma B.3 in the supplementary appendix shows $\frac{1}{G(\widetilde{d})}\left(\frac{1}{m}\sum_{j=1}^{[mr]}v_j^kI_j\lambda_j^{2\widetilde{d}}\right)=\frac{1}{m}\sum_{j=1}^{[mr]}v_j^k+o_p(1)$ uniformly in $r\in[0,1]$ for $k=0,1,2$. Applying this result, along with $\frac{1}{G(\widetilde{d})}(\frac{1}{m}\sum_{k=1}^{m}I_k\lambda_k^{2\widetilde{d}})=1$, (A.1) can be written as

$$2m^{-1}\sum_{j=1}^{[mr]}v_j^2-\left(2m^{-1}\sum_{j=1}^{[mr]}v_j\right)\left(m^{-1}\sum_{k=1}^{m}v_k\right)+o_p(1)\rightarrow 2\int_0^r(1+\log s)^2ds=2\Phi(r). \tag{A.3}$$

Combining (A.2) and (A.3), we have

$$\begin{aligned}Y_m(r,\hat{d})&=m^{-1/2}\frac{G_0}{G(d_0)}\sum_{j=1}^{[mr]}v_j\left(\frac{I_j}{G_0\lambda_j^{-2d_0}}-1\right)\\&\quad-m^{-3/2}\frac{G_0}{G(d_0)}\left(\sum_{j=1}^{[mr]}v_j\right)\sum_{j=1}^{m}\left(\frac{I_j}{G_0\lambda_j^{-2d_0}}-1\right)+2\Phi(r)m^{1/2}(\hat{d}-d_0)+o_p(1).\end{aligned} \tag{A.4}$$

**Step 2:** *Prove finite dimensional convergence.* Consider the first term in (A.4). Because $G_0/G(d_0)\rightarrow^p 1$, it suffices to study $m^{-1/2}\sum_{j=1}^{[mr]}v_j(I_j/(G_0\lambda_j^{-2d_0})-1)$ which, by Lemma B.6, equals to

$$m^{-1/2}\sum_{j=1}^{[mr]}v_j\left(2\pi I_{\varepsilon j}-1\right)+o_p(1) \tag{A.5}$$

uniformly over $r\in[\varepsilon,1]$. Rewrite the leading term in (A.5) as

$$\sum_{t=1}^{n}z_{t,r}\quad\text{with }z_{t,r}=\varepsilon_t\sum_{s=1}^{t-1}c_{t-s,r}\varepsilon_s\quad\text{and}\quad c_{t,r}=2n^{-1}m^{-1/2}\sum_{j=1}^{[mr]}v_j\cos(t\lambda_j).$$

Let $p$ denote an integer and suppose $0\le r_1\le...\le r_p\le 1$. It suffices to show

$$(\sum_{t=1}^{n}z_{t,r_s})_{s=1,...,p}\Rightarrow(\int_0^{r_s}(1+\log x)dW(x))_{s=1,...,p}. \tag{A.6}$$

First, the asymptotic normality of $(\sum_{t=1}^{n} z_{t,r_s})_{s=1,...,p}$ follows from Robinson (1995b, Theorem 2) and Shao and Wu (2007, Theorem 3.1). Second, for $0 \le r_1 \le r_2 \le 1$,

$$Cov\left(\sum_{t=1}^{n} z_{t,r_1}, \sum_{t=1}^{n} z_{t,r_2}\right) = 4n^{-2}m^{-1}\sum_{j=1}^{[mr_1]}\sum_{k=1,k\neq j}^{[mr_2]} v_j v_k \left(-\frac{n}{2}\right) + 4n^{-2}m^{-1}\sum_{j=1}^{[mr_1]} v_j^2 \left(\frac{n^2}{4}\right).$$

The first term is of order $O(\sum_{j=1}^{m} v_j^2/n) = O(m/n) \to 0$, see Robinson (1995b, p.1645). The second term converges to $\int_0^{r_1}(1+\log s)^2 ds$ by Lemma B.1 in the supplementary appendix. Therefore, (A.6) holds.

Consider the second term in (A.4). Because of Lemma B.7,

$$m^{-1/2}\sum_{j=1}^{m}(I_j/(G_0\lambda_j^{-2d_0}) - 1) = m^{-1/2}\sum_{j=1}^{m}(2\pi I_{\varepsilon j} - 1) + o_p(1) \Rightarrow W(1).$$

Also, because of Lemma B.1,

$$m^{-1}\sum_{j=1}^{[mr]} v_j \to \int_0^r (1+\log s)ds.$$

For the third term in (A.4), note that

$$m^{1/2}(\hat{d} - d_0) = -2^{-1}m^{-1/2}\sum_{j=1}^{m} v_j(I_j/(G_0\lambda_j^{-2d_0}) - 1) + o_p(1).$$

Thus, it can be analyzed in the same way as the second term. The detail is omitted.

**Step 3:** *Prove tightness.* We follow Nielsen (2004) and use Theorem 13.5 of Billingsley (1999). That is, we show that for every $m$ and $r_1 \le r \le r_2$,

$$E\left(\left|\sum_{t=1}^{n} z_{t,r} - \sum_{t=1}^{n} z_{t,r_1}\right|^2 \left|\sum_{t=1}^{n} z_{t,r_2} - \sum_{t=1}^{n} z_{t,r}\right|^2\right) \le K(\varphi_m(r_2) - \varphi_m(r_1))^2 \qquad (A.7)$$

where $K$ is some constant and $\varphi_m(.)$ is a function on $[0,1]$ that is finite, nondecreasing, and satisfies $\lim_{\delta \to 0}\limsup_{m\to\infty}|\varphi_m(s+\delta) - \varphi_m(s)| \to 0$ uniformly in $s \in [0,1]$. First, Lemma B.8 shows that the left hand side of (A.7) is bounded from above by

$$C(\sum_{t=1}^{n}\sum_{s=1}^{t-1} c_{t-s}(r_1,r)^2)(\sum_{t=1}^{n}\sum_{h=1}^{t-1} c_{t-h}(r,r_2)^2),$$

where $C$ is some constant and $c_{t-s}(r_1,r) = c_{t-s,r} - c_{t-s,r_1}$. Second,

$$\sum_{t=1}^{n}\sum_{s=1}^{t-1} c_{t-s}(r_1,r)^2 \le \frac{1}{nm}\sum_{j=[mr_1]+1}^{[mr]}\sum_{k\neq j}^{[mr]}\left(v_j^2 + v_k^2\right) + \frac{1}{m}\sum_{j=[mr_1]+1}^{[mr]} v_j^2 \le \frac{3}{m}\sum_{j=[mr_1]+1}^{[mr]} v_j^2.$$

Similarly,

$$\sum_{t=1}^{n}\sum_{h=1}^{t-1} c_{t-h}(r,r_2)^2 \le 3m^{-1}\sum_{j=[mr]+1}^{[mr_2]} v_j^2.$$

Finally, let $\varphi_m(s) = m^{-1}\sum_{j=1}^{[ms]} v_j^2$. It satisfies the conditions stated above because

$$\lim_{\delta\to 0}\limsup_{m\to\infty}|\varphi_m(s+\delta)-\varphi_m(s)| = \lim_{\delta\to 0}\int_s^{s+\delta}(1+\log x)^2 dx \to 0.$$

This completes the proof. ∎

**Proof of Theorem 2**. The proof uses the property that, for the two processes considered, $I_j = O_p(1)$ when $jn^{-1/2} \to \infty$. Note that $v_j$ is monotonically increasing in $j$ with $v_1 < 0$ and $v_m > 0$.

Define $j^* = \min\{j : v_j \ge 0\}$. Then, Lemma B.1 implies $j^* = \kappa m$ with $\kappa \approx 0.37$. Let

$$A = \left(\sum_{j=1}^{m} v_j^2\right)^{-1/2}\left|\sum_{j=j^*}^{m} v_j\left(\frac{I_j}{G(\hat{d})\lambda_j^{-2\hat{d}}}-1\right)\right| \text{ and } B = \left(\sum_{j=1}^{m} v_j^2\right)^{-1/2}\left|\sum_{j=1}^{j^*-1} v_j\left(\frac{I_j}{G(\hat{d})\lambda_j^{-2\hat{d}}}-1\right)\right|.$$

Using the reverse triangle inequality, we have $W \ge \max(A-B,B)$. Now, if $A \ge 2B$, then $W \ge A - B = A/2 + (A/2 - B) \ge A/2$. If $A < 2B$, then $W \ge B > A/2$. Thus $W \ge A/2$ always holds. Hence, to prove the result, it suffices to show $A \to^p \infty$ as $n \to \infty$.

Applying the reverse triangle inequality to term $A$,

$$A \ge \left(\sum_{j=1}^{m} v_j^2\right)^{-1/2}\sum_{j=j^*}^{m} v_j - \left(\sum_{j=1}^{m} v_j^2\right)^{-1/2}\sum_{j=j^*}^{m}\left(v_j\frac{I_j}{G(\hat{d})\lambda_j^{-2\hat{d}}}\right) \tag{A.8}$$

We study the two terms separately. The first term satisfies

$$\left(\sum_{j=1}^{m} v_j^2\right)^{-1/2}\sum_{j=j^*}^{m} v_j = m^{1/2}\int_\kappa^1 (1+\log s)ds + o(m^{1/2}), \tag{A.9}$$

which is strictly positive and of exact order $m^{1/2}$. For the second term, because $m/n^{1/2} \to \infty$, $j^*/n^{1/2} \to \infty$. Thus, for $j^* \le j \le m$, $I_j = O_p(1)$. Further, $v_j I_j/(G(\hat{d})\lambda_j^{-2\hat{d}}) = O_p(\lambda_j^{2\hat{d}}) = O_p(\lambda_j^{2\epsilon}) = o_p(1)$, where the first equality uses $v_j = O_p(1)$ for $j^* \le j \le m$, $G(\hat{d}) = m^{-1}\sum_{j=1}^{m} I_j\lambda_j^{2\hat{d}}$ and the variance condition in Theorem 2, the second equality uses $P(\hat{d} > \epsilon) \to 1$, and the last equality is because $\lambda_j = o(1)$. Thus, the second term is of order lower than $m^{1/2}$ and is dominated asymptotically by the first term (A.9). This shows $A \to^p \infty$ as $n \to \infty$. ∎

Table 2. Empirical size at 5% nominal level

| n | The proposed W test | | | | | | | | Other existing tests | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $m=n^{0.60}$ | | $m=n^{0.65}$ | | $m=n^{0.70}$ | | $m=n^{0.75}$ | | ORT | $W_c$ | $\eta_\mu$ | $Z_t$ | mean-$t_d$ |
| | $\varepsilon=2\%$ | $\varepsilon=5\%$ | $\varepsilon=2\%$ | $\varepsilon=5\%$ | $\varepsilon=2\%$ | $\varepsilon=5\%$ | $\varepsilon=2\%$ | $\varepsilon=5\%$ | | | | | |
| DGP1: $(1-L)^{0.4}x_t= e_t$ | | | | | | | | | | | | | |
| 500 | 0.010 | 0.013 | 0.015 | 0.017 | 0.013 | 0.014 | 0.016 | 0.019 | 0.059 | 0.082 | 0.035 | 0.007 | 0.062 |
| 1000 | 0.017 | 0.023 | 0.018 | 0.026 | 0.020 | 0.022 | 0.023 | 0.024 | 0.052 | 0.075 | 0.038 | 0.015 | 0.050 |
| 2000 | 0.024 | 0.027 | 0.021 | 0.026 | 0.024 | 0.026 | 0.028 | 0.029 | 0.057 | 0.068 | 0.037 | 0.023 | 0.055 |
| DGP2: $(1-0.4L)(1-L)^{0.4}x_t= e_t$ | | | | | | | | | | | | | |
| 500 | 0.011 | 0.017 | 0.016 | 0.020 | 0.019 | 0.019 | 0.013 | 0.018 | 0.056 | 0.080 | 0.048 | 0.003 | 0.051 |
| 1000 | 0.016 | 0.020 | 0.018 | 0.021 | 0.019 | 0.024 | 0.017 | 0.019 | 0.050 | 0.074 | 0.042 | 0.013 | 0.051 |
| 2000 | 0.020 | 0.027 | 0.024 | 0.032 | 0.027 | 0.028 | 0.024 | 0.024 | 0.059 | 0.074 | 0.040 | 0.026 | 0.055 |
| DGP2: $(1-0.8L)(1-L)^{0.4}x_t= e_t$ | | | | | | | | | | | | | |
| 500 | 0.010 | 0.013 | 0.011 | 0.011 | 0.006 | 0.009 | 0.008 | 0.013 | 0.062 | 0.083 | 0.052 | 0.026 | 0.071 |
| 1000 | 0.018 | 0.019 | 0.012 | 0.014 | 0.011 | 0.014 | 0.010 | 0.012 | 0.055 | 0.077 | 0.059 | 0.008 | 0.076 |
| 2000 | 0.019 | 0.020 | 0.017 | 0.016 | 0.018 | 0.016 | 0.013 | 0.015 | 0.050 | 0.066 | 0.047 | 0.005 | 0.065 |
| DGP3: $(1-L)^{0.4}x_t=(1+0.4L)e_t$ | | | | | | | | | | | | | |
| 500 | 0.011 | 0.014 | 0.015 | 0.019 | 0.018 | 0.021 | 0.021 | 0.024 | 0.062 | 0.086 | 0.036 | 0.019 | 0.052 |
| 1000 | 0.017 | 0.021 | 0.021 | 0.028 | 0.023 | 0.033 | 0.028 | 0.032 | 0.057 | 0.078 | 0.036 | 0.031 | 0.053 |
| 2000 | 0.020 | 0.027 | 0.025 | 0.030 | 0.031 | 0.034 | 0.038 | 0.043 | 0.057 | 0.072 | 0.040 | 0.036 | 0.054 |
| DGP3: $(1-L)^{0.4}x_t=(1+0.8L)e_t$ | | | | | | | | | | | | | |
| 500 | 0.011 | 0.018 | 0.014 | 0.017 | 0.018 | 0.024 | 0.024 | 0.028 | 0.062 | 0.084 | 0.032 | 0.027 | 0.057 |
| 1000 | 0.012 | 0.019 | 0.022 | 0.029 | 0.022 | 0.033 | 0.032 | 0.040 | 0.051 | 0.075 | 0.035 | 0.030 | 0.053 |
| 2000 | 0.019 | 0.023 | 0.027 | 0.031 | 0.029 | 0.033 | 0.035 | 0.038 | 0.058 | 0.070 | 0.039 | 0.032 | 0.046 |
| DGP4: $(1-0.5L)(1-0.3L)(1-L)^{0.4}x_t= e_t$ | | | | | | | | | | | | | |
| 500 | 0.009 | 0.013 | 0.005 | 0.009 | 0.010 | 0.018 | 0.014 | 0.036 | 0.063 | 0.086 | 0.025 | 0.044 | 0.063 |
| 1000 | 0.026 | 0.026 | 0.019 | 0.024 | 0.015 | 0.024 | 0.035 | 0.069 | 0.062 | 0.074 | 0.061 | 0.021 | 0.096 |
| 2000 | 0.046 | 0.042 | 0.033 | 0.033 | 0.026 | 0.028 | 0.044 | 0.073 | 0.057 | 0.067 | 0.089 | 0.012 | 0.097 |
| DGP5: $x_t=(1-L)^{-0.4}e_t+\eta_t$ | | | | | | | | | | | | | |
| 500 | 0.020 | 0.028 | 0.029 | 0.031 | 0.033 | 0.035 | 0.058 | 0.056 | 0.063 | 0.083 | 0.079 | 0.011 | 0.066 |
| 1000 | 0.029 | 0.028 | 0.036 | 0.043 | 0.046 | 0.053 | 0.059 | 0.054 | 0.055 | 0.076 | 0.091 | 0.007 | 0.063 |
| 2000 | 0.041 | 0.043 | 0.050 | 0.054 | 0.059 | 0.058 | 0.078 | 0.065 | 0.051 | 0.064 | 0.098 | 0.011 | 0.067 |
| DGP6: $(1-L)^{0.4}x_t= u_t$ with $u_t= \sigma_t e_t$, $\sigma_t^2=1+0.10u_{t-1}^2+0.85\sigma_{t-1}^2$ | | | | | | | | | | | | | |
| 500 | 0.019 | 0..023 | 0.023 | 0.024 | 0.032 | 0.033 | 0.032 | 0.033 | 0.054 | 0.100 | 0.038 | 0.007 | 0.059 |
| 1000 | 0.022 | 0.024 | 0.024 | 0.031 | 0.025 | 0.032 | 0.029 | 0.037 | 0.058 | 0.108 | 0.042 | 0.012 | 0.068 |
| 2000 | 0.023 | 0.029 | 0.025 | 0.036 | 0.029 | 0.038 | 0.036 | 0.037 | 0.056 | 0.108 | 0.040 | 0.025 | 0.057 |

Note: $\varepsilon$ is the trimming proportion. ORT: the test of Ohanissian et al. (2008); $W_c, \eta_\mu$ and $Z_t$: tests of Shimotsu (2006); mean-$t_d$: the test of Perron and Qu (2010); the mean-$t_d$ test is constructed using unfiltered series.

Table 3. Finite sample power of the tests at a 5% nominal level

| $n$ | Test statistics | | | | | | | Mean of $\hat{d}$ |
|---|---|---|---|---|---|---|---|---|
| | W ($\varepsilon=0.02$) | W ($\varepsilon=0.05$) | ORT | $W_c$ | $\eta_\mu$ | $Z_t$ | mean-$t_d$ | |
| Nonstationary random level shift | | | | | | | | |
| 500 | 0.19 | 0.16 | 0.09 | 0.17 | **0.37** | 0.00 | 0.19 | 0.23 |
| 1000 | 0.27 | 0.21 | 0.11 | 0.26 | **0.48** | 0.00 | 0.25 | 0.24 |
| 3000 | **0.76** | 0.48 | 0.26 | 0.51 | 0.74 | 0.00 | 0.51 | 0.25 |
| 5000 | **0.92** | 0.76 | 0.34 | 0.76 | 0.84 | 0.00 | 0.64 | 0.26 |
| 7000 | **0.97** | 0.89 | 0.44 | 0.75 | 0.89 | 0.00 | 0.74 | 0.26 |
| 9000 | **0.98** | 0.95 | 0.53 | 0.81 | 0.93 | 0.00 | 0.77 | 0.26 |
| Stationary random level shift | | | | | | | | |
| 500 | 0.21 | 0.20 | 0.08 | **0.33** | 0.32 | 0.00 | 0.14 | 0.24 |
| 1000 | 0.44 | 0.41 | 0.10 | **0.60** | 0.42 | 0.00 | 0.19 | 0.31 |
| 3000 | **0.92** | 0.89 | 0.11 | 0.82 | 0.33 | 0.00 | 0.15 | 0.42 |
| 5000 | **0.98** | 0.98 | 0.10 | 0.76 | 0.21 | 0.00 | 0.07 | 0.46 |
| 7000 | **1.00** | 1.00 | 0.14 | 0.71 | 0.13 | 0.00 | 0.04 | 0.49 |
| 9000 | **1.00** | 1.00 | 0.16 | 0.66 | 0.08 | 0.00 | 0.02 | 0.50 |
| Markov switching with iid regimes | | | | | | | | |
| 500 | 0.17 | 0.17 | 0.07 | **0.32** | 0.21 | 0.02 | 0.07 | 0.15 |
| 1000 | 0.46 | 0.46 | 0.09 | **0.59** | 0.36 | 0.01 | 0.16 | 0.25 |
| 3000 | **0.91** | 0.91 | 0.12 | 0.89 | 0.47 | 0.00 | 0.19 | 0.41 |
| 5000 | **0.98** | 0.98 | 0.10 | 0.79 | 0.41 | 0.00 | 0.15 | 0.45 |
| 7000 | **1.00** | 1.00 | 0.09 | 0.66 | 0.32 | 0.00 | 0.08 | 0.47 |
| 9000 | **1.00** | 1.00 | 0.08 | 0.51 | 0.22 | 0.00 | 0.05 | 0.49 |
| Markov switching with GARCH regimes | | | | | | | | |
| 500 | 0.02 | 0.02 | 0.07 | **0.12** | 0.08 | 0.06 | 0.06 | 0.14 |
| 1000 | 0.06 | 0.05 | 0.07 | 0.11 | **0.20** | 0.03 | 0.13 | 0.14 |
| 3000 | **0.47** | 0.23 | 0.15 | 0.20 | 0.42 | 0.01 | 0.34 | 0.17 |
| 5000 | **0.78** | 0.55 | 0.21 | 0.27 | 0.47 | 0.00 | 0.45 | 0.19 |
| 7000 | **0.92** | 0.77 | 0.23 | 0.29 | 0.48 | 0.00 | 0.47 | 0.19 |
| 9000 | **0.98** | 0.90 | 0.26 | 0.30 | 0.47 | 0.00 | 0.48 | 0.20 |
| White noise with a monotonic trend | | | | | | | | |
| 500 | 0.02 | 0.02 | 0.06 | 0.10 | **0.37** | 0.00 | 0.08 | 0.10 |
| 1000 | 0.03 | 0.02 | 0.07 | 0.12 | **0.43** | 0.00 | 0.14 | 0.11 |
| 3000 | 0.13 | 0.06 | 0.13 | 0.22 | **0.91** | 0.00 | 0.32 | 0.12 |
| 5000 | 0.44 | 0.21 | 0.19 | 0.35 | **0.99** | 0.00 | 0.44 | 0.13 |
| 7000 | 0.77 | 0.46 | 0.24 | 0.51 | **1.00** | 0.00 | 0.55 | 0.13 |
| 9000 | 0.93 | 0.73 | 0.30 | 0.62 | **1.00** | 0.00 | 0.62 | 0.13 |
| White noise with a non-monotonic trend | | | | | | | | |
| 500 | **0.83** | 0.51 | 0.09 | 0.05 | 0.01 | 0.00 | 0.11 | 0.41 |
| 1000 | **0.95** | 0.60 | 0.09 | 0.03 | 0.00 | 0.00 | 0.17 | 0.45 |
| 3000 | **1.00** | 0.80 | 0.12 | 0.01 | 0.06 | 0.00 | 0.27 | 0.44 |
| 5000 | **1.00** | 0.97 | 0.15 | 0.00 | 0.48 | 0.00 | 0.34 | 0.43 |
| 7000 | **1.00** | 0.99 | 0.16 | 0.00 | 0.95 | 0.00 | 0.39 | 0.42 |
| 9000 | **1.00** | 1.00 | 0.17 | 0.00 | 1.00 | 0.00 | 0.43 | 0.41 |

Note. $\varepsilon$: the trimming proportion. ORT: the test of Ohanissian et al. (2008); $W_c, \eta_\mu$ and $Z_t$: tests of Shimotsu (2006); mean-$t_d$: the test of Perron and Qu (2010); the mean-$t_d$ test is constructed using unfiltered series.

Table 1. Asymptotic critical values of the W test

| $\alpha$ | $\varepsilon=0.02$ | $\varepsilon=0.05$ |
|---|---|---|
| 10% | 1.118 | 1.022 |
| 5% | 1.252 | 1.155 |
| 2.5% | 1.374 | 1.277 |
| 1% | 1.517 | 1.426 |

Note: $\varepsilon$ is the trimming proportion.

Table 4. Test statistics values for the three time series

| Series | W ($\varepsilon=0.02$) | W ($\varepsilon=0.05$) | ORT | $W_c$ | $\eta_\mu$ | $Z_t$ | mean-$t_d$ |
|---|---|---|---|---|---|---|---|
| Temperature | 2.09** | 2.09** | 1.67(7.82, 11.35) | 11.55**(7.82, 11.35) | 1.56**(0.44,0.71) | 0.93(-2.92,-3.46) | 0.31(1.51,2.15) |
| Inflation | 1.51* | 1.46** | 1.82(7.82, 11.35) | 16.34**(7.82, 11.35) | 0.45*(0.44,0.71) | -0.62(-2.92,-3.46) | -1.11 (1.56,2.23) |
| Yen/$ volatility | 0.41 | 0.37 | 0.15(7.82, 11.35) | 2.18(7.82, 11.35) | 0.19(0.44,0.70) | -1.30(-2.85,-3.41) | 0.01(1.54,2.20) |

Note. * and ** denote significance at 5% and 1% level. The critical values of the W test are reported in Table 1. Other relevant critical values (at 5% and 1% level) are included in parentheses for reference. Note that the critical values of $\eta_\mu$, $Z_t$ and mean-$t_d$ are model-dependent.

Table 5. Rejection frequencies for simulations calibrated to empirical estimates

| Panel (a): Design based on the temperature series (n=1632) | | | | | | | |
|---|---|---|---|---|---|---|---|
| DGP | W ($\varepsilon=0.02$) | W ($\varepsilon=0.05$) | ORT | $W_c$ | $\eta_\mu$ | $Z_t$ | mean-$t_d$ |
| Temp-$H_0$ | 0.049 | 0.066 | 0.099 | 0.035 | 0.038 | 0.036 | 0.063 |
| Temp-$H_1$ | 0.894 | 0.742 | 0.256 | 0.106 | 0.990 | 0.000 | 0.487 |
| Panel (b): Design based on the inflation rate series (n=612) | | | | | | | |
| DGP | W ($\varepsilon=0.02$) | W ($\varepsilon=0.05$) | ORT | $W_c$ | $\eta_\mu$ | $Z_t$ | mean-$t_d$ |
| Infl-$H_0$ | 0.022 | 0.033 | 0.086 | 0.097 | 0.099 | 0.007 | 0.070 |
| Infl-$H_1$ | 0.274 | 0.302 | 0.112 | 0.313 | 0.067 | 0.000 | 0.213 |

Note. $\varepsilon$ is the trimming proportion. ORT: the test of Ohanissian et al. (2008); $W_c$, $\eta_\mu$ and $Z_t$: tests of Shimotsu (2006); mean-$t_d$: the test of Perron and Qu (2010); the mean-$t_d$ test is constructed using unfiltered series.

**Figure 1(a): Spectral domain properties of processes with level shifts**
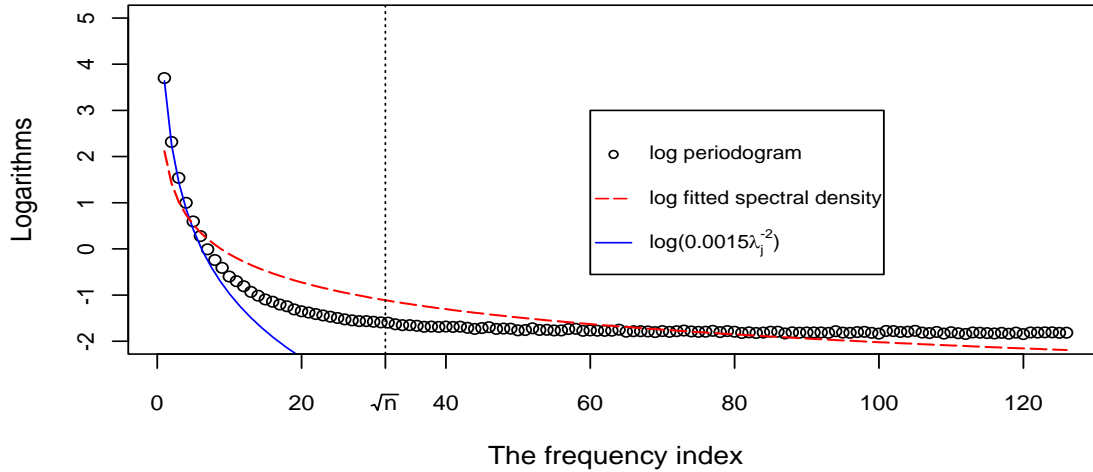


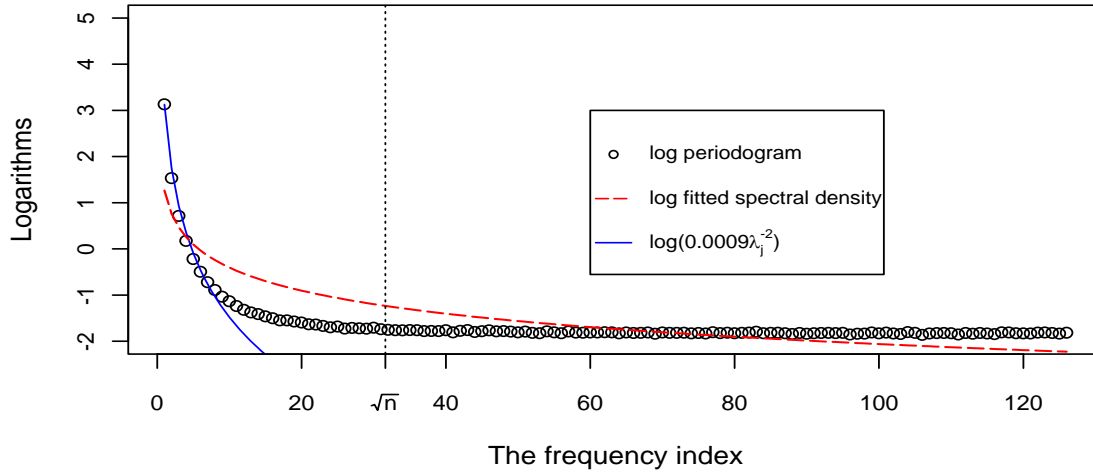**Figure 1(b): Local Whittle estimates with different m for processes with level shifts**
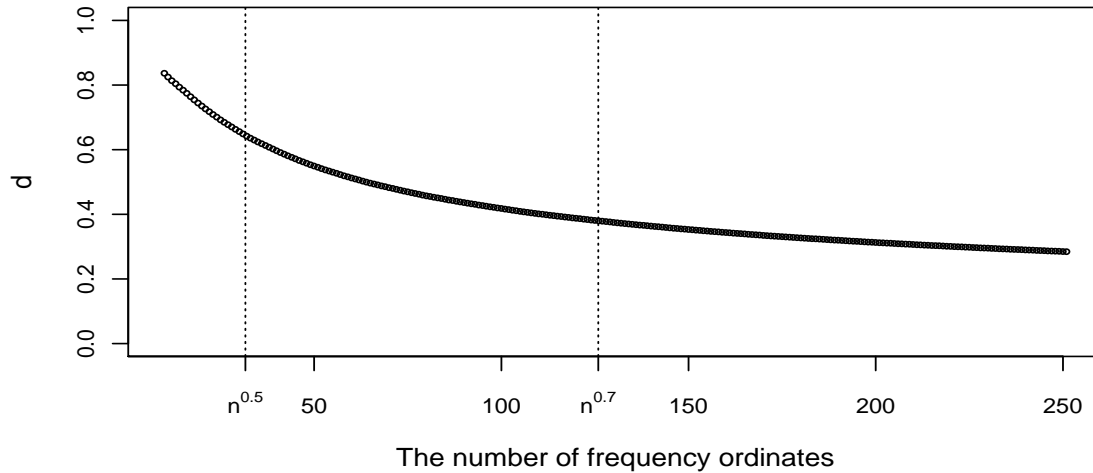


Note. The DGP is (3) with $z_t, \eta_t \sim i.i.d.N(0,1)$ and $\pi_t \sim i.i.d.B(1, 5/n)$. The sample size n=1000 and the results reported are empirical means based on 5000 replications. Figure 1(a) depicts the log periodograms as a function of the frequency index. An extra line, $\log(0.0015\lambda_j^{-2})$, is superimposed to highlight the slope of the periodogram when $j = o(n^{1/2})$. The dashed line is the estimated spectral density function, imposing the false restriction that the series are of true long memory. Figure 1(b) reports the local Whittle estimates using different bandwidths $m$ ranging between $n^{1/3}$ and $n^{0.8}$.

**Figure 2(a): Spectral domain properties of processes with smooth trends**



The frequency index

**Figure 2(b): Local Whittle estimates with different m for processes with smooth trends**



The number of frequency ordinates
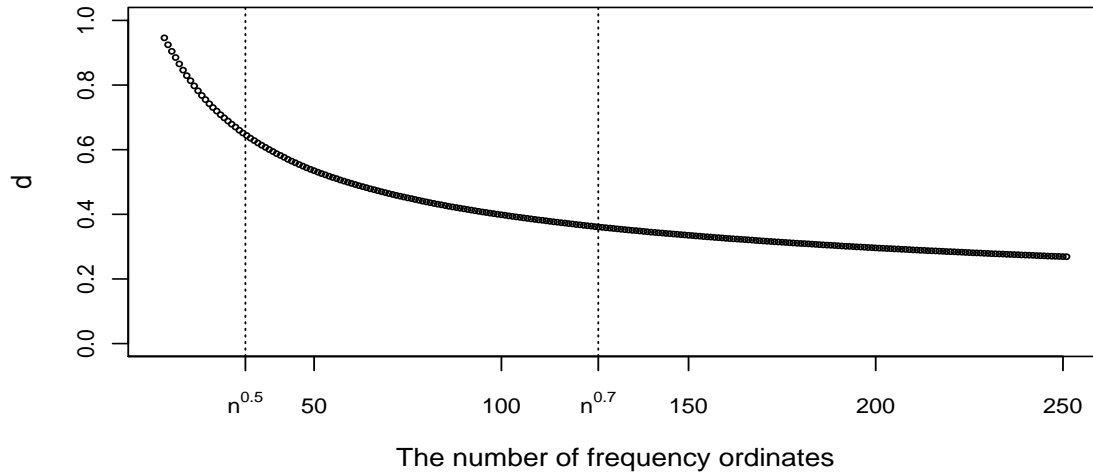
Note. The DGP is (4) with $h(x) = 2x - 4x^2$ and $z_t$ being $i.i.d. N(0, 1)$. The sample size n=1000 and the results reported are empirical means based on 5000 replications. Figure 2(a) depicts the log periodograms as a function of the frequency index. An extra line, $\log(0.0009\lambda_j^{-2})$, is superimposed to highlight the slope of the periodogram when $j = o(n^{1/2})$. The dashed line is the estimated spectral density function, imposing the false restriction that the series are of true long memory. Figure 2(b) reports the local Whittle estimates using different bandwidths $m$ ranging between $n^{1/3}$ and $n^{0.8}$.

# Figure 3: Results for the northern hemisphere temperature series

## (a) The series: 1854-1989



## (b) Memory parameter estimates with different numbers of frequency ordinates (m)



The number of frequency ordinates

## (c) Memory parameter estimates using different subsamples



The size of the subsample
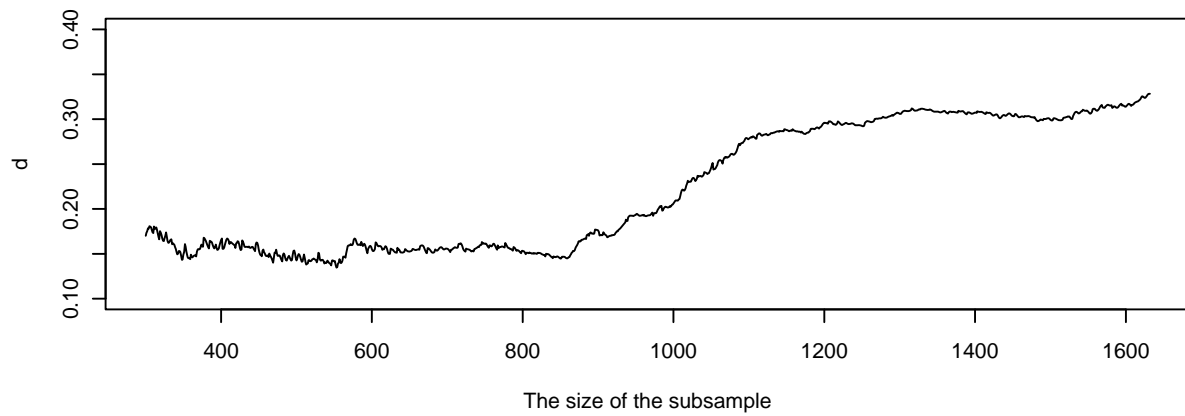
# Figure 4: Results for the US inflation rate series

## (a) The series: 1958-2008
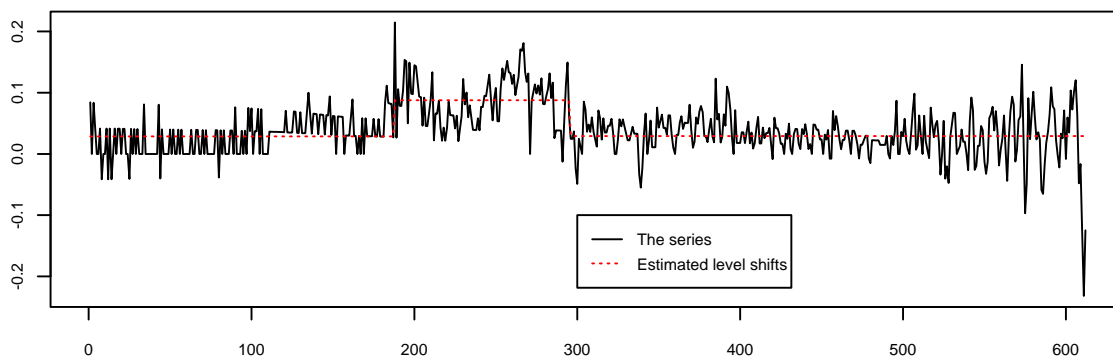


## (b) The memory parameter estimates with different numbers of frequency ordinates (m)
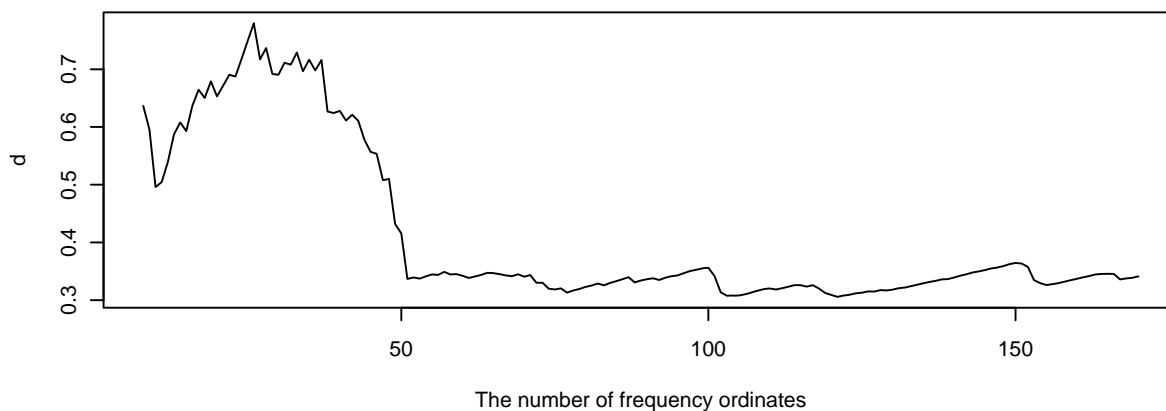


## (c) The memory parameter estimates using different subsamples
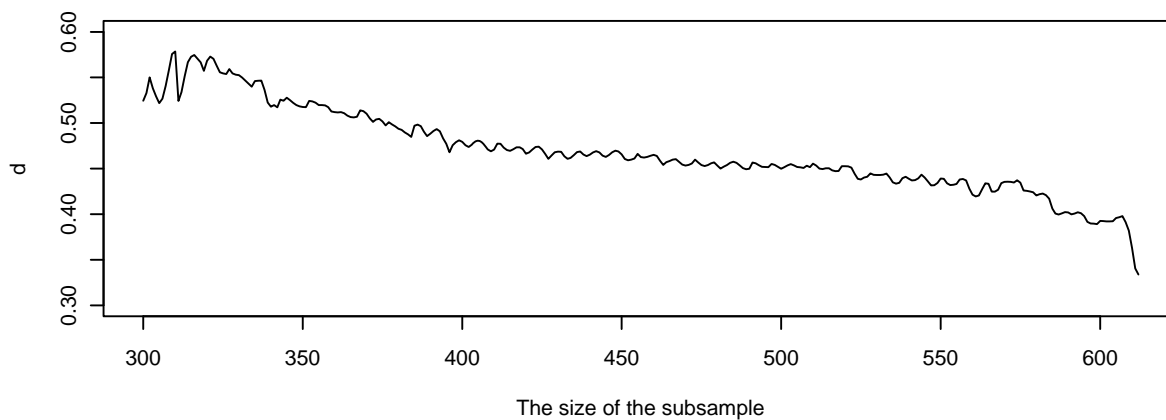
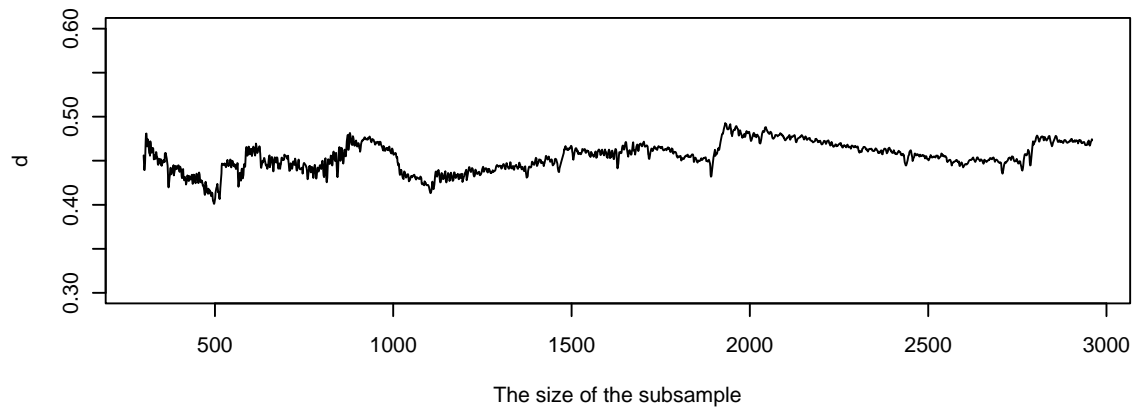# Figure 5: Results for the Yen/Dollar Spot Exchange Rate Series

## (a) The log realized volatility: 1986-1999
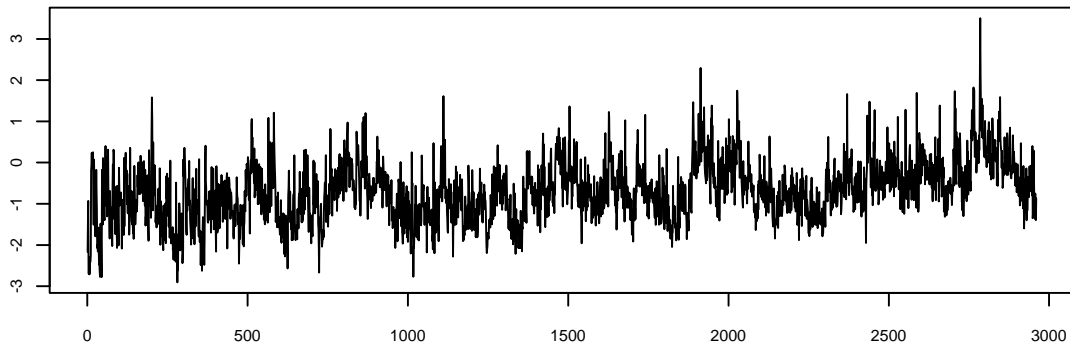


## (b) Memory parameter estimates with different numbers of frequency ordinates (m)
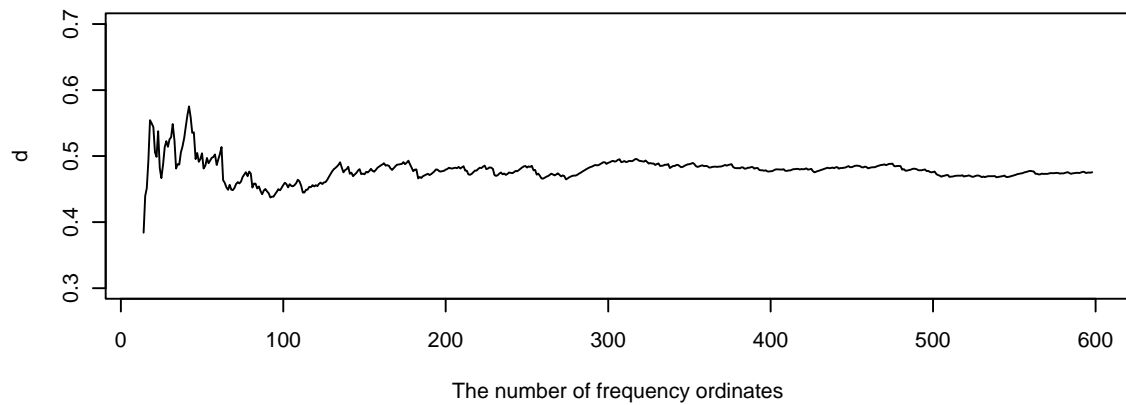


The number of frequency ordinates

## (c) Memory parameter estimates using different subsamples



The size of the subsample

This appendix will be made available electronically. It contains proofs for Lemma 2, Corollary 1 and 2 and some auxiliary lemmas. As a matter of notation, $f_{\xi j}$ and $I_{\xi j}$ denote the spectral density and periodogram of $\xi_t$ at frequency $\lambda_j$. $f_j$ and $I_j$ are short for $f_{xj}$ and $I_{xj}$ with $x_t$ being the long memory process defined in the paper.

**Proof of Lemma 2:** We first establish a uniform weak law of large numbers. For any fixed $r_1 \in [0,1]$, $Q_m(r_1) \rightarrow^p r_1$ by Robinson (1995b, p. 1638) and Shao and Wu (2007, p. 920, A.13). Using Theorem 2.1 of Newey (1991), it then suffices to show that the process $Q_m(r)$ is stochastically equicontinuous on $[0,1]$, i.e., for an arbitrary $r \in [0,1]$ and any $\epsilon, \eta > 0$, there exists a $\delta > 0$, such that $P\left( \sup_{s \in [\delta]_r} |Q_m(r) - Q_m(s)| > \eta \right) < \epsilon$ for large $m$, where $[\delta]_r = \{s : 0 \leq s \leq 1, |s - r| \leq \delta\}$. To show this, let $[\delta]_r^+$ denote the elements in the set $[\delta]_r$ that are no less than $r$. Then,

$$
P\left( \sup_{s \in [\delta]_r^+} \frac{1}{m} \sum_{j=[mr]+1}^{[ms]} I_j \lambda_j^{2d_0} > \eta \right) \leq P\left( \frac{1}{m} \sum_{j=[mr]+1}^{[m(r+\delta)]} I_j \lambda_j^{2d_0} > \eta \right) \leq \frac{1}{m\eta} \sum_{j=[mr]+1}^{[m(r+\delta)]} E(I_j \lambda_j^{2d_0})
$$

$$
\leq \frac{1}{m\eta} \sum_{j=[mr]+1}^{[m(r+\delta)]} C \rightarrow \frac{\delta C}{\eta}, \quad \text{for some } 0 < C < \infty,
$$

where the first inequality holds because the summands are non-negative, the second is due to Markov's inequality, and the third is due to Theorem 2 of Robinson (1995a) and Shao and Wu (2007, p. 920, A.13). $\delta C/\eta$ can be made arbitrarily small by choosing a small $\delta$. This proves the first result.

Let $F_m(r) = m^{1/2}(Q_m(r) - r)$. We will first show finite dimensional convergence and then prove tightness. For the former, it suffices to show that for $0 \leq r_1 \leq r_2 \leq 1$,

$$
F_m(r_1) \longrightarrow^d N(0, r_1) \text{ and } Cov(F_m(r_1), F_m(r_2)) \rightarrow r_1. \tag{B.1}
$$

The first expression of (B.1) follows immediately. Consider the second result. Lemma B.7 shows that $F_m(r)$ can be approximated by $m^{-1/2} \sum_{j=1}^{[mr]} (2\pi I_{\varepsilon j} - 1)$ with the approximation error being uniformly $o_p(1)$ over $r \in [0,1]$. Note that

$$
m^{-1/2} \sum_{j=1}^{[mr]} (2\pi I_{\varepsilon j} - 1) = \sum_{t=1}^{n} z_{t,r}^*
$$

with

$$z_{t,r}^* \equiv \varepsilon_t \sum_{k=1}^{t-1} b_{t-k,r}\varepsilon_k \quad \text{and} \quad b_{t,r} = 2n^{-1}m^{-1/2}\sum_{j=1}^{[mr]}\cos(t\lambda_j). \tag{B.2}$$

Thus, it suffices to analyze $Cov(\sum_{t=1}^n z_{t,r_1}^*, \sum_{t=1}^n z_{t,r_2}^*)$ :

$$\begin{aligned}
E\left(\sum_{t=1}^n \sum_{s=1}^{t-1} b_{t-s,r_1}\varepsilon_s \sum_{l=1}^{t-1} b_{t-l,r_2}\varepsilon_l\right) &= \frac{4}{n^2 m}\sum_{j=1}^{[mr_1]}\sum_{l=1,l\neq j}^{[mr_2]}\sum_{t=1}^n\sum_{s=1}^{t-1}\cos((t-s)\lambda_j)\cos((t-s)\lambda_l) \\
&\quad + \frac{4}{n^2 m}\sum_{j=1}^{[mr_1]}\sum_{t=1}^n\sum_{s=1}^{t-1}\cos^2((t-s)\lambda_j) \\
&= \frac{4}{n^2 m}\sum_{j=1}^{[mr_1]}\sum_{l=1,l\neq j}^{[mr_2]}\left(-\frac{n}{2}\right) + \frac{4}{n^2 m}\sum_{j=1}^{[mr_1]}\frac{n^2}{4} \to r_1.
\end{aligned}$$

To prove tightness, we follow Nielsen (2004) and use Theorem 13.5 of Billingsley (1999). That is, we show that for every $m$ and $r_1 \leq r \leq r_2$,

$$E\left(\left|\sum_{t=1}^n z_{t,r}^* - \sum_{t=1}^n z_{t,r_1}^*\right|^2 \left|\sum_{t=1}^n z_{t,r_2}^* - \sum_{t=1}^n z_{t,r}^*\right|^2\right) \leq K(\psi_m(r_2) - \psi_m(r_1))^2, \tag{B.3}$$

where $K$ is some finite constant and $\psi_m(.)$ is a function on $[0,1]$ that is finite, nondecreasing, and satisfies $\lim_{\delta\to 0}\limsup_{m\to\infty}|\psi_m(s+\delta) - \psi_m(s)| \to 0$ uniformly in $s \in [0,1]$. To verify this, we apply Lemma B.8, leading to

$$E\left(\left|\sum_{t=1}^n z_{t,r}^* - \sum_{t=1}^n z_{t,r_1}^*\right|^2 \left|\sum_{t=1}^n z_{t,r_2}^* - \sum_{t=1}^n z_{t,r}^*\right|^2\right) \leq C\left(\sum_{t=1}^n\sum_{s=1}^{t-1}b_{t-s}(r_1,r)^2\right)\left(\sum_{k=1}^n\sum_{l=1}^{t-1}b_{k-l}(r,r_2)^2\right),$$

where $C$ is some constant, $b_{t-s}(r_1,r) = b_{t-s,r} - b_{t-s,r_1}$, and $b_{k-l}(r,r_2)$ is defined analogously. Consider the first term:

$$\begin{aligned}
\sum_{t=1}^n\sum_{s=1}^{t-1}b_{t-s}(r_1,r)^2 &= \frac{4}{n^2 m}\sum_{t=1}^n\sum_{s=1}^{t-1}\left(\sum_{j=[mr_1]+1}^{[mr]}\cos((t-s)\lambda_j)\right)^2 \\
&= \frac{4}{n^2 m}\sum_{j=[mr_1]+1}^{[mr]}\sum_{t=1}^n\sum_{s=1}^{t-1}\cos^2((t-s)\lambda_j) \\
&\quad + \frac{4}{n^2 m}\sum_{j=[mr_1]+1}^{[mr]}\sum_{k=[mr_1]+1,k\neq j}^{[mr]}\sum_{t=1}^n\sum_{s=1}^{t-1}\cos((t-s)\lambda_j)\cos((t-s)\lambda_k) \\
&\leq \frac{2}{m}([mr] - [mr_1]).
\end{aligned}$$

B-2

The second term can be bounded analogously. Hence, the left hand side of (B.3) is bounded from the above by $C \left\{ \frac{2}{m} \left( [mr_2] - [mr_1] \right) \right\}^2$. Let $\psi_m(s) = [ms]/m$. The proof is complete. ∎

**Proof of Corollary 1**. The proof is similar to that of Theorem 2. The argument preceding (A.8) goes through without modification. Thus we only need to consider the second term in (A.8). It is easy to verify that the true long memory component dominates the level shift or the trend component for $j \in [j^*, m]$. Thus, $I_j = O_p(\lambda_j^{-2d_0})$ for $j \in [j^*, m]$. Consequently, $v_j I_j / (G(\hat{d}) \lambda_j^{-2\hat{d}}) = O_p(\lambda_j^{2(\hat{d}-d_0)}) = O_p(\lambda_j^{2\epsilon}) = o_p(1)$, where the first equality uses $G(\hat{d}) = m^{-1} \sum_{j=1}^m I_j \lambda_j^{2\hat{d}}$ and the variance condition in Corollary 1, the second equality is due to $P(\hat{d} - d_0 > \epsilon) \to 1$, and the last equality is because $\lambda_j = o(1)$. This implies $A \to^p \infty$ as $n \to \infty$. ∎

**Proof of Corollary 2**. Because $\hat{\mu} - \mu_0 = O_p(n^{d_0-1/2}) = o_p(1)$ and the correction involves an ARMA(1,1) filter, replacing $\mu_0$ by $\hat{\mu}$ has no effect on the asymptotic property of the procedure. Without loss of generality, we can further assume $\mu_0$ is zero because the Fourier transform is invariant to the mean when evaluated at non-zero frequencies.

First suppose Assumptions 1-4 and F(a) hold for the original series. Let $\phi(L) = (1 - \hat{a}_1 L)/(1 + \hat{b}_1 L)$ and $\psi(L) = \sum_{j=0}^\infty \alpha_j L^j$. It suffices to analyze the process $x_t^* = \phi(L)\psi(L)\varepsilon_{t-j}$. Because $-1 + \delta \le \hat{a}_1, \hat{b}_1 \le 1 - \delta$, the power transfer function $||\phi(e^{-i\lambda})||^2$ satisfies

$$\left\| \phi(e^{-i\lambda}) \right\|^2 = O(1 + \lambda^2) \quad \text{as } \lambda \to 0+ . \tag{B.4}$$

Thus Assumption 1 holds for the transformed process. To verify Assumption 2, we note that

$$x_t^* = \sum_{j=0}^\infty \alpha_j^* \varepsilon_{t-j} \quad \text{with} \quad |\alpha_j^*| \le |\alpha_j| + 2 \sum_{s=0}^{j-1} |\alpha_{j-s-1}| (1-\delta)^s .$$

Because $\alpha_j = O(j^{-1/2-c})$ for $j \to \infty$, there exists a constant M, such that $|\alpha_j| \le M j^{-1/2-c}$ for all $j > 0$. Hence,

$$
\begin{aligned}
|\alpha_j^*| &\le M j^{-1/2-c} + 2M \sum_{s=0}^{j-2} (j-s-1)^{-1/2-c} (1-\delta)^s + 2M (1-\delta)^{j-1} \\
&\le 2M j^{-1/2-c} \left\{ 1 + \sum_{s=0}^{j-2} \left| 1 - \frac{s+1}{j} \right|^{-1/2-c} (1-\delta)^s \right\} + 2M (1-\delta)^{j-1} .
\end{aligned}
$$

The summation inside of the curly brackets is bounded for all j, because $(1-\delta)^s$ decreases to zero at an exponential rate as $s$ increases, while $|1 - (s+1)/j|^{-1/2-c}$ diverges only hyperbolically.

The second term, $2M(1-\delta)^{j-1}$, is of lower order than $j^{-1/2-c}$ for the same reason. Thus, $|\alpha_j^*| = O(j^{-1/2-c})$ as $j \to \infty$ and Assumption 2 is satisfied. Assumption 3 can be verified by differentiating $\phi(e^{-i\lambda})\psi(e^{-i\lambda})$ and applying the chain rule.

Now suppose Assumption 1, H and F(b) hold for the original series. The filtered series satisfies Assumption 1 because of (B.4). It also satisfies F(b) because the GMC property is preserved under ARMA filter. Because F(b) implies H(a) and H(b), Assumption F continues to hold. This completes the proof. ∎

**Lemma B.1** $(1/m)\sum_{j=1}^{[mr]} v_j = \int_0^r (1+\log x)dx + O(\frac{1}{m^{1-\epsilon}})$ and $(1/m)\sum_{j=1}^{[mr]} v_j^2 = \int_0^r (1+\log x)^2 dx + O(\frac{1}{m^{1-\epsilon}})$ uniformly in $r \in [0,1]$, where $\epsilon$ is some arbitrarily small positive number.

**Proof:** We prove the result using the Euler-Maclaurin formula which states that if $k$ is a natural number and $g(\cdot)$ is a twice differentiable function defined for all real numbers between 0 and k, then the following equality holds:

$$\sum_{j=1}^{k} g(j) = \int_1^k g(x)dx + \frac{g(1)+g(k)}{2} + \frac{1}{12}(g'(k)-g'(1)) + R,$$

where R satisfies $|R| \le \frac{2}{(2\pi)^2}\int_1^k |g''(x)|\,dx$. Let $k = [mr]$ and apply the above result with

$$g(x) = \log(\frac{x}{m}),$$

then

$$\frac{1}{m}\sum_{j=1}^{[mr]} \log\frac{j}{m} = \int_{1/m}^r \log x dx + \int_r^{\frac{[mr]}{m}} \log x dx + \frac{\log[mr]-2\log m}{2m} + \frac{1}{12m}\left(\frac{1}{[mr]}-1\right) + \frac{1}{m}R,$$

where $|R| \le \frac{2}{(2\pi)^2}\int_1^{[mr]}\frac{1}{x^2}dx = O(1)$. Thus, $\frac{1}{m}\sum_{j=1}^{[mr]}\log\frac{j}{m} = \int_{1/m}^r \log x dx + O(\frac{\log m}{m}) = \int_0^r \log x dx + O(\frac{1}{m^{1-\epsilon}})$, where the last equality is due to $\int_0^{1/m}|\log x|\,dx = \int_m^\infty \left(\frac{|\log x|}{x^\epsilon}\right)(\frac{1}{x^{2-\epsilon}})dx = O(m^{\epsilon-1})$. Hence,

$$\frac{1}{m}\sum_{j=1}^{[mr]} v_j = \int_0^r \log x dx - r\int_0^1 \log x dx + O(\frac{1}{m^{1-\epsilon}}) = \int_0^r (1+\log x)dx + O(\frac{1}{m^{1-\epsilon}}).$$

The second result can be proved in a similar way. ∎

**Lemma B.2** *Let $\hat{d}$ denote the local Whittle estimate of $d$ and $\delta(d_0, \hat{d})$ the interval between $d_0$ and $\hat{d}$. Under Assumptions 1-4 or Assumption H, for k=0,1, and 2, we have*

$$\sup_{d \in \delta(d_0,\hat{d})} \sup_{r \in [0,1]} \left| \frac{1}{m} \sum_{j=1}^{[mr]} j^{2d_0} I_j \left( \log j \right)^k - \frac{1}{m} \sum_{j=1}^{[mr]} j^{2d} I_j \left( \log j \right)^k \right| = o_p(n^{2d_0}).$$

**Proof:** The arguments follow Robinson (1995b, pp.1642-1643). Fix $\epsilon > 0$ and choose $n$ and $m$ such that $(\log m)^2 > 2\epsilon$. Define $M = \left\{ d \colon (\log m)^5 |d - d_0| \leq \epsilon \right\}$. Then, for $\eta > 0$

$$P \left( \sup_{d \in \delta(d_0,\hat{d})} \sup_{r \in [0,1]} \frac{1}{m} \left| \sum_{j=1}^{[mr]} j^{2d_0} I_j \left( \log j \right)^k - \sum_{j=1}^{[mr]} j^{2d} I_j \left( \log j \right)^k \right| > \eta \left( \frac{2\pi}{n} \right)^{-2d_0} \right)$$

$$\leq \quad P(\hat{d} \notin M) + P \left( \sup_{r \in [0,1], d \in M} \frac{1}{m} \left| \sum_{j=1}^{[mr]} j^{2d_0} I_j \left( \log j \right)^k - \sum_{j=1}^{[mr]} j^{2d} I_j \left( \log j \right)^k \right| > \eta \left( \frac{2\pi}{n} \right)^{-2d_0} \right).$$

The first term is $o(1)$ by Robinson (1995b), with a correction by Andrews and Sun (2004, p.600), and Shao and Wu (2007). For the second term,

$$\sup_{r \in [0,1], d \in M} m^{-1} \left| \sum_{j=1}^{[mr]} j^{2d_0} I_j \left( \log j \right)^k - \sum_{j=1}^{[mr]} j^{2d} I_j \left( \log j \right)^k \right| \leq \sup_{d \in M} m^{-1} \sum_{j=1}^{m} j^{2d_0} I_j \left( \log j \right)^k \left| 1 - j^{2(d-d_0)} \right|$$

$$\leq \quad 2e\epsilon \left( \log m \right)^{k-2} m^{-1} \sum_{j=1}^{m} j^{2d_0} I_j = 2e\epsilon \left( \log m \right)^{k-2} \left( \frac{2\pi}{n} \right)^{-2d_0} (G_0 + o_p(1)),$$

where the last equality follows from $m^{-1} \sum_{j=1}^{m} \lambda_j^{2d_0} I_j \to^p G_0$. The second term can be made arbitrarily small by choosing a small $\epsilon$. This completes the proof. ∎

**Lemma B.3** *Let $\hat{d}$ denote the local Whittle estimate of $d$ and $\delta(d_0, \hat{d})$ the interval between $d_0$ and $\hat{d}$. Under Assumptions 1-4 or Assumption H, for k=0,1, and 2, we have*

$$\sup_{d \in \delta(d_0,\hat{d})} \sup_{r \in [0,1]} \left| \frac{1}{m} \sum_{j=1}^{[mr]} v_j^k \frac{I_j}{G(d)\lambda_j^{-2d}} - \frac{1}{m} \sum_{j=1}^{[mr]} v_j^k \right| = o_p \left( 1 \right). \tag{B.5}$$

**Proof.** We have, uniformly in $r \in [0,1]$,

$$\frac{1}{m} \sum_{j=1}^{[mr]} v_j^k \frac{I_j}{G(d)\lambda_j^{-2d}}$$

$$= \quad \frac{m^{-1} \sum_{j=1}^{[mr]} v_j^k I_j \lambda_j^{2d}}{m^{-1} \sum_{j=1}^{m} I_j \lambda_j^{2d}} = \frac{m^{-1} \sum_{j=1}^{[mr]} v_j^k I_j j^{2d_0} + o_p(n^{2d_0})}{m^{-1} \sum_{j=1}^{m} I_j j^{2d_0} + o_p(n^{2d_0})} = \frac{1}{m} \sum_{j=1}^{[mr]} v_j^k \frac{I_j}{G_0 \lambda_j^{-2d_0}} + o_p(1),$$

B-5

where the first equality is because $G(d) = m^{-1} \sum_{j=1}^{m} \lambda_j^{2d} I_j$, the second equality follows from Lemma B.2, and the last equality follows from $m^{-1} \sum_{j=1}^{m} \lambda_j^{2d_0} I_j \rightarrow^p G_0$. The above result implies

$$\sup_{d \in \delta(d_0, \hat{d})} \sup_{r \in [0,1]} \left| \frac{1}{m} \sum_{j=1}^{[mr]} v_j^k \frac{I_j}{G(d)\lambda_j^{-2d}} - \frac{1}{m} \sum_{j=1}^{[mr]} v_j^k \right| = \sup_{r \in [0,1]} \left| \frac{1}{m} \sum_{j=1}^{[mr]} v_j^k \left( \frac{I_j}{G_0 \lambda_j^{-2d_0}} - 1 \right) \right| + o_p(1).$$

Using summation by parts, the leading term on the right hand side is bounded from above by

$$\sup_{r \in [0,1]} \left( \frac{1}{m} \sum_{s=1}^{[mr]-1} \left| v_{s+1}^k - v_s^k \right| \left| \sum_{j=1}^{s} \left( \frac{I_j}{G_0 \lambda_j^{-2d_0}} - 1 \right) \right| + \frac{1}{m} (\log [mr])^k \left| \sum_{j=1}^{[mr]} \left( \frac{I_j}{G_0 \lambda_j^{-2d_0}} - 1 \right) \right| \right).$$

The first term in the preceding expression is bounded by (see Robinson 1995b, p.1643)

$$\frac{1}{m} \sum_{s=1}^{m} \frac{\log(s+1) + 2}{s} \left| \sum_{j=1}^{s} \left( \frac{I_j}{G_0 \lambda_j^{-2d_0}} - 1 \right) \right| = O_p \left( \frac{(\log m)^2}{m^{1/2}} \right).$$

For the second term, using Lemma 2, we have

$$\sup_{r \in [0,1]} \left( \frac{1}{m} (\log [mr])^k \left| \sum_{j=1}^{[mr]} \left( \frac{I_j}{G_0 \lambda_j^{-2d_0}} - 1 \right) \right| \right)$$

$$= \sup_{r \in [0,1]} \frac{(\log [mr])^2}{m^{1/2}} |W(r)| + o_p \left( \frac{(\log [mr])^2}{m^{1/2}} \right) = O_p \left( \frac{(\log m)^2}{m^{1/2}} \right).$$

This proves (B.5). ∎

The next lemma concerns the behavior of the periodograms under Assumption H. Let $Z_k = \sum_{t=0}^{\infty} E(\xi_{t+k} | \mathcal{F}_k)$ and

$$\varepsilon_k = \frac{1}{2\pi G_0} (Z_k - E(Z_k | \mathcal{F}_{k-1})). \tag{B.6}$$

Then, $\{\varepsilon_k\}_{k=0}^{\infty}$ forms a martingale difference sequence in $\mathcal{L}^q$ with $E(\varepsilon_k^2) = 2\pi f_\xi(0)/(2\pi G_0) \simeq 1$.

**Lemma B.4** *Under Assumption H, for any fixed $r \in (0, 1]$.*

$$m^{-1/2} \sum_{j=1}^{[mr]} v_j \left( \frac{I_j}{G_0 \lambda_j^{-2d_0}} - 1 \right) - m^{-1/2} \sum_{j=1}^{[mr]} v_j (2\pi I_{\varepsilon j} - 1) = o_p(1).$$

**Proof:** This is immediate from Theorem 3.1 in Shao and Wu (2007). ∎

The next result proves tightness and is needed to strengthen the pointwise convergence result in Lemma (B.4) to uniform convergence.

B-6

**Lemma B.5** *Under Assumption H, for any $0 < \delta < 1$,*

$$\sup_{r \in [\delta, 1]} \left| m^{-1/2} \sum_{j=1}^{[mr]} v_j \left( \frac{I_j}{G_0 \lambda_j^{-2d_0}} - 2\pi I_{\varepsilon j} \right) \right| = O_p(1)$$

**Proof:** We have, for any $\delta \in (0, r]$,

$$\left| m^{-1/2} \sum_{j=1}^{[mr]} v_j \left( \frac{I_j}{G_0 \lambda_j^{-2d_0}} - 2\pi I_{\varepsilon j} \right) \right|$$

$$\leq \left| m^{-1/2} \sum_{j=[m\delta]+1}^{[mr]} v_j \left( \frac{I_j}{G_0 \lambda_j^{-2d_0}} - 2\pi I_{\varepsilon j} \right) \right| + \left| m^{-1/2} \sum_{j=1}^{[m\delta]} v_j \left( \frac{I_j}{G_0 \lambda_j^{-2d_0}} - 2\pi I_{\varepsilon j} \right) \right|$$

$$= \left| m^{-1/2} \sum_{j=[m\delta]+1}^{[mr]} v_j \left( \frac{I_j}{G_0 \lambda_j^{-2d_0}} - 2\pi I_{\varepsilon j} \right) \right| + o_P(1),$$

where the first inequality is due to the triangle inequality, and the $o_p(1)$ term is due to Lemma B.4.

Upon recursive substitution, we have

$$\left| \frac{1}{\sqrt{m}} \sum_{j=[m\delta]+1}^{[mr]} v_j \left( \frac{I_j}{G_0 \lambda_j^{-2d_0}} - 2\pi I_{\varepsilon j} \right) \right|$$

$$\leq \frac{1}{\sqrt{m}} \sum_{j=[m\delta]+1}^{[mr]} \left| v_j \frac{I_j}{f_j} \left( \frac{f_j}{G_0 \lambda_j^{-2d_0}} - 1 \right) + v_j \left( \frac{I_j}{f_j} - \frac{I_{\xi j}}{f_{\xi j}} \right) + v_j \left( \frac{I_{\xi j}}{f_{\xi j}} - 2\pi I_{\varepsilon j} \right) \right|$$

$$\leq \frac{1}{\sqrt{m}} \sum_{j=[m\delta]+1}^{m} |v_j| \frac{I_j}{f_j} \left| \frac{f_j}{G_0 \lambda_j^{-2d_0}} - 1 \right| + \frac{1}{\sqrt{m}} \sum_{j=[m\delta]+1}^{m} |v_j| \left| \frac{I_j}{f_j} - \frac{I_{\xi j}}{f_{\xi j}} \right| + \frac{1}{\sqrt{m}} \sum_{j=[m\delta]+1}^{m} |v_j| \left| \frac{I_{\xi j}}{f_{\xi j}} - 2\pi I_{\varepsilon j} \right|$$

$$= (a) + (b) + (c),$$

where the last inequality is again due to the triangle inequality. For term (a),

$$\Pr \left( (a) > C_1 \right)$$

$$\leq \frac{1}{C_1 m^{1/2}} \sum_{j=1}^{m} |v_j| E \left( \frac{I_j}{f_j} \right) \left| \frac{f_j}{G_0 \lambda_j^{-2d_0}} - 1 \right|$$

$$\leq \frac{C_2}{C_1 m^{1/2}} \sum_{j=1}^{m} |v_j| \left| \frac{f_j}{G_0 \lambda_j^{-2d_0}} - 1 \right|$$

$$\leq \frac{C_2}{C_1 m^{1/2}} \sum_{s=1}^{m-1} |v_{s+1} - v_s| \sum_{j=1}^{s} \left| \frac{f_j}{G_0 \lambda_j^{-2d_0}} - 1 \right| + \frac{C_2}{C_1 m^{1/2}} (\log m) \sum_{j=1}^{m} \left| \frac{f_j}{G_0 \lambda_j^{-2d_0}} - 1 \right|$$

where the first inequality is due to Markov's inequality with $C_1$ being an arbitrary constant, the

second inequality is because $E(I_j/f_j) \leq C_2$ uniformly over $j = 0, ..., m$ (c.f. Lemma A.2 in Shao and

Wu, 2007), and the third inequality is due to summation by parts. Because $|v_{s+1} - v_s| \le 1/s$ and $f_j/(G_0\lambda_j^{-2d_0}) - 1 = O(\lambda_j^\beta)$ by Assumption 1, the first term in the last line of the display is of order $O(m^{\beta+1/2}/n) = o(1)$ in view of Assumption H. The second term is of order $O(m^{\beta+1/2}\log(m)/n) = o(1)$. Thus, term (a) is $O_p(1)$. For term (b)

$$
\begin{aligned}
\Pr\left((b) > C_1\right) &\le \frac{1}{C_1 m^{1/2}} \sum_{j=[m\delta]+1}^{m} |v_j| E\left|\frac{I_j}{f_j} - \frac{I_{\xi j}}{f_{\xi j}}\right| \\
&\le \frac{C_3}{C_1 m^{1/2}} \sum_{j=[m\delta]+1}^{m} j^{-1/2} |v_j| \le \frac{C_3}{C_1 m^{1/2} [m\delta]^{1/2}} \sum_{j=[m\delta]+1}^{m} |v_j|,
\end{aligned}
$$

where the second inequality uses Lemma A.3. in Shao and Wu (2007), stating that $E\left|\frac{I_j}{f_j} - \frac{I_{\xi j}}{f_{\xi j}}\right| = O(j^{-1/2})$ uniformly over j=1,...,m. The preceding display can be made arbitrarily small by choosing a large $C_1$, implying term (b) is $O_p(1)$. Term (c) can be analyzed similarly and is also $O_p(1)$. ∎

**Lemma B.6** *Under Assumptions 1-4, or Assumption H, Lemma B.4 holds with the approximation error being uniformly $o_p(1)$ over $r \in [\delta, 1]$ for any $0 < \delta < 1$.*

**Proof:** Under Assumptions H, Lemma B.4 implies finite dimensional convergence and Lemma B.5 tightness. The uniformity follows. Under Assumptions 1-4, Robinson (1995b) proved Lemma B.4, see Eq. (4.8) and Eq. (4.11). It is easy to verify that Lemma B.5 still holds. This completes the proof. ∎

**Lemma B.7** *Lemma B.4-B.6 hold when $v_j$ and $\delta$ are replaced by 1 and 0, respectively.* ∎

**Proof:** Lemma B.4, with $v_j$ replaced by 1, is proved in Nielsen (2004, p.157). The proof of the rest involves repeating the same argument as in Lemma B.5 and B.6. The detail is omitted.

**Lemma B.8** *Let $z_{t,r} = \varepsilon_t \sum_{s=1}^{t-1} c_{t-s,r}\varepsilon_s$ with $c_{t,r} = 2n^{-1}m^{-1/2}\sum_{j=1}^{[mr]} v_j \cos(t\lambda_j)$ and $\varepsilon_t$ defined in Assumption 2 or (B.6). Then, for $0 \le r_1 \le r \le r_2 \le 1$, we have*

$$
E\left(\left|\sum_{t=1}^{n} z_{t,r} - \sum_{t=1}^{n} z_{t,r_1}\right|^2 \left|\sum_{t=1}^{n} z_{t,r_2} - \sum_{t=1}^{n} z_{t,r}\right|^2\right) \le C\left(\sum_{t=1}^{n}\sum_{s=1}^{t-1} c_{t-s}(r_1, r)^2\right)\left(\sum_{t=1}^{n}\sum_{h=1}^{t-1} c_{t-h}(r, r_2)^2\right),
$$

*where $C$ is a constant that does not depend on $r_1$, $r$ or $r_2$; $c_{t-s}(r_1,r) = c_{t-s,r} - c_{t-s,r_1}$; and $c_{t-h}(r, r_2)$ is defined analogously. This result also holds when $z_{t,r}$ is replaced by $z_{t,r}^*$ (c.f. B.2) and $c_t$ by $b_t$.*

**Proof:** Let $z_t(s,r) = z_{t,r} - z_{t,s}$. Then,

$$E\left(\left|\sum_{t=1}^n z_{t,r} - \sum_{t=1}^n z_{t,r_1}\right|^2 \left|\sum_{t=1}^n z_{t,r_2} - \sum_{t=1}^n z_{t,r}\right|^2\right)$$

$$= E\sum_{t=1}^n z_t(r_1,r)^2 z_t(r,r_2)^2 + E\sum_{t=1}^n \sum_{k\neq t} z_t(r_1,r)^2 z_k(r,r_2)^2 + 2E\sum_{t=1}^n \sum_{l\neq t} z_t(r_1,r) z_l(r_1,r) z_t(r,r_2) z_l(r,r_2).$$

$$= (T.1)+(T.2)+(T.3)$$

We analyze the three terms separately.

$$(T.2) = E\sum_{t=1}^n \varepsilon_t^2 \left(\sum_{s=1}^{t-1} c_{t-s}(r_1,r)\varepsilon_s\right)^2 \sum_{k<t} \varepsilon_k^2 \left(\sum_{h=1}^{k-1} c_{k-h}(r,r_2)\varepsilon_h\right)^2$$

$$+ E\sum_{k=1}^n \varepsilon_k^2 \left(\sum_{h=1}^{k-1} c_{k-h}(r,r_2)\varepsilon_h\right)^2 \sum_{t<k} \varepsilon_t^2 \left(\sum_{s=1}^{t-1} c_{t-s}(r_1,r)\varepsilon_s\right)^2.$$

Due to symmetry, it suffices to consider the first term, which equals to

$$\sum_{t=1}^n \sum_{k<t} E\left(\sum_{s=1}^{t-1} c_{t-s}(r_1,r)^2\varepsilon_s^2\right)\varepsilon_k^2\left(\sum_{h=1}^{k-1} c_{k-h}(r,r_2)^2\varepsilon_h^2\right) \qquad (I)$$

$$+ \sum_{t=1}^n \sum_{k<t} \sum_{s=1}^{k-1} \sum_{h=1,h\neq s}^{k-1} c_{t-s}(r_1,r)c_{t-h}(r_1,r)c_{k-s}(r,r_2)c_{k-h}(r,r_2) \quad (II).$$

Because $E(\varepsilon_t^4) < \infty$, term (I) is bounded by

$$C_1 \sum_{t=1}^n \sum_{k<t}\left(\sum_{s=1}^{t-1} c_{t-s}(r_1,r)^2\right)\left(\sum_{h=1}^{k-1} c_{k-h}(r,r_2)^2\right) \leq C_1 \sum_{t=1}^n\left(\sum_{s=1}^{t-1} c_{t-s}(r_1,r)^2\right)\sum_{t=1}^n\left(\sum_{h=1}^{t-1} c_{t-h}(r,r_2)^2\right) \quad (B.7)$$

for some $0 < C_1 < \infty$. Applying the Cauchy-Schwarz inequality to the elements of (II), we have

$\sum_{s=1}^{k-1} |c_{t-s}(r_1,r)c_{k-s}(r,r_2)| \leq \left(\sum_{s=1}^{k-1} c_{t-s}(r_1,r)^2\right)^{1/2}\left(\sum_{s=1}^{k-1} c_{k-s}(r,r_2)^2\right)^{1/2}$ and

$$\sum_{h=1,h\neq s}^{k-1} |c_{t-h}(r_1,r)c_{k-h}(r,r_2)| \leq \left(\sum_{h=1}^{k-1} c_{t-h}(r_1,r)^2\right)^{1/2}\left(\sum_{h=1}^{k-1} c_{k-h}(r,r_2)^2\right)^{1/2}.$$

Combining these two results:

$$|(II)| \leq \sum_{t=1}^n \sum_{k<t}\left(\sum_{s=1}^{k-1} c_{t-s}(r_1,r)^2\right)\left(\sum_{h=1}^{k-1} c_{k-h}(r,r_2)^2\right) \qquad (B.8)$$

$$\leq \sum_{t=1}^n\left(\sum_{s=1}^{t-1} c_{t-s}(r_1,r)^2\right)\sum_{t=1}^n\left(\sum_{h=1}^{t-1} c_{t-h}(r,r_2)^2\right),$$

which is proportional to (B.7). Hence,

$$|(I) + (II)| \leq C_2 \sum_{t=1}^{n} \left( \sum_{s=1}^{t-1} c_{t-s}(r_1, r)^2 \right) \sum_{t=1}^{n} \left( \sum_{h=1}^{t-1} c_{t-h}(r, r_2)^2 \right) \text{ with } C_2 = C_1 + 1. \qquad (B.9)$$

Apply the Cauchy-Schwarz inequality to (T.3):

$$|E(T.3)| \leq E\left( \sum_{t=1}^{n} z_t(r_1, r)^2 \right) \left( \sum_{t=1}^{n} z_t(r, r_2)^2 \right)$$

$$= E\left( \sum_{t=1}^{n} z_t(r_1, r)^2 \sum_{k=1, k \neq t}^{n} z_k(r, r_2)^2 \right) + E \sum_{t=1}^{n} z_t(r_1, r)^2 z_t(r, r_2)^2, \qquad (B.10)$$

where the first term is the same as (T.2) and the second term equals (T.1).

Finally, we turn to (T.1). It equals to

$$\mu_4 \sum_{t=1}^{n} \left( E \sum_{s=1}^{t-1} \sum_{k=1}^{t-1} \sum_{h=1}^{t-1} \sum_{l=1}^{t-1} c_{t-s}(r_1, r) c_{t-k}(r_1, r) c_{t-h}(r, r_2) c_{t-l}(r, r_2) \varepsilon_s \varepsilon_k \varepsilon_h \varepsilon_l \right)$$

$$= \mu_4^2 \sum_{t=1}^{n} \left( \sum_{s=1}^{t-1} c_{t-s}(r_1, r)^2 c_{t-s}(r, r_2)^2 \right) + \mu_4 \sum_{t=1}^{n} \left( \sum_{s=1}^{t-1} \sum_{h=1, h \neq s}^{t-1} c_{t-s}(r_1, r)^2 c_{t-h}(r, r_2)^2 \right)$$

$$+ 2\mu_4 \sum_{t=1}^{n} \left( \sum_{s=1}^{t-1} c_{t-s}(r_1, r) c_{t-s}(r, r_2) \sum_{k=1, k \neq s}^{t-1} c_{t-k}(r_1, r) c_{t-k}(r, r_2) \right)$$

$$\leq C_3 \sum_{t=1}^{n} \left( \sum_{s=1}^{t-1} \sum_{h=1}^{t-1} c_{t-s}(r_1, r)^2 c_{t-h}(r, r_2)^2 \right) \qquad (III)$$

$$+ 2\mu_4 \sum_{t=1}^{n} \left( \sum_{s=1}^{t-1} c_{t-s}(r_1, r) c_{t-s}(r, r_2) \sum_{k=1, k \neq s}^{t-1} c_{t-k}(r_1, r) c_{t-k}(r, r_2) \right) \qquad (IV)$$

for some constant $C_3$. Applying the same arguments that lead to (B.8), we have $|(IV)| \leq 2\mu_4 \sum_{t=1}^{n} \sum_{s=1}^{t-1} c_{t-s}(r_1, r)^2 \sum_{k=1}^{t-1} c_{t-k}(r, r_2)^2$. Hence,

$$|(III) + (IV)| \leq C_4 \left( \sum_{t=1}^{n} \sum_{s=1}^{t-1} c_{t-s}(r_1, r)^2 \right) \left( \sum_{t=1}^{n} \sum_{h=1}^{t-1} c_{t-h}(r, r_2)^2 \right), \qquad (B.11)$$

where $C_4 = C_3 + 2\mu_4$. Combining (B.9), (B.10), and (B.11) leads to the desired result. ∎