Queen's Economics Department Working Paper No. 1101

Simple (but effective) tests of long memory versus structural breaks

Katsumi Shimotsu
Department of Economics, Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

12-2006

Simple (but effective) tests of long memory versus structural breaks

Katsumi Shimotsu*
Department of Economics
Queen's University

December 3, 2006

Abstract

This paper proposes two simple tests that are based on certain time domain properties of $I(d)$ processes. First, if a time series follows an $I(d)$ process, then each subsample of the time series also follows an $I(d)$ process with the same value of d. Second, if a time series follows an $I(d)$ process, then its dth differenced series follows an $I(0)$ process. Simple as they may sound, these properties provide useful tools to distinguish the true and spurious $I(d)$ processes.

In the first test, we split the sample into b subsamples, estimate d for each subsample, and compare them with the estimate of d from the full sample. In the second test, we estimate d, use the estimate to take the d th difference of the sample, and apply the KPSS test and Phillips-Perron test to the differenced data and its partial sum. Both tests are applicable to both stationary and nonstationary $I(d)$ processes.

Simulations show that the proposed tests have good power against the spurious long memory models considered in the literature. The tests are applied to the daily realized volatility of the S&P 500 index.

JEL Classification Number: C12, C13, C14, C22
Keywords: long memory; fractional integration; structural breaks; realized volatility.

1 Introduction

Long memory (fractionally integrated, $I(d)$) processes have been used extensively in modeling the strong persistence observed in volatility of asset prices. Important contributions include Ding, Granger and Engle (1993), Baillie, Bollerslev and Mikkelsen (1996), Andersen and Bollerslev (1997), Breidt et al. (1998), Bollerslev and Wright (2000), Andersen et al. (2003), Deo, Hurvich

*e-mail: shimotsu@econ.queensu.ca. Phone: +1-613-536-6546. Fax: +1-613-533-6668. The author thanks Torben Andersen, Rohit Deo, Christian Gourieroux, Alex Maynard, Nour Meddahi, Peter Phillips, and the participants at the CIRANO-CIREQ Conference on Time Series, 2006 International Symposium on Econometric Theory and Applications (SETA), and the seminar at Yokohama National University for helpful comments. The author thanks Qian Liu for research assistance. Financial support from SSHRC is gratefully acknowledged.

and Lu (2005), Hurvich, Moulines, and Soulier (2005), and Christensen and Nielsen (2006), just to name a few.

At the same time, many studies show that the time series with structural breaks can induce a strong persistence in the autocorrlation function and hence generate "spurious" long memory (Diebold and Inoue, 2001, Gourieroux and Jasiak, 2001, Granger and Hyung, 2004, Perron and Qu, 2006). By providing examples in which long memory can be easily confused with structural breaks, this literature concludes that it is very difficult to distinguish between true and spurious long memory processes. We are left with the impression that long memory and structural breaks are almost observationally equivalent and that long memory may fall into an "empty box" category. Perron and Qu (2006) analytically show how a stationary short memory process with level shifts can generate spurious long memory.

On the other hand, there is evidence that long memory processes successfully model some economics and financial data. Andersen et al. (2003) and Bhardwaj and Swanson (2006) find that the long memory models provide significantly better out-of-sample predictions than AR, MA, ARMA, GARCH and related models. Granger and Hyung (2004) also report that long memory models have better out-of-sample forecasting performance than the occasional break models, although the evidence is statistically insignificant. Mayoral (2006) develops a time-domain test of $I(d)$ versus $I(0)$ plus trends and/or breaks, and finds that the null of $I(d)$ is not rejected with the U.S. inflation data. Hsu (2005) reports that the U.S. inflation rates have strong dependence even after the breaks in the mean are allowed. Choi and Zivot (2005) estimate the d of an exchange rate forward discount series after adjusting for breaks in their mean. Choi and Zivot find that allowing for structural breaks reduces the persistence of the forward discount but there is still evidence of long memory.

This paper proposes two simple tests that are based on certain time domain properties of $I(d)$ processes. First, if a time series follows an $I(d)$ process, then each subsample of the time series also follows an $I(d)$ process with the same value of d. Second, if a time series follows an $I(d)$ process, then its dth differenced series follows an $I(0)$ process. Simple as they may sound, these properties provide useful tools to distinguish the true and spurious $I(d)$ processes. Both tests are applicable to both stationary and nonstationary $I(d)$ processes.

In the first test, we split the sample into b subsamples, estimate d for each subsample, and compare them with the estimate of d from the full sample. We use the local Whittle (Gaussian semiparametric) estimator (Robinson, 1995) and the exact local Whittle (ELW) estimator (Shimotsu and Phillips, 2005; Shimotsu, 2006) to estimate d, mainly because of their computational simplicity and the invariance of their limiting distribution with respect to d. For spurious $I(d)$ models, it turns out that the averaged estimates from the subsamples tend to differ from the full sample estimate, and their difference increases as the degree of sample splitting increases. Even when the spurious $I(d)$ models can generate certain features of the true $I(d)$ models, such as autocorrelation's rate of decay, the "implied value" of d from these models depends on their parameter values (e.g. the switching probability) and the number of the observations. Once we split the sample, the combination of the number of observations and parameter values changes, and it generates a different "implied value" of d. Because this test involves only splitting the sample and estimating d for each subsample, an empirical researcher can routinely use it as a reality check before proceeding to a further analysis of the data at hand. We also develop a formal test statistic that tests the null of parameter constancy.

In the second test, we estimate d on the whole sample, use the estimate to take the dth difference of the sample, and apply the KPSS test and Phillips-Perron test to the differenced data and its partial sum. Despite its simplicity, it provides a very powerful tool to distinguish between the true and spurious $I(d)$ processes. The spurious long memory processes are essentially $I(0)$ or $I(1)$ in their nature, and taking their dth difference magnifies their non-$I(d)$ properties. The limiting distribution of these test statistics depends on d, and its simulated critical values are provided.

Other tests to distinguish between the long memory and structural breaks have been developed recently. Ohanissian et al. (2004) propose a test of true versus spurious long memory by comparing the estimates of d obtained from temporally aggregated series with different degrees of aggregation. It exploits the invariance of the memory parameter under temporal aggregation proven by Chambers (1998). Other contributions include Berkes et al. (2006), Dolado et al. (2005), Giraitis et al. (2006), Mayoral (2006), and Perron and Qu (2006). Banerjee and Urga (2005) provide a comprehensive survey of the literature on both long memory and structural breaks.

The rest of the paper is organized as follows. Section 2 reviews stationary long memory processes and local Whittle estimation. Section 3 introduces our sample splitting-based tests. Section 4 extends the results in Sections 2 and 3 to nonstationary $I(d)$ processes. Section 5 discusses our second test based on the dth differencing. Section 6 reports the simulation results and applies the proposed tests to the daily realized volatility of the S&P 500 index. Section 7 concludes. Proofs and technical results are collected in Appendix in Section 8.

2 Long-memory process and local Whittle estimation

In this and the following section, we consider covariance stationary long memory processes. We assume the spectral density $f(\lambda)$ of the process X_t satisfies

$$f(\lambda) \sim G\lambda^{-2d}, \quad \text{as } \lambda \to 0+,$$

where $d \in (-1/2, 1/2)$ and $G \in (0, \infty)$. The most widely used long memory process is a fractionally integrated process, given by

$$(1-L)^d X_t = u_t, \tag{1}$$

where L is the lag operator and u_t is a covariance stationary process whose spectral density is bounded and bounded away from zero at the zero frequency $\lambda = 0$.

Define the discrete Fourier transform (dft) and the periodogram of X_t evaluated at the fundamental frequencies as

$$\begin{aligned} w_x(\lambda_j) &= (2\pi n)^{-1/2} \sum_{t=1}^{n} X_t e^{it\lambda_j}, \quad \lambda_j = \frac{2\pi j}{n}, j = 1, \ldots, n, \\ I_x(\lambda_j) &= |w_x(\lambda_j)|^2. \end{aligned} \tag{2}$$

Local Whittle (Gaussian semiparametric) estimation was developed by Künsch (1987) and Robinson (1995). Specifically, it starts with the following Gaussian objective function, defined in terms of the parameters d and G

$$Q_m(G,d) = \frac{1}{m} \sum_{j=1}^{m} \left[\log\left(G\lambda_j^{-2d}\right) + \frac{\lambda_j^{2d}}{G} I_x(\lambda_j) \right], \tag{3}$$

where m is some integer less than n. The local Whittle procedure estimates G and d by minimizing $Q_m(G,d)$, so that

$$(\widehat{G}, \widehat{d}) = \underset{G \in (0,\infty), d \in [\Delta_1, \Delta_2]}{\arg\min} Q_m(G,d),$$

where Δ_1 and Δ_2 are numbers such that $-1/2 < \Delta_1 < \Delta_2 < \infty$. It will be convenient in what follows to distinguish the true values of the parameters by the notation $G_0 = f_u(0)$ and d_0. Concentrating (3) with respect to G gives

$$\widehat{d} = \underset{d \in [\Delta_1, \Delta_2]}{\arg\min} R(d),$$

where

$$R(d) = \log \widehat{G}(d) - 2d \frac{1}{m} \sum_{1}^{m} \log \lambda_j, \quad \widehat{G}(d) = \frac{1}{m} \sum_{1}^{m} \lambda_j^{2d} I_x(\lambda_j).$$

Robinson (1995) shows $\sqrt{m}(\hat{d} - d_0) \to_d N(0, 1/4)$ as $n \to \infty$ under the conditions stated in Section 3.

3 Sample splitting-based diagnosis

This section provides two diagnoses based on sample splitting; one is based on visual examination, and the other is a formal statistical test. Let b be an integer and split the sample into b blocks, so that each block has n/b observations. We assume n/b is integer. Define $\hat{d}^{(a)}$, $a = 1, \ldots, b$, be the local Whittle estimator of d computed from the ath block of the observations, $\{X_t : t = (a-1)n/b + 1, \ldots, an/b\}$.

The choice of m, the number of the periodogram ordinates used in the objective function, plays an important role in the local Whittle and other semiparametric estimators, because it determines the width of the frequency band used in estimating d. We set the number of the periodogram ordinates used in the subsample estimation as m/b. With this choice, the subsample estimation uses the same amount of frequency domain information as the estimation by the total sample. Specifically, the subsample uses the periodogram ordinates $2\pi/(n/b), \ldots, 2\pi(m/b)/(n/b)$. This mitigates the effect of short-run dynamics on the test statistic, because the estimates using the subsamples have the same amount of bias from short-run dynamics of X_t as the estimate using the total sample.

For the ath subsample, define

$$\widehat{d}^{(a)} = \underset{d \in [\Delta_1, \Delta_2]}{\arg\min}\, R^{(a)}(d),$$

where the objective function is constructed from the ath block of the observations:

$$\begin{aligned} R^{(a)}(d) &= \log \widehat{G}^{(a)}(d) - 2d \frac{b}{m} \sum_{j=1}^{m/b} \log \tilde{\lambda}_j, \quad \widehat{G}^{(a)}(d) = \frac{b}{m} \sum_{j=1}^{m/b} \tilde{\lambda}_j^{2d} I_x^{(a)}\left(\tilde{\lambda}_j\right), \\ I_x^{(a)}\left(\tilde{\lambda}_j\right) &= (2\pi n)^{-1} \left| \sum_{t=(a-1)n/b+1}^{an/b} X_t e^{it\tilde{\lambda}_j} \right|^2, \quad \tilde{\lambda}_j = \frac{2\pi j}{n/b}, \quad j = 1, \ldots, n/b. \end{aligned}$$

3.1 Visual examination based on sample-splitting

First, we introduce a simple visual examination of whether a given (possibly spurious long-memory) process can reproduce one feature of the true long memory process. Consider estimating d by taking the average of $\hat{d}^{(1)}, \ldots, \hat{d}^{(b)}$. Then, if X_t is an $I(d)$ process, then the average of $\hat{d}^{(1)}, \ldots, \hat{d}^{(b)}$ should be close to $\hat{d}$. As it turns out, it is not necessarily the case with spurious long-memory processes.

We introduce assumptions on X_t, $f(\lambda)$, and m. They are the same as Assumptions A1′-A4′ of Robinson (1995).

Assumption A1 For some $\beta \in (0, 2]$,

$$f(\lambda) \sim G_0 \lambda^{-2d_0} \left(1 + O(\lambda^{\beta})\right) \quad \text{as } \lambda \to 0+,$$

where $G_0 \in (0, \infty)$ and $d_0 \in [\Delta_1, \Delta_2] \cap (-1/2, 1/2)$.

Assumption A2

$$X_t - EX_0 = C(L)\,\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} c_j^2 < \infty,$$

where $E(\varepsilon_t | F_{t-1}) = 0$, $E(\varepsilon_t^2 | F_{t-1}) = 1$ a.s., $t = 0, \pm 1, \ldots$, in which F_t is the σ-field generated by ε_s, $s \leq t$, and there exists a random variable ε such that $E\varepsilon^2 < \infty$ and for all $\eta > 0$ and some $K > 0$, $\Pr(|\varepsilon_t| > \eta\ ) \leq K \Pr(|\varepsilon| > \eta)$. Furthermore,

$$E(\varepsilon_t^3 | F_{t-1}) = \mu_3, \quad E(\varepsilon_t^4 | F_{t-1}) = \mu_4, \quad \text{a.s.}, \quad t = 0, \pm 1, \ldots, \quad \mu_3, \mu_4 < \infty.$$

Assumption A3 In a neighborhood $(0, \delta)$ of the origin, $C(e^{i\lambda})$ is differentiable and

$$\frac{d}{d\lambda} C(e^{i\lambda}) = O\left(\frac{|C(e^{i\lambda})|}{\lambda}\right) \quad \text{as } \lambda \to 0+.$$

Assumption A4 As $n \to \infty$,

$$\frac{1}{m} + \frac{m^{1+2\beta} \log^2 m}{n^{2\beta}} \to 0.$$

The following Lemma is a simple consequence of the results in Robinson (1995):

Lemma 1 *Suppose Assumptions A1-A4 hold. Then* $\sqrt{m}(\bar{d} - d_0) \to_d N(0, 1/4)$ *as* $n \to \infty$.

The limiting variance of $\bar{d}$ is the same as that of $\widehat{d}$, the estimator of d constructed from the full (non-split) sample, because $\hat{d}^{(a)}$ for different a are asymptotically independent due to the asymptotic uncorrelatedness of $I_x^{(a)}(\tilde{\lambda}_j)$. As shown in the simulation below, in finite samples $\bar{d}$ has a larger variance than $\widehat{d}$. We cannot use $\sqrt{m}(\hat{d} - \bar{d})$ to construct a test statistic, because it has a degenerate limiting distribution as shown in Section 3.2. However, even a simple visual examination provides useful information regarding the nature of the data.

We compare the behavior of $\bar{d}$ for true and spurious long memory processes through simulations. We consider three processes that have been shown to exhibit spurious long memory properties by Diebold and Inoue (2001) and Granger and Hyung (2004). The first is the mean-plus-noise model, which is given by

$$X_t = \mu_t + \varepsilon_t, \quad \mu_t = \mu_{t-1} + v_t, \quad v_t = \begin{cases} 0 & \text{with probability } 1 - p, \\ w_t & \text{with probability } p, \end{cases}$$

where $\varepsilon_t \sim iidN(0, \sigma_\varepsilon^2)$ and $w_t \sim iidN(0,1)$. The second model is the stochastic permanent break (STOPBREAK) model (Engle and Smith, 1999):

$$X_t = \mu_t + \varepsilon_t, \quad \mu_t = \mu_{t-1} + \frac{\varepsilon_{t-1}^2}{\gamma + \varepsilon_{t-1}^2}\varepsilon_{t-1}, \quad \varepsilon_t \sim iidN(0,1).$$

The third model is the Markov-switching model:

$$X_t = \begin{cases} N(0,\sigma^2) & \text{if } s_t = 0, \\ N(1,\sigma^2) & \text{if } s_t = 1, \end{cases} \qquad M = \begin{pmatrix} p_{00} & 1-p_{00} \\ 1-p_{11} & p_{11} \end{pmatrix},$$

where s_t is a first-order Markov process taking the value 0 or 1 and the with transition probability matrix is given by M.

Figures 1 and 2 plot kernel estimates of the density of $\bar{d}$ for an $I(0.4)$ process given by (1) with $u_t \sim iidN(0,1)$ and the three spurious long memory processes. We choose $d = 0.4$, since the typical estimates of d from realized volatility data lie around 0.4. The sample size is set to $n = 5,000$. This approximates the number of daily observations in 20 years. m and b are chosen as $m = 200$ and $b = \{1,2,4,8\}$. $b = 1$ corresponds to no sample splitting. The parameter values of the spurious long memory processes are chosen so that the mean of $\bar{d}$ when $b = 1$ is approximately equal to 0.4. The upper panel of Figure 1 shows the density of $\bar{d}$ with $I(0.4)$ process. The mean of $\bar{d}$ remain unchanged even when b changes, but the distribution of $\bar{d}$ becomes more dispersed as b increases. The second panel of Figure 1 shows the density of $\bar{d}$ when X_t is generated by the mean-plus-noise model. As b increases, the distribution of $\bar{d}$ shifts toward left and appears to more concentrate around its peak. Figure 2 shows the density of $\bar{d}$ for STOPBREAK and Markov-switching models. With the STOPBREAK model, the distribution of $\bar{d}$ changes similarly to the mean-plus-noise model. As illustrated by Diebold and Inoue (2001), both models are $I(1)$, and sample splitting appears to reduce their $I(1)$ characteristics. With the Markov-switching models, the density of $\bar{d}$ does not appear to shift significantly but is more concentrated around its peak for larger values of b.

In sum, the behavior of $\bar{d}$ from spurious long memory models differs substantially from the behavior of $\bar{d}$ from a true long-memory process. In the following, we develop a formal test statistic to test the null of the constancy of d across subsamples.

3.2 Test statistic for the parameter constancy

We construct a test statistic for formally testing true $I(d)$ versus spurious $I(d)$. It tests the hypothesis $H_0 : d_0 = d_{0,1} = \cdots = d_{0,b}$, where $d_{0,a}$ is the true value of d from the ath subsample.

Define a $b+1$ vector $\hat{d}_b$ and a $b \times (b+1)$ matrix A as

$$\hat{d}_b = \begin{pmatrix} \hat{d} - d_0 \\ \hat{d}^{(1)} - d_0 \\ \vdots \\ \hat{d}^{(b)} - d_0 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & -1 \end{pmatrix}.$$

In the proof of Lemma 2, we show that, under H_0,

$$\sqrt{m}\hat{d}_b = Z_n + bias(m), \quad Z_n \to_d N\left(0, \frac{1}{4}\Omega\right), \quad \Omega = \begin{pmatrix} 1 & \iota_b' \\ \iota_b & bI_b \end{pmatrix},$$

where I_b is a $b \times b$ identity matrix and ι_b is a $b \times 1$ vector of ones. We ignore the bias term for the moment. Ω is singular with rank b, and simple algebra shows $A\Omega A' = bI_b - \iota_b\iota_b'$, which has rank $b-1$. Define the Wald statistic for testing H_0 as

$$W = 4mA\hat{d}_b(A\Omega A')^{+}(A\hat{d}_b)',$$

where $(A\Omega A')^{+}$ denotes a generalized inverse of $A\Omega A'$. Then W has a chi-squared limiting distribution with $b-1$ degrees of freedom.

Hurvich and Chen (2000, p. 164) report that the finite sample variance of the local Whittle estimator tends to be larger than $1/(4m)$ and the Wald test based on the asymptotic variance tends to overreject the null. Hurvich and Chen find that replacing m in the variance estimate by a number c_m improves approximation, where c_m is defined as[1]

$$c_m = \sum_{j=1}^{m} \nu_j^2, \quad \nu_j = \log\lambda_j - \frac{1}{m}\sum_{j=1}^{m}\log\lambda_j = \log j - \frac{1}{m}\sum_{j=1}^{m}\log j.$$

Since $c_m/m \to 1$ as $m \to \infty$, this modification does not alter the asymptotic distribution of the test statistic. Following Hurvich and Chen, we introduce the adjusted Wald statistic[2]

$$W_c = 4m(c_{m/b}/(m/b))A\hat{d}_b(A\Omega A')^{+}(A\hat{d}_b)'.$$

One feature of our test is that each subsample-based estimator uses the same amount of frequencies. This implies that the bias of the all elements of $\hat{d}_b$ are the same and allows us to choose larger values of m than in estimating d.

We introduce two additional assumptions. Assumption A1$'$ strengthens Assumption A1, and

[1] Hurvich and Chen use $2\sin(\lambda_j/2)$ instead of λ_j, but their difference is small for $\lambda_j \sim 0$. For more details, see Hurvich and Chen (2000).

[2] We also experimented with more detailed modifications using both c_m and $c_{m/b}$, but the resulting test statistics did not improve over W_c.

Assumption A4$'$ relaxes Assumption A4 slightly. The maximum allowable expansion rate of m when $\beta = 2$ is $n^{6/7-\delta}$ for any $\delta > 0$, contrary to the rate $n^{4/5-\delta}$ allowed in Assumption A4.

Assumption A1$'$ For some $\beta \in (0, 2]$,

$$f(\lambda) = G_0 \lambda^{-2d_0}(1 + E_\beta \lambda^\beta + O(\lambda^{\beta+1})) \quad \text{as } \lambda \to 0+,$$

where $G_0, E_\beta \in (0, \infty)$ and $d_0 \in [\Delta_1, \Delta_2] \cap (-1/2, 1/2)$.

Assumption A4$'$ As $n \to \infty$,

$$\frac{1}{m} + \frac{m^{1+3\beta} \log^{30} m}{n^{3\beta}} \to 0.$$

The following Lemma establishes the limiting distribution of W.

Lemma 2 *Suppose Assumptions A1$'$, A2, A3, and A4$'$ hold. Then $W, W_c \to_d \chi^2(b-1)$ as $n \to \infty$.*

A result corresponding to Lemma 2 but using the parametric Whittle estimator (Fox and Taqqu, 1986; Giraitis and Surgailis, 1990) has been proven by Beran and Terrin (1994). Beran and Terrin (1994) also suggest the possibility of extending their results to the log-periodogram regression estimator, albeit without a formal justification.

Ohanissian et al. (2004) propose a test of true against spurious long memory by comparing the estimates of d obtained from temporally aggregated series with different degrees of aggregation. While they report very good power of their test, their results are not directly comparable to ours. Their test compares the memory parameter estimates that use $(n/a)^{0.5}$ periodogram ordinates, where a is the degree of temporal aggregation. Therefore, they compare the memory parameter estimates that use different amount of frequency domain information depending on the degree of temporal aggregation. Second, they are interested in analyzing 5-minute returns with around 600,000 observations.

4 Extension to nonstationary $I(d)$ processes

In this section, we extend Lemmas 1 and 2 to $I(d)$ processes with initialization at $t = 0$ (the so-called Type II processes). Type II processes are attractive in empirical applications, because of their ability to accommodate both (asymptotically) stationary and nonstationary $I(d)$ processes in a single dgp.

Assumption B0

$$X_t - \mu_0 = (1-L)^{-d_0} u_t \mathbf{1}\{t \geq 1\}, \tag{4}$$

where μ_0 is a non-random unknown finite number, $\mathbf{1}\{\cdot\}$ is the indicator function, u_t is stationary with zero mean and spectral density $f_u(\lambda)$, and $d_0 \in (-1/2, 2)$.

Assumption B1 $f_u(\lambda) \sim G_0(1 + O(\lambda^\beta))$ as $\lambda \to 0$ for some $\beta \in (0, 2]$ and $G_0 \in (0, \infty)$.

Assumption B2 u_t satisfies Assumption A2.

Assumption B3 Assumption A3 holds.

Assumption B4 Assumption A4 holds, and $\sum_{j \geq k} E(u_t u_{t+j}) = O((\log(k+1))^{-4})$, $\sum_{j \geq k} c_j = O((\log(k+1))^{-4})$ uniformly in $k = 0, 1, \ldots$.

Assumption B5 (a) For any $\gamma > 0$,

$$\frac{1}{m} + \frac{m^{(1+2\beta)} \log^2 m}{n^{2\beta}} + \frac{\log n}{m^\gamma} \to 0 \quad \text{as } n \to \infty,$$

and (b) the spectral density of u_t is bounded.

With X_t defined by (4), we can use both the local Whittle estimator and the exact local Whittle estimator of Shimotsu and Phillips (2005), albeit the local Whittle estimator requires $d_0 \in (-1/2, 3/4)$. Phillips and Shimotsu (2005, 2006) show that the local Whittle estimator is consistent and asymptotically $N(0, 1/4)$ for $d_0 \in (-1/2, 3/4)$. As shown by Shimotsu and Phillips (2005), the exact local Whittle estimator is consistent and asymptotically $N(0, 1/4)$ for $d_0 \in (\Delta_1, \Delta_2)$ if $\Delta_2 - \Delta_1 \leq 9/2$ and $\mu_0 = 0$. Shimotsu (2006) extends the exact local Whittle estimator to develop the two-step feasible exact local Whittle (FELW) estimator that accommodates unknown μ_0 and a polynomial time trend in X_t.

Assumptions B1-B3 are analogous to assumptions A1-A3 except that they are imposed on u_t rather than X_t. The summability condition in Assumption B4 is fairly mild and allows for poles and discontinuity of $f_u(\lambda)$ at $\lambda \neq 0$; see Shimotsu and Phillips (2006, p. 217). The rate condition in Assumption B5(a) is slightly stronger than Assumption A5. Assumption B5(b) excludes the poles in $f_u(\lambda)$ outside the origin. The FELW estimator uses the objective function

$$R_F(d) = \log \widehat{G}_F(d) - 2d \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j, \quad \widehat{G}_F(d) = \frac{1}{m} \sum_{j=1}^{m} I_{\Delta^d(x - \widetilde{\mu}(d))}(\lambda_j),$$

where $I_{\Delta^d(x-\widetilde{\mu}(d))}(\lambda_j)$ denotes the periodograms of $(1-L)^d(X_t - \widetilde{\mu}(d))$. $\widetilde{\mu}(d)$ is an estimate of

μ_0 defined by

$$\widetilde{\mu}\left(d\right)=w(d)\overline{X}+(1-w(d))X_1,$$

where $w(d)$ is a twice continuously differentiable weight function such that $w(d) = 1$ for $d \leq \frac{1}{2}$ and $w(d) = 0$ for $d \geq \frac{3}{4}$. This form of $\widetilde{\mu}\left(d\right)$ is chosen so that the effect of estimation of μ_0 on estimation of d becomes negligible. See Shimotsu (2006) for details. The FELW estimator is defined as

$$\widehat{d}_F=\tilde{d}-R_F(\tilde{d})'/R_F(\tilde{d})'', \tag{5}$$

where $\tilde{d}$ is a first-stage $m^{1/2}$-consistent estimator of d, for example the tapered estimators of Velasco (1999) Hurvich and Chen (2003). The tapered estimators are $m^{1/2}$-consistent but not as efficient as the FELW estimator. A polynomial trend in X_t can be handled by detrending X_t and applying the FELW estimator to the detrended data.

The following assumption corresponds to Assumptions A1$'$.

Assumption B1$'$ $f_u(\lambda)$ satisfies Assumption A1$'$ with $d_0 = 0$.

The following lemmas show we can apply the asymptotic results of Lemmas 1 and 2 to X_t generated by (4). We may use both the local Whittle and FELW estimator. Part (b) of Lemma 4 is proven only under the "standard" rate condition $m = o(n^{2\beta/(2\beta+1)})$, which is stronger than Assumption A4$'$. This is because of the difficulty in deriving the higher-order asymptotics for the FELW estimator, although we conjecture that Lemma 4 (b) holds under Assumption A4$'$.

Lemma 3 *(a) Suppose Assumptions B0-B4 hold and $d_0 \in (-1/2, 3/4)$. Then $\sqrt{m}(\bar{d} - d_0) \to_d N(0, 1/4)$ as $n \to \infty$. (b) Suppose Assumptions B0-B3 and B5 hold. Then $\sqrt{m}(\bar{d}_F - d_0) \to_d N(0, 1/4)$ as $n \to \infty$, where $\bar{d}_F$ is defined analogously to $\bar{d}$ using the FELW estimator.*

Lemma 4 *(a) Suppose Assumptions B0, B1$'$, B2, B3, B4 and A4$'$ hold. Then $W, W_c \to_d \chi^2(b-1)$ as $n \to \infty$. (b) Suppose Assumptions B0-B3 and B5 hold and we construct W and W_c using the FELW estimator. Then $W, W_c \to_d \chi^2(b-1)$ as $n \to \infty$.*

5 Test using dth differencing

The second test utilizes another time-domain property of an $I(d)$ process; if an $I(d)$ process is differenced d times, then the resulting time series is an $I(0)$ process. While this may seem trivial, some spurious long-memory processes do not mimic this property.

We propose a test statistic that uses the Z_t unit root test (Phillips and Perron, 1988) and KPSS test (Kwiatkowski et al., 1992). The idea is simple; we first demean the data, then apply the Z_t and KPSS tests to its $\hat{d}$th difference, where $\hat{d}$ is a consistent estimate of d. The demeaning needs to be done carefully, however, if we want to allow for nonstationary processes.

We assume X_t follows a truncated $I(d)$ process with initialization at $t = 0$, as in Assumption B0:

$$X_t - \mu_0 = (1 - L)^{-d_0} u_t \mathbf{1}\{t \geq 1\},$$

where μ is a non-random unknown finite number. We use this specification, because it accommodates both (asymptotically) stationary and nonstationary X_t in a single model, and the $\hat{d}$th difference of $X_t - \mu_0$ is a truncated $I(d - \hat{d})$ process for any pair of $(d, \hat{d})$. Here, μ is the mean of the process X_t in the sense $EX_t = \mu_0$, but when $d > 1/2$ it may be better interpreted as the initial condition of X_t.

We need to subtract an estimate of μ_0 from X_t before taking its $\hat{d}$th difference. When d_0 is known to be no larger than 1, we can use the sample average $\overline{X} = n^{-1} \sum_{t=1}^{n} X_t$ as an estimate of μ_0. When $d_0 \geq 1$, however, using the sample average induces a nonnegligible error. This is because $\overline{X} - \mu_0 = O_p(n^{d_0 - 1/2})$ and, as shown in the proof,

$$(1 - L)^{d_0}(X_t - \overline{X}) = u_t + O_p(n^{d_0 - 1/2} t^{-d_0}).$$

If $d_0 \geq 1$, the second term on the right has a nonnegligible effect on the sample statistics of $(1 - L)^{d_0}(X_t - \overline{X})$, in particular, the sample variance and autocovariances. Following Shimotsu (2006), we use X_1 as an estimate of μ_0 when d_0 takes a large value. Although X_1 is not a consistent estimate of μ_0, it turns out

$$(1 - L)^{d_0}(X_t - X_1) = u_t + O_p(t^{-d_0}),$$

and the error from estimating μ_0 becomes negligible.

As in Shimotsu (2006), we use a linear combination of $\overline{X}$ and X_1 to estimate μ_0 :

$$\hat{\mu}(d) = w(d)\overline{X} + (1 - w(d))X_1,$$

where $w(d)$ is a smooth (twice continuously differentiable) weight function such that $w(d) = 1$ for $d \leq 1/2$ and $w(d) = 0$ for $d \geq \frac{3}{4}$. An example of $w(d)$ for $d \in [1/2, 3/4]$ is $(1/2)[1 + \cos(4\pi d)]$. Then the difference between $(1 - L)^{d_0}(X_t - \hat{\mu}(d_0))$ and u_t becomes negligible for any value of d_0.

With this estimate of μ_0 in hand, define the dth differenced series as

$$\hat{u}_t = (1 - L)^{\hat{d}}(X_t - \hat{\mu}(\hat{d})) = \sum_{k=0}^{t-1} \frac{\Gamma(-\hat{d} + k)}{\Gamma(-\hat{d})k!}(X_{t-k} - \hat{\mu}(\hat{d})). \tag{6}$$

We then apply the Z_t and KPSS tests to $\hat{u}_t$. The Z_t test with an intercept term is applied to the partial sum process of $\hat{u}_t$. The KPSS test statistic is defined as

$$\hat{\eta}_\mu = n^{-2}\sum_{t=1}^{n} S_t^2/s^2(l), \quad S_t = \sum_{k=1}^{t} e_k, \quad s^2(l) = \frac{1}{n}\sum_{t=1}^{n} e_t^2 + \frac{2}{n}\sum_{s=1}^{l}\left(1-\frac{s}{l+1}\right)\sum_{t=s+1}^{n} e_t e_{t-s},$$

where e_t is the residuals from regressing $\hat{u}_t$ on an intercept, i.e., $e_t = \hat{u}_t - n^{-1}\sum_{t=1}^{n}\hat{u}_t$. $s^2(l)$ is a consistent estimate of the long-run variance of $(1-L)^{d_0}(X_t - \mu_0)$. Including an intercept stabilizes the finite sample behavior of the test statistic.

Assumption C1 X_t satisfies Assumption B0.

Assumption C2 $\hat{d} - d_0 = o_p((\log n)^{-1})$.

Assumption C3 For $0 \le r \le 1$,

$$n^{-d_0-1/2}\sum_{t=1}^{[nr]} X_t \Rightarrow \frac{\omega}{\Gamma(d_0+1)} W_{d_0+1}(r), \quad W_d(r) \equiv \int_0^r (r-s)^{d-1} dW(s), \tag{7}$$

$$n^{-1/2}\sum_{t=1}^{[nr]} u_t \Rightarrow \omega W(r), \tag{8}$$

where $W(r)$ is a standard Brownian motion, $\omega^2 = 2\pi f_u(0)$, and $\Rightarrow$ signifies weak convergence in the space $D[0,1]$.

Assumption C4 (a) $l(\hat{d} - d_0) \to_p 0$. (b) $\tilde{s}^2(l) \to_p \omega^2$, where $\tilde{s}^2(l)$ is the version of $s^2(l)$ constructed with u_t instead of $\hat{u}_t$.

Under Assumptions B0-B5, the local Whittle estimator satisfies Assumption C2 by Theorem 4.1 of Phillips and Shimotsu (2006), and the FELW estimator satisfies Assumption C2 by Theorem 5 of Shimotsu (2006). Marrinuchi and Robinson (2000) show the convergence (7) when $\sum_{j=1}^{\infty} j|c_j| < \infty$ and ε_t is iid with $E|\varepsilon_t|^q < \infty$, where $q = \max\{2, 2/(2d_0+1)\}$. In our context, $q = 2$ suffices in most cases, because $d_0 < 0$ is unlikely to occur. Hosoya (2005, Theorems 2 and 3) show that the convergence (7) holds when u_t is a white noise (not necessarily iid) with $E|u_t|^q < \infty$ and has a bounded fourth-order cumulant spectra density. The convergence (8) holds under weaker assumptions on u_t than those required for (7) in general. Assumption C4(a) limits the divergence rate of l when the convergence rate of $\hat{d}$ is slow. A similar condition between the expansion rate of the bandwidth and the convergence rate of an estimator is employed in other works on estimation of ω^2, e.g. Hansen (1992) and de Jong and Davidson (2000). The convergence $\tilde{s}^2(l) \to_p \omega^2$ has been shown under various conditions on the temporal dependence of u_t (for example, Hansen (1992), de Jong and Davidson (2000)).

If $\hat{u}_t$ were a level-stationary process without a linear time trend, then it follows from Phillips and Perron (1988) and Kwiatkowski et al. (1992) that $Z_t \to_d P(W(r))$ and $\hat{\eta}_\mu \to_d K(W(r)) = \int_0^1 (W(r) - rW(1))^2 dr$, where $P(W(r))$ denotes the standard Dickey-Fuller (DF) distribution when an intercept is included. The following lemma establishes the limiting distribution of Z_t and $\hat{\eta}_\mu$ when they are applied to $\hat{u}_t$.

Lemma 5 *Suppose Assumptions C1-C4 hold. Then Z_t and $\hat{\eta}_\mu$ converge to $P(W(r;d_0))$ and $K(W(r;d_0))$ in distribution as $n \to \infty$, where*

$$W(r;d) = W(r) - w(d)(\Gamma(2-d)\Gamma(d+1))^{-1} r^{1-d} W_{d+1}(1).$$

The limiting distribution of Z_t and $\hat{\eta}_\mu$ is affected by the limiting behavior of $\hat{\mu}(d_0)$. When $d = 0$, $W(r;d)$ reduces to a standard Brownian bridge. Table 1 lists the 1%, 5% and 10% lower tail quantiles of $P(W(r;d))$ and the upper tail quantiles of $K(W(r;d))$ with $w(d) = (1/2)[1 + \cos(4\pi d)]$ for $d \in [1/2, 3/4]$ from 500,000 replications. We set $n = 5,000$ and use the true value of d in taking the dth difference of X_t. For small d, the critical values are similar to those of the DF distribution and the ones in Table 1 of Kwiatkowski et al. (1992). As d increases, the critical values decrease in the absolute value initially, then increases once d is larger than 0.6, and changes little for $d \geq 1$. Figure 3 plots the density estimate of $P(W(r;d))$ for some values of d. The distribution of $P(W(r;d))$ is more concentrated to the left than that of $P(W(r))$.

When X_t has a linear time trend

$$X_t - \mu_0 - \beta_0 t = (1-L)^{-d_0} u_t \mathbf{1}\{t \geq 1\},$$

we simply apply the above procedure to the demeaned and detrended data. First, let $\dot{X}_t$ denote the demeaned detrended data:

$$\dot{X}_t = X_t - \tilde{\mu} - \tilde{\beta} t,$$

where $\tilde{\mu} = n^{-1}\sum_{t=1}^n X_t - \tilde{\beta}\bar{t}$, $\tilde{\beta} = (\sum_{t=1}^n (t-\bar{t})^2)^{-1}\sum_{t=1}^n (t-\bar{t})X_t$, and $\bar{t} = n^{-1}\sum_{t=1}^n t$. Then, apply the Z_t and KPSS test to

$$\tilde{u}_t = (1-L)^{\hat{d}}(\dot{X}_t - \dot{\mu}(\hat{d})) = \sum_{k=0}^{t-1} \frac{\Gamma(-\hat{d}+k)}{\Gamma(-\hat{d})k!}(X_{t-k} - \dot{\mu}(\hat{d})), \tag{9}$$

where $\dot{\mu}(\hat{d}) = w(d)n^{-1}\sum_{t=1}^n \dot{X}_t + (1-w(d))\dot{X}_1$. The FELW estimator satisfies assumption C2 if applied to $\dot{X}_t$ (Shimotsu, 2006, Theorem 5). The following lemma provides the limiting distribution of Z_t and $\hat{\eta}_\mu$ in this case.

Lemma 6 *Suppose Assumptions C1-C4 hold. Then Z_t and $\hat{\eta}_\mu$ constructed with $\tilde{u}_t$ in place of*

$\hat{u}_t$ *converge to* $P(W_2(r;d_0))$ *and* $K(W_2(r;d_0))$ *in distribution as* $n \to \infty$*, where*

$$W_2(r;d) = W(r) + \frac{r^{2-d}\xi(d)}{\Gamma(3-d)} + \frac{w(d)r^{1-d}}{\Gamma(2-d)}\left[\frac{W_{d+1}(1)}{\Gamma(d+1)} - \frac{\xi(d)}{2}\right],$$

where $\xi(d) = 12(\Gamma(d+1))^{-1}[-\int_0^1 W_{d+1}(r)dr + (1/2)W_{d+1}(1)]$.

Table 2 lists the 1%, 5% and 10% lower tail quantiles of $P(W_2(r;d))$ and the upper tail quantiles of $K(W_2(r;d))$ with $w(d) = (1/2)[1+\cos(4\pi d)]$ for $d \in [1/2, 3/4]$. Again, the critical values are similar to those of the DF distribution and the ones in Table 1 of Kwiatkowski et al. (1992) for small d.

6 Simulations and application to S&P 500 realized volatility

6.1 Simulation results from sample-splitting

Tables 3 and 4 report the mean of $\hat{d}$ and $\bar{d}$ and the rejection frequencies of the tests based on W and W_c with 5% asymptotic critical values for $n = 5,000$ and selected values of m. 10,000 replications were used.

Table 3 shows the simulation results when X_t follows an $I(d)$ process. The local Whittle estimator with $[\Delta_1, \Delta_2] = [-1/2, 1]$ is used in the first and second panels, and the FELW estimator is used in the third panel. The first panel report the results when $X_t = (1-L)^{-0.4}u_t$ with $u_t \sim iidN(0,1)$. Concurring Lemma 1, the mean of $\hat{d}$ and $\bar{d}$ are close to each other. The unmodified Wald statistic W tends to overreject, and its size distortion is substantial when $b = 8$. The modified Wald statistic W_c also tends to overreject, but its size distortion is much smaller than that of W, and it has a good size except when $b = 8$ and m is small. As discussed in Section 3, the test is expected to be relatively robust against the bias in semiparametric estimation arising from the short-run dynamics of the data. The second panel of Table 3 shows the results when $X_t = (1-L)^{-0.1}(1-0.6L)u_t$. The bias in $\hat{d}$ is substantial, exceeding 0.2 when m is large. On the other hand, the difference between $\hat{d}$ and $\bar{d}$ is small and not severely affected by the short-run dynamics. The test is not completely free from the effect of short-run dynamics, however, as the size distortion of W and W_c is larger than those in the first panel for larger values of m. The third panel of Table 3 shows the results when X_t follows a unit root process: $X_t = (1-L)^{-1}u_t\mathbf{1}\{t \geq 1\}$. The rejection frequencies of the W and W_c tests are similar to those in the first panel and not affected by the value of d_0.

Table 4 shows the results when X_t is generated by spurious $I(d)$ processes. The local Whittle estimator with $[\Delta_1, \Delta_2] = [-1/2, 1]$ is used in estimating d. The first panel report the results when X_t follows the mean-plus-noise model. The value of σ_ε^2 is fixed to 2, and the value of p is adjusted depending on m so that the mean of $\hat{d}$ is close to 0.4. The mean of $\bar{d}$ decreases monotonically as b increases, and the change in $\bar{d}$ is larger when m is smaller. The difference between the mean of $\hat{d}$ and $\bar{d}$ becomes as large as 0.2 when $m \leq 400$ and $b = 8$. The test based on W_c appears to have a good power. The power of the test does not increase substantially even when m increases. The test has larger power for larger b, but, in view of the size distortion documented in Table 3, the test with $b = 8$ may not be very useful.

In the second panel, X_t is generated by the STOPBREAK model with $\sigma_\varepsilon^2 = 2$. The value of γ is changed depending on m so that the mean of $\hat{d}$ is close to 0.4. Similar to the first panel, the mean of $\bar{d}$ decreases monotonically as b increases, and the change in $\bar{d}$ is larger when m is smaller. Again, the difference between the mean of $\hat{d}$ and $\bar{d}$ is close to 0.2 when $b = 8$ and $m = 200$. The power of the test is weaker than in the first panel, particularly for small values of m. Interestingly, larger values of b do not necessarily lead to increases in the power. This reinforces the view that choosing $b = 2$ or 4 may be preferable.

The third panel reports the results from the Markov-switching model. As in the other spurious long-memory models, the mean of $\bar{d}$ decreases as b increases, but its magnitude is smaller than the other two models. Despite the smaller difference between the mean of $\hat{d}$ and $\bar{d}$, the test based on W_c has good power, particularly when m is moderate to large. This is because the value of each $\hat{d}^{(a)}$ differs more across a in the Markov switching model than in other models.

Gourieroux and Robert (2005) develop a stochastic unit root model. This model has two regimes, a random walk regime and a stationary regime, and the time spent in each regime is determined endogenously. Specifically, X_t is modeled by

$$X_t = \begin{cases} \mu + X_{t-1} + \varepsilon_t, & \text{with probability } \pi_{rw}(X_{t-1}), \\ \mu + \varepsilon_t & \text{with probability } 1 - \pi_{rw}(X_{t-1}), \end{cases}$$

where ε_t is iid with mean zero, and $\pi_{rw}(x)$ is a cadlag increasing function with $\lim_{x\to-\infty} \pi_{rw}(x) > 0$ and $\lim_{x\to\infty} \pi_{rw}(x) < 1$. Gourieroux and Robert (2005) show that this model can produce hyperbolically decaying autocorrelogram of X_t. The fourth panel shows the results when X_t is generated by a stochastic unit root model with $\mu = 0$, $\varepsilon_t \sim iidN(0, 1)$ and $\pi_{rw}(x) = \alpha(1 - x^{-2})1\{x \geq 1\}$ (cf. Gourieroux and Robert, 2005, p. 18). The results are similar to those in the third panel. The difference between the mean of $\hat{d}$ and $\bar{d}$ is small, but test based on W_c has good power.

6.2 Simulation results from dth differencing

Table 5 reports the size of the test based on Z_t and $\hat{\eta}_\mu$ for the three types $I(d)$ processes used in Table 3. Lag length l is set to $m^{0.4}$, and the critical values are chosen by interpolating the critical values in Table 1. The first panel shows the results when $X_t = (1-L)^{-0.4} u_t \mathbf{1}\{t \geq 1\}$ with $u_t \sim iidN(0, 1)$. The test has good size overall, albeit they are undersized in some cases. The underrejection is due to the error in estimating d, and it improves for larger m. The second panel shows the results with $X_t = (1 - L)^{-0.2} \varepsilon_t \mathbf{1}\{t \geq 1\}$ with $(1 - 0.6L)\varepsilon_t \sim iidN(0, 1)$. The estimates of d are upwardly biased from the short-run dynamics of the process, and the bias increases as m increases. The tests correctly reject the null of $I(\hat{d})$ with those biased values of $\hat{d}$. This suggests the possibility of using the dth differencing tests to check whether an (semiparametric) estimate of d is contaminated from short-run dynamics. The third panel reports the results when X_t follows a unit root process. The rejection frequencies are similar to those in the first panel.

Table 6 reports the rejection frequencies when X_t is generated by spurious $I(d)$ processes. In this table, the value of the tuning parameters such as p and γ is fixed across different m. The mean of $\hat{d}$ changes as m changes, which is another feature of spurious $I(d)$ processes. As shown in the first and second panels, the KPSS test has very strong power against the mean-plus-noise model and the STOPBREAK model, while the Z_t test has no power. This is because these processes are essentially $I(1)$ processes, and the residuals $\hat{u}_t$ from $\hat{d}$th differencing are $I(\alpha)$ processes with $\alpha > 0$. Consequently, the KPSS test correctly rejects the null of $I(0)$ of $\hat{u}_t$. On the other hand, the partial sums of $\hat{u}_t$ behave like an unstable AR(1) process, and the majority of the Z_t statistic take *positive* values. The third and fourth panels show the results with the Markov-switching model and the stochastic unit root model. Z_t test has very strong power, but the KPSS test has no power. This is because these processes are essentially an $I(0)$ process, and its $\hat{d}$th difference is an overdifferenced process.

In sum, combining the Z_t and KPSS test provides a powerful test against both types ($I(0)$ and $I(1)$) of spurious long memory models.

6.3 Simulation results with $n = 240$

Although the primary motivation of the proposed tests is application to financial datasets, we investigate their applicability to datasets with a smaller sample size, $n = 240$. Table 7 reports the rejection frequency of the proposed tests with the three types of $I(d)$ processes used in Table 5. The overall results are similar to those in Tables 3 and 5. The modified Wald statistic W_c tends to overreject but is relatively robust against the bias arising from the short-run dynamics. The test based on Z_t and $\hat{\eta}_\mu$ tend to underreject and also correctly reject the null of $I(\hat{d})$ when $\hat{d}$ is biased.

Table 8 reports the rejection frequency with spurious $I(d)$ processes. Not surprisingly, the test based on W_c does not have strong power, with the exception of the stochastic unit root model. The Z_t and KPSS tests are able to detect spurious $I(d)$ models, however, in particular when $m \geq 40$.

6.4 Application to S&P 500 realized volatility

We apply the proposed tests to the logarithm of the realized volatility of the S&P 500 index. We construct a realized volatility series (Andersen et al. (2001a)) from thirty-minute intraday returns on the S&P 500 index obtained from Tick Data, Inc. Thirty-minute interval was chosen to balance between the accuracy of the approximation of the continuous record asymptotics and the bias from market microstructure problems (see Andersen et al. (2003), Deo et al. (2006)).

The returns are computed as the difference of the logarithm of the prices using the closing prices from 10:00 a.m. to 3:00 p.m. Let p_τ denote the logarithm of the price of the S&P 500 index at time τ, and define the h-period return at time τ as $r_{\tau+h,h} = p_{\tau+h} - p_\tau$. Then, the realized daily volatility of day t is calculated as $v_t^2 = \sum_{j=1}^{1/\Delta} r_{t+j\Delta,\Delta}^2$ (Andersen et al. (2001b)). In other words, the realized volatility of day t is defined as the sum of the squares of 11 returns in day t, where the 11th return is defined as the overnight return computed using the closing price at 3:00 p.m. on day t and the 10:00 a.m. closing price on day $t+1$. The dataset consists of 5,000 observations spanning from 1/2/1985 to 10/25/2004.

Table 9 shows the estimates $\hat{d}$ and $\bar{d}$, the value of W_c, Z_t, and $\hat{\eta}_\mu$ statistic from $X_t = \log |v_t|$ for various values of $m \in [200, 800]$ and $b = \{1, 2, 4\}$. The value of $\hat{d}$ and $\bar{d}$ are close to each other, and the W_c test rejects the null of the constancy of d only for $m = 600$. The Z_t and $\hat{\eta}_\tau$ statistic reject the null of $I(d)$ only with $d \leq 0.41$. Overall, the evidence against true $I(d)$ is not strong in view of the results in Table 4. However, both $\hat{d}$ and $\bar{d}$ decrease as m increases, which is consistent with the mean-plus-noise model and STOPBREAK model. This suggests a possibility of a presence of jumps and/or structural breaks in the data.

In Table 10, we split the data into five subperiods of length 1,000 and apply the same tests as in Table 9 to each subperiod. The estimates of d vary across subperiods, partly due to sampling variation and small m. Note that the asymptotic standard error is 0.155 and 0.098 for $m = 40$ and 160, respectively. The Z_t and $\hat{\eta}_\tau$ statistic do not reject the null of $I(\hat{d})$ in most cases, but the null of the constancy of d is strongly rejected in the first and fourth subperiods.

The results in Tables 9 and 10 do not show a strong evidence against true $I(d)$, even though they suggest a strong possibility of local variation in d. They also suggest some possibility of the presence of jumps and/or structural breaks, but not as overwhelming as would be observed if the data were generated by those spurious $I(d)$ models suggested in the literature. These results appear to concur with the view of Granger and Hyung (2004) and Zivot and Choi (2005); a pure $I(d)$ process may not explain all of the persistence in the logarithm of the realized volatility, but the data do not support an extreme view that structural breaks account for all the observed persistence.

7 Conclusion

This paper proposes two simple tests to distinguish true and spurious $I(d)$ processes that accommodate both (asymptotically) stationary and nonstationary $I(d)$ processes. The proposed tests are based on certain time domain properties of true $I(d)$ processes which spurious $I(d)$ processes do not mimic. Simulation results show that both tests have good power for the sample size of 5,000. One of the tests has good power even with a sample size as small as 240.

When applied to the daily realized volatility of the S&P 500 index, these test show some evidence of infrequent structural breaks in the data. The evidence against true $I(d)$ is not strong, however, and the overall results suggest that the data generating process is somewhere in between these two polar models of observed long memory property. Conducting a similar experiment to Zivot and Choi (2005) with a realized volatility dataset may provide further insight in this respect.

8 Appendix: Proofs and technical results

8.1 Proof of Lemma 1

The stated result follows if we show

$$\sqrt{m/b}\begin{pmatrix} \hat{d}^{(1)} - d_0 \\ \vdots \\ \hat{d}^{(b)} - d_0 \end{pmatrix} \to_d N\left(0, \frac{1}{4}I_b\right).$$

From the Cramér-Wald device, this holds if, for any $b \times 1$ vector $\eta = (\eta_1, \ldots, \eta_b)'$,

$$\sqrt{m/b}\left(\eta_1(\hat{d}^{(1)} - d_0) + \cdots + \eta_b\,(\hat{d}^{(b)} - d_0)\right) \to_d N\left(0, \frac{1}{4}\left(\eta_1^2 + \cdots + \eta_b^2\right)\right). \tag{10}$$

Since each subsample and the bandwidth m/b satisfy Assumptions A1′-A4′ of Robinson (1995), we can apply the arguments in Robinson (1995) leading to its page 1644 to obtain, for $a = 1, \ldots, b$,

$$\sqrt{m/b}(\widehat{d}^{(a)} - d_0) = -\left(\frac{1}{4} + o_p(1)\right) 2(m/b)^{-1/2} \sum_{j=1}^{m/b} \tilde{\nu}_j \left(2\pi I_\varepsilon^{(a)}(\tilde{\lambda}_j) - 1\right), \tag{11}$$

where $\tilde{\nu}_j = \log j - (m/b)^{-1}\sum_{j=1}^{m/b}\log j$ with $\sum_{j=1}^{m/b}\tilde{\nu}_j = 0$ and $I_\varepsilon^{(a)}(\tilde{\lambda}_j) = (2\pi n)^{-1}|\sum_{t=(a-1)n/b+1}^{an/b}\varepsilon_t e^{it\tilde{\lambda}_j}|^2$. From Robinson (1995, p.1644),

$$(m/b)^{-1/2}\sum_{j=1}^{m/b}\tilde{\nu}_j\left(2\pi I_\varepsilon^{(a)}(\tilde{\lambda}_j) - 1\right) = \sum_{t=(a-1)n/b+1}^{an/b} z_t^{(a)} + o_p(1), \tag{12}$$

where $z_{(a-1)n/b+1}^{(a)} = 0$ and, for $(a-1)n/b + 2 \le t \le an/b$, $z_t^{(a)} = \varepsilon_t \sum_{s=(a-1)n/b+1}^{t-1}\varepsilon_s\tilde{c}_{t-s}$ and $\tilde{c}_s = 2(n/b)^{-1}(m/b)^{-1/2}\sum_{j=1}^{m/b}\tilde{\nu}_j\cos(s\tilde{\lambda}_j)$.

Hence, (10) follows if we show

$$\sum_{a=1}^{b} \eta_a \sum_{t=(a-1)n/b+1}^{an/b} z_t^{(a)} \to_d N\left(0, \left(\eta_1^2 + \cdots + \eta_b^2\right)\right). \tag{13}$$

By a standard martingale CLT, (13) follows if

$$\sum_{a=1}^{b} \eta_a^2 \sum_{t=(a-1)n/b+1}^{an/b} E((z_t^{(a)})^2|F_{t-1}) - \left(\eta_1^2 + \cdots + \eta_b^2\right) \quad \to \quad {}_p 0, \tag{14}$$

$$\sum_{a=1}^{b} \sum_{t=(a-1)n/b+1}^{an/b} E\left((z_t^{(a)})^2 \mathbf{1}\{|z_t| > \delta\}\right) \quad \to \quad 0 \quad \text{for all } \delta > 0, \tag{15}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. Replacing (n, m) in the proof of equations (4.12) and (4.13) of Robinson (1995) with $(n/b, m/b)$ and applying it to $\sum_{t=(a-1)n/b+1}^{an/b} E((z_t^{(a)})^2|F_{t-1})$ and $\sum_{t=(a-1)n/b+1}^{an/b} E((z_t^{(a)})^2 \mathbf{1}\{|z_t| > \delta\})$ gives

$$\sum_{t=(a-1)n/b+1}^{an/b} E((z_t^{(a)})^2|F_{t-1}) - 1 \to_p 0, \qquad \sum_{t=(a-1)n/b+1}^{an/b} E\left((z_t^{(a)})^2 \mathbf{1}\{|z_t| > \delta\}\right) \to 0 \quad \text{for all } \delta > 0,$$

for $a = 1, \ldots, b$. Therefore, (14) and (15) follow, and we establish (10). □

8.2 Proof of Lemma 2

From Slutzky's theorem and Theorem 9.2.2 of Rao and Mitra (1971), W and W_c have the stated limiting distribution if, for some constant B,

$$\sqrt{m} \begin{pmatrix} \hat{d} - d_0 \\ \hat{d}^{(1)} - d_0 \\ \vdots \\ \hat{d}^{(b)} - d_0 \end{pmatrix} = \frac{Z_n}{2} + B_n + o_p(1), \quad Z_n \to_d N(0, \Omega), \quad B_n = B\iota_{b+1} m^{1/2+\beta} n^{-\beta}. \tag{16}$$

Let $DR_0 = ((\partial/\partial d)R(d_0), (\partial/\partial d)R^{(1)}(d_0), \cdots, (\partial/\partial d)R^{(b)}(d_0))'$. Using Taylor expansion, (16) is proven if we show

$$\begin{aligned} DR_0 &= 2Z_n + 4B\iota_{b+1} m^{1/2+\beta} n^{-\beta} + o_p(1), && (17) \\ (\partial^2/\partial^2 d)R(\tilde{d}) &= 4 + O_p(m^{-1/6}), \quad \tilde{d} \in [\hat{d}, d_0] && (18) \\ (\partial^2/\partial^2 d)R^{(a)}(\tilde{d}^{(a)}) &= 4 + O_p(m^{-1/6}), \quad \tilde{d}^{(a)} \in [\hat{d}^{(a)}, d_0], \quad a = 1, \ldots, b. && (19) \end{aligned}$$

The order of magnitude, $O_p(m^{-1/6})$, on the right hand side of (18) and (19) is required so that $m^{1/2+\beta}n^{-\beta}[(\partial^2/\partial^2 d)R(\tilde{d})-4]^{-1} \to_p 0$ holds.

We prove (18) and (19) first. Define $\hat{E}_k(d) = m^{-1}\sum_{j=1}^m (\log j)^k j^{2d} I_x(\lambda_j)$, which corresponds to $\hat{E}_k(H)$ on p. 1641 of Robinson (1995) with $d = H - 1/2$. Then, as in (4.3) in Robinson (1995), we obtain

$$(\partial^2/\partial^2 d)R(d) = 4[\hat{E}_2(d)\hat{E}_0(d) - \hat{E}_1^2(d)]/\hat{E}_0^2(d).$$

Consequently, if we show

$$\hat{E}_k(\tilde{d}) - \hat{E}_k(d_0) = o_p(n^{2d_0}m^{-1/6}\log^{-k} m), \tag{20}$$

then (18) follows from repeating the argument in pp. 1643-44 of Robinson (1995) in conjunction with $m^{-1}\sum_{j=1}^m(\log j)^2 - (m^{-1}\sum_{j=1}^m \log j)^2 = 1 + O(m^{-1}\log^2 m)$, which follows straightforwardly from the Euler-Maclaurin summation formula. See Andrews and Sun (2004, p. 600) for the explanation on why the term $\log^{-k} m$ is necessary on the right hand side of (20).

In order to show (20), fix $\varepsilon > 0$ first. Following the argument in Robinson (1995) p. 1642, on the set $M = \{d : m^{1/6}(\log m)^5|d - d_0| \leq \varepsilon\}$ we have

$$|\hat{E}_k(d) - \hat{E}_k(d_0)| \leq 2e\varepsilon m^{-1/6}\log^{k-4} m \hat{E}_0(d_0).$$

Therefore, in place of (4.4) of Robinson (1995), we have

$$\begin{aligned} &\Pr\left(|\hat{E}_k(\tilde{d}) - \hat{E}_k(d_0)| > \eta(2\pi/n)^{-2d_0}(\log m)^{-k}m^{-1/6}\right) \\ \leq\ &\Pr\left(\hat{G}(d_0) > \eta(2e\varepsilon)^{-1}\log^{4-2k} m\right) + \Pr\left(m^{1/6}(\log^5 m)|\tilde{d} - d_0| > \varepsilon\right). \end{aligned} \tag{21}$$

The first probability in (21) tends to 0 for sufficiently small ε because $\hat{G}(d_0) \to_p G_0$. In view of (4.4), (4.5), and (3.4) of Robinson (1995), the second probability in (21) tends to 0 if

$$\sup_{d\in\Theta_1} |[\hat{G}(d) - G(d)]/G(d)| = o_p(m^{-1/3}\log^{-10} m), \tag{22}$$

where $G(d) = G_0 m^{-1}\sum_{j=1}^m \lambda_j^{2d-2d_0}$ and $\Theta_1 = \{d : d_0 - 1/2 + \Delta \leq d \leq \Delta_2\}$ for arbitrary small $\Delta > 0$. (22) follows from equations (4.7), (4.8) and (4.9) of Robinson (1995) and the assumption $m^{\beta+1/3}n^{-\beta}\log^{10} m \to 0$, and we show (20). (19) is proven in the same way.

We proceed to show (17). Taking the derivative of $R(d_0)$ and $R^{(a)}(d_0)$ and using $\hat{G}(d_0) = G_0 + O_p(m^{-1/6})$ (from (22)) gives

$$\sqrt{m}\frac{\partial R(d_0)}{\partial d} = (G_0 + O_p(m^{-1/6}))^{-1} 2m^{-1/2} \sum_{j=1}^{m} \nu_j \lambda_j^{2d_0} I_x(\lambda_j), \tag{23}$$

$$\sqrt{m/b}\frac{\partial R^{(a)}(d_0)}{\partial d} = (G_0 + O_p(m^{-1/6}))^{-1} 2(m/b)^{-1/2} \sum_{j=1}^{m/b} \tilde{\nu}_j \tilde{\lambda}_j^{2d_0} I_x^{(a)}(\tilde{\lambda}_j), \tag{24}$$

where $\nu_j = \log j - m^{-1}\sum_{j=1}^m \log j$ with $\sum_{j=1}^m \nu_j = 0$, and $\tilde{\nu}_j$ is defined in the proof of Proposition 1. Define $Z_n^{(0)} = m^{-1/2}\sum_{j=1}^m \nu_j(2\pi I_\varepsilon(\lambda_j)-1)$ and $Z_n^{(a)} = bm^{-1/2}\sum_{j=1}^{m/b} \tilde{\nu}_j(2\pi I_\varepsilon^{(a)}(\tilde{\lambda}_j)-1)$. Then, (17) follows if we show, for $a = 1, \ldots, b$,

$$G_0^{-1} m^{-1/2} \sum_{j=1}^{m} \nu_j \lambda_j^{2d_0} I_x(\lambda_j) = Z_n^{(0)} + 2Bm^{1/2+\beta} n^{-\beta} + o_p(1), \tag{25}$$

$$bG_0^{-1} m^{-1/2} \sum_{j=1}^{m/b} \tilde{\nu}_j \tilde{\lambda}_j^{2d_0} I_x^{(a)}(\tilde{\lambda}_j) = Z_n^{(a)} + 2Bm^{1/2+\beta} n^{-\beta} + o_p(1), \tag{26}$$

and we show

$$(Z_n^{(0)}, Z_n^{(1)}, \cdots, Z_n^{(b)})' \to_d N(0, \Omega). \tag{27}$$

We show (25) by showing

$$G_0^{-1} m^{-1/2} \sum_{j=1}^{m} \nu_j \lambda_j^{2d_0} I_x(\lambda_j) - m^{-1/2} \sum_{j=1}^{m} \nu_j f(\lambda_j)^{-1} I_x(\lambda_j) = 2Bm^{1/2+\beta} n^{-\beta} + o_p(1), \tag{28}$$

$$m^{-1/2} \sum_{j=1}^{m} \nu_j f(\lambda_j)^{-1} I_x(\lambda_j) - m^{-1/2} \sum_{j=1}^{m} \nu_j 2\pi I_\varepsilon(\lambda_j) = o_p(1). \tag{29}$$

Since $f(\lambda) = G_0 \lambda^{-2d}(1 + E_\beta \lambda^\beta + O(\lambda^{\beta+1}))$, the left hand side of (28) is

$$m^{-1/2} E_\beta \sum_{j=1}^{m} \nu_j \lambda_j^\beta f(\lambda_j)^{-1} I_x(\lambda_j) + m^{-1/2} \sum_{j=1}^{m} \nu_j O(\lambda_j^{\beta+1}) f(\lambda_j)^{-1} I_x(\lambda_j).$$

The second term is $O_p(m^{3/2+\beta} n^{-1-\beta} \log m) = o_p(1)$. Rewrite the first term as

$$m^{-1/2} E_\beta \sum_{j=1}^{m} \nu_j \lambda_j^\beta \left[f(\lambda_j)^{-1} I_x(\lambda_j) - 2\pi I_\varepsilon(\lambda_j)\right] + m^{-1/2} E_\beta \sum_{j=1}^{m} \nu_j \lambda_j^\beta 2\pi I_\varepsilon(\lambda_j). \tag{30}$$

Using the proof of (4.8) in Robinson (1995, p. 1648), we can show, for $1 \le r \le m$,

$$\sum_{j=1}^{r} \left(f(\lambda_j)^{-1} I_x(\lambda_j) - 2\pi I_\varepsilon(\lambda_j)\right) = O_p(r^{1/3}(\log r)^{2/3} + r^{1/2} n^{-1/4}). \tag{31}$$

Applying (31), (4.9) of Robinson (1995), and summation by parts, we deduce that (30) equals (cf. Henry and Robinson, 1996, p. 223, Phillips and Shimotsu, 2004, p. 686)

$$m^{-1/2}E_\beta \sum_{j=1}^{m} \nu_j \lambda_j^\beta + o_p(1) = m^{1/2}\frac{E_\beta \beta}{(\beta+1)^2}\lambda_m^\beta + o_p(1),$$

and (28) holds with $2B = (\beta+1)^{-2}E_\beta \beta (2\pi)^\beta$. (29) follows from (31), and we establish (25). An identical argument gives (26).

It remains to show (27). From (12) and Robinson (1995, p. 1644), for $a = 1, \dots, b$,

$$Z_n^{(0)} = \sum_{t=1}^{n} z_t + o_p(1), \quad Z_n^{(a)} = b^{1/2} \sum_{t=(a-1)n/b+1}^{an/b} z_t^{(a)} + o_p(1),$$

where $z_1 = 0$ and, for $t \geq 2$, $z_t = \varepsilon_t \sum_{s=1}^{t-1} \varepsilon_s c_{t-s}$ with $c_s = 2n^{-1}m^{-1/2}\sum_{j=1}^{m} \nu_j \cos(s\lambda_j)$. $z_t^{(a)}$ is defined in the proof of Proposition 1. Therefore, it suffices to show, for any $b \times 1$ vector $\eta = (\eta_1, \dots, \eta_b)'$,

$$\sum_{t=1}^{n} z_t + \sum_{a=1}^{b} \eta_a \sum_{t=(a-1)n/b+1}^{an/b} b^{1/2} z_t^{(a)} \to_d N\left(0, H(b,\eta)\right), \tag{32}$$

where $H(b,\eta) = 1 + 2(\eta_1 + \cdots + \eta_b) + b(\eta_1^2 + \cdots + \eta_b^2)$. By a standard martingale CLT, (32) holds if

$$\sum_{a=1}^{b} \sum_{t=(a-1)n/b+1}^{an/b} E((z_t + \eta_a b^{1/2} z_t^{(a)})^2 | F_{t-1}) - H(b,\eta) \quad \to \quad _p 0, \tag{33}$$

$$\sum_{a=1}^{b} \sum_{t=(a-1)n/b+1}^{an/b} E\left((z_t + z_t^{(a)})^2 \mathbf{1}\{|z_t + z_t^{(a)}| > \delta\}\right) \quad \to \quad 0 \quad \text{for all } \delta > 0. \tag{34}$$

(34) follows from (4.13) of Robinson (1995) and (15). Because $\sum_{a=1}^{b}\sum_{t=(a-1)n/b+1}^{an/b} E(z_t^2|F_{t-1}) \to_p 1$ from (4.12) of Robinson (1995), and $\sum_{a=1}^{b}\sum_{t=(a-1)n/b+1}^{an/b} E(\eta_a b^{1/2} z_t^{(a)})^2 | F_{t-1}) \to_p b(\eta_1^2 + \cdots + \eta_b^2)$ from (14), we establish (32) if we show, for $a = 1, \dots, b$,

$$\sum_{t=(a-1)n/b+1}^{an/b} E(z_t b^{1/2} z_t^{(a)} | F_{t-1}) \to_p 1. \tag{35}$$

From Proposition 1, the left hand side is equal to $b^{1/2}\sum_{t=(a-1)n/b+1}^{an/b-1}\sum_{s=(a-1)n/b+1}^{an/b-t} c_s \tilde{c}_s + o_p(1)$, which converges to 1 in probability from Proposition 2. Thus we show (32) and complete the proof. □

8.3 Proof of Lemma 3

For part (a), it suffices to show that (11) holds when X_t is generated by (4). The Taylor expansion gives

$$0 = \partial R^{(a)}(\hat{d}^{(a)})/\partial d = \partial R^{(a)}(d_0)/\partial d + [\partial^2 R^{(a)}(\breve{d})/\partial d^2](\hat{d}^{(a)} - d_0),$$

where $\breve{d} \in [\hat{d}^{(a)}, d_0]$. Shimotsu and Phillips (2006, pp. 230-231) show $\partial^2 R^{(a)}(\breve{d})/\partial d^2 \to_p 4$. Shimotsu and Phillips (2006) show in line 13 on page 232 that

$$\partial R^{(a)}(d_0)/\partial d = 2(m/b)^{-1/2} \sum_{j=1}^{m/b} \tilde{\nu}_j \left(2\pi I_{\varepsilon}^{(a)}(\tilde{\lambda}_j) - 1\right) + o_p(1),$$

and (11) follows.

For part (b), it suffices to show that (11) holds when X_t is generated by (4) and $\hat{d}^{(a)}$ is the FELW estimator from the ath subsample. The Taylor expansion gives

$$\widehat{d}^{(a)} - d_0 = -\frac{\partial R_F^{(a)}(d_0)/\partial d}{\partial^2 R_F^{(a)}(\breve{d})/\partial d^2} + \left[1 - \frac{\partial^2 R_F^{(a)}(\breve{d})/\partial d^2}{\partial^2 R_F^{(a)}(\tilde{d})/\partial d^2}\right](\tilde{d} - d_0).$$

where $R_F^{(a)}(d)$ is constructed using the ath subsample and $\breve{d} \in [\hat{d}^{(a)}, d_0]$. From the proof of Theorem 5 of Shimotsu (2006, p. 12), we have $\partial^2 R_F^{(a)}(d)/\partial d^2 \to_p 4$ for all d such that $|d - d_0| \leq |\tilde{d} - d_0|$. Consequently, we obtain $\widehat{d}^{(a)} - d_0 = -(1/4 + o_p(1))\ \partial R_F^{(a)}(d_0)/\partial d + o_p(m^{-1/2})$.

We derive the limit of $\partial R_F^{(a)}(d_0)$ in two steps. Let $R_E^{(a)}(d)$ denote the objective function $R_F^{(a)}(d)$ when $\tilde{\mu}(d) = \mu_0$, i.e., when the mean is known. In other words, $R_E^{(a)}(d)$ is the objective function of the exact local Whittle estimator from the ath subsample. First, Shimotsu and Phillips (2005, p.1918) show $(m/b)^{1/2} \partial R_E^{(a)}(d_0) = 2(m/b)^{-1/2} \sum_{j=1}^{m/b} \tilde{\nu}_j (2\pi I_{\varepsilon}^{(a)}(\tilde{\lambda}_j) - 1)$. Second, the proof of Theorem 3b of Shimotsu (2006, p. 23) in conjunction with its proof of Theorems 1b and 2b shows $\partial R_F^{(a)}(d_0) = \partial R_E^{(a)}(d_0) + o_p(m^{-1/2})$. Therefore, (11) follows. □

8.4 Proof of Lemma 4

We show part (a) first. In view of the proof of Lemma 2, it suffices to show (17)-(19) hold under the current set of assumptions. First, note that (18) and (19) follow if (22) holds. To show (22), the equation in the last line of page 230 of Shimotsu and Phillips (2006) provide a counterpart of (4.8) of Robinson (1995) for X_t generated by (4):

$$E \left| \sum_{j=1}^{r} (G_0^{-1} \lambda_j^{2d_0} I_x(\lambda_j) - 2\pi I_{\varepsilon}(\lambda_j)) \right| = O(r^{1/2} \log r + r^{\beta+1} n^{-\beta}). \tag{36}$$

Since (4.7) and (4.9) of Robinson (1995) still hold under the current set of assumptions, (22) follows from combining (36) with (4.7) and (4.9) of Robinson (1995).

The proof of part (a) completes if we show (17), which holds if (25) and (26) hold with $Z_n^{(0)}$ and $Z_n^{(a)}$ defined above them. First, observe that (25) holds if $\lambda_j^{2d_0} I_x(\lambda_j)$ is replaced with $I_u(\lambda_j)$, because u_t satisfies Assumptions A1$'$, A2 and A3 with $d_0 = 0$. Second, we have (from Shimotsu and Phillips, 2006, pp. 231-232 for $d_0 \in (-1/2, 1/2)$, and from Shimotsu and Phillips, 2005, pp. 685-686 for $d_0 \in (1/2, 3/4)$)

$$\sum_{j=1}^{m} \nu_j \lambda_j^{2d_0} I_x(\lambda_j) - \sum_{j=1}^{m} \nu_j I_u(\lambda_j) = o_p(m^{-1/2}).$$

Therefore, (25) follows. An identical argument gives (26), and we complete the proof of part (a).

For part (b), let $\hat{d}$ denote the FELW estimator, and let $\hat{d}^{(a)}$ denote the FELW estimator from the ath subsample. The stated result follows if we show

$$\sqrt{m}(\hat{d} - d_0, \hat{d}^{(1)} - d_0, \cdots \hat{d}^{(b)} - d_0)' = W_n/2 + o_p(1), \quad W_n \to_d N(0, \Omega). \tag{37}$$

From the proof of Lemma 3, we have, for $a = 1, \ldots, b$,

$$\sqrt{m/b}(\widehat{d}^{(a)} - d_0) = -\frac{1}{2}(m/b)^{-1/2} \sum_{j=1}^{m/b} \tilde{\nu}_j (2\pi I_\varepsilon^{(a)}(\tilde{\lambda}_j) - 1) + o_p(1),$$

and using a similar argument gives $\sqrt{m}(\widehat{d} - d_0) = -(1/2)m^{-1/2} \sum_{j=1}^{m} \nu_j (2\pi I_\varepsilon(\lambda_j) - 1) + o_p(1)$. Consequently, the left hand side of (37) can be written as $-(W_n^{(0)}, W_n^{(1)}, \cdots, W_n^{(b)})/2 + o_p(1)$, where $W_n^{(0)} = m^{-1/2} \sum_{j=1}^{m} \nu_j (2\pi I_\varepsilon(\lambda_j) - 1)$ and $W_n^{(a)} = bm^{-1/2} \sum_{j=1}^{m/b} \tilde{\nu}_j (2\pi I_\varepsilon^{(a)}(\tilde{\lambda}_j) - 1)$. The stated result then follows because $(W_n^{(0)}, W_n^{(1)}, \cdots, W_n^{(b)}) \to_d N(0, \Omega)$ from (27) in the proof of Lemma 2. □

8.5 Proof of Lemma 5

From Phillips and Perron (1988) and Kwiatkowski et al. (1992), the stated result follows if

$$n^{-1/2} \sum_{t=1}^{[nr]} \left[(1-L)^{\hat{d}} (X_t - \hat{\mu}(\hat{d})) \right] \Rightarrow \omega W(r; d_0), \quad \omega^2 = 2\pi f_u(0), \quad r \in [0,1], \tag{38}$$

$$s^2(l) \to_p \omega^2. \tag{39}$$

First, we show (38). Assume $\mu_0 = 0$ without loss of generality. Since $X_t = (1-L)^{-d_0} u_t \mathbf{1}\{t \geq 1\}$, we have

$$\overline{X} = n^{-1} \sum_{t=1}^{n} X_t = n^{-1}(1-L)^{-d_0-1} u_n \mathbf{1}\{t \geq 1\}.$$

Define $\theta = \hat{d} - d_0$ and $v_t = \mathbf{1}\{t \geq 1\}$. If we apply $(1-L)^{\hat{d}}$ to $(X_t - \hat{\mu}(\hat{d}))$, its partial sum equals

$$\sum_{t=1}^{[nr]} \left[(1-L)^{\hat{d}} (X_t - \hat{\mu}(\hat{d})) \right] = (1-L)^{\theta-1} u_{[nr]} \mathbf{1}\{t \geq 1\} - \hat{\mu}(\hat{d})(1-L)^{\hat{d}-1} v_{[nr]}. \tag{40}$$

Therefore, (38) follows if we show

$$n^{-1/2}(1-L)^{\theta-1} u_{[nr]} \mathbf{1}\{t \geq 1\} \Rightarrow \omega W(r), \quad r \in [0,1], \tag{41}$$

$$n^{-1/2} \hat{\mu}(\hat{d})(1-L)^{\hat{d}-1} v_{[nr]} \Rightarrow w(d_0)(\Gamma(2-d_0)\Gamma(d_0+1))^{-1} \omega r^{1-d_0} W_{d_0+1}(1). \tag{42}$$

We show (41) first. Define $a_k(\theta) = (1-\theta)_k / k! - 1 = \Gamma(1-\theta+k)/[\Gamma(1-\theta)\Gamma(k+1)] - 1$, then the left hand side of (41) is

$$n^{-1/2}(1-L)^{\theta-1} u_{[nr]} \mathbf{1}\{t \geq 1\} = n^{-1/2} \sum_{k=0}^{[nr]-1} a_k(\theta) u_{[nr]-k} + n^{-1/2} \sum_{k=0}^{[nr]-1} u_{[nr]-k}.$$

The second term on the right converges to $\omega W(r)$ from (8). For the first term on the right, summation by parts gives

$$\sum_{k=0}^{[nr]-1} a_k(\theta) u_{[nr]-k} = \sum_{k=0}^{[nr]-2} (a_k(\theta) - a_{k+1}(\theta)) \sum_{q=0}^{k} u_{[nr]-q} + a_{[nr]-1}(\theta) \sum_{q=0}^{[nr]-1} u_{[nr]-q}. \tag{43}$$

From Phillips and Shimotsu (2004, p. 670),

$$a_k(\theta) - a_{k+1}(\theta) = \frac{\theta \Gamma(k+1-\theta)}{\Gamma(1-\theta)\Gamma(k+2)}.$$

Recall, if $|\theta| \leq 1/\log n$, for $1 \leq k \leq n$ we have $k^{|\theta|} \leq k^{1/\log n} \leq n^{1/\log n} = e$. Since $\Gamma(k+\alpha)/\Gamma(k+\beta) = k^{\alpha-\beta}(1+O(k^{-1}))$ (Erdélyi, 1953, p.47), we have

$$a_k(\theta) - a_{k+1}(\theta) = \theta \cdot O(k^{-\theta-1}) = \theta \cdot O(k^{-1}) \text{ for } |\theta| \leq 1/\log n,$$

and Taylor expansion gives $a_k(\theta) = (\Gamma(1-\theta))^{-1} k^{-\theta}(1+O(k^{-1})) - 1 = \theta \cdot O(\log(k+1)) + O(k^{-1})$.

In view of $E(\sum_{k=0}^{q} u_{[nr]-k})^2 = O(q)$ for $q = 1, \ldots, [nr]-1$, it follows that

$$\sum_{k=0}^{[nr]-1} a_k(\theta) u_{[nr]-k} = \theta \cdot O_p\left(\sum_{k=0}^{[nr]-2} (k+1)^{-1/2} + n^{1/2}\log n\right) + O_p\left(n^{-1/2}\right) = o_p\left(n^{1/2}\right),$$

and (41) follows.

For (42), first note that

$$n^{1/2-d_0}\hat{\mu}(\hat{d}) = n^{1/2-d_0} w(\hat{d})\overline{X} + n^{1/2-d_0}(1-w(\hat{d}))X_1.$$

$n^{1/2-d_0} w(\hat{d})\overline{X} \to_d w(d_0)(\Gamma(d_0+1))^{-1}\omega W_{d_0+1}(1)$ from (7) and $\hat{d} \to_p d_0$. Furthermore, $n^{1/2-d_0}(1-w(\hat{d}))X_1 \to_p 0$ because $X_1 = O_p(1)$, $w(d_0) = 1$ if $d_0 \leq 1/2$, and $\hat{d} \to_p d_0$. Therefore, $n^{1/2-d_0}\hat{\mu}(\hat{d}) \to_d w(d_0)(\Gamma(d_0+1))^{-1}\omega W_{d_0+1}(1)$. Furthermore, applying Phillips and Shimotsu (2004, p.676, line 10) with replacing their $-\alpha+1$ with $\hat{d}-1$ gives

$$\begin{aligned}(1-L)^{\hat{d}-1} v_{[nr]} &= \frac{(2-\hat{d})_{[nr]-1}}{([nr]-1)!} = \frac{\Gamma(1-\hat{d}+[nr])}{\Gamma(2-\hat{d})\Gamma([nr])} = (\Gamma(2-\hat{d}))^{-1}[nr]^{1-\hat{d}}(1+O([nr]^{-1})) \\ &= (\Gamma(2-d_0))^{-1}(nr)^{1-d_0}(1+o_p(1)),\end{aligned} \tag{44}$$

and (42) follows.

For (39), define $b_k(\theta) = \Gamma(-\theta+k)/[\Gamma(-\theta)\Gamma(k+1)]$, then we obtain

$$\begin{aligned}\hat{u}_t &= (1-L)^{\theta} u_t \mathbf{1}\{t \geq 1\} - \hat{\mu}(\hat{d})(1-L)^{\hat{d}} v_t \\ &= \sum_{k=0}^{t-1} \frac{\Gamma(-\theta+k)}{\Gamma(-\theta)\Gamma(k+1)} u_{t-k} - \hat{\mu}(\hat{d})(1-L)^{\hat{d}} v_t \\ &= u_t + \sum_{k=1}^{t-1} b_k(\theta) u_{t-k} - \hat{\mu}(\hat{d})(1-L)^{\hat{d}} v_t.\end{aligned} \tag{45}$$

Applying summation by parts as in (43), proceeding as above, and using $-\theta\Gamma(-\theta) = \Gamma(1-\theta)$ gives $\sum_{k=1}^{t-1} b_k(\theta)u_{t-k} = \sum_{k=1}^{t-2} \theta O(k^{-\theta-2}) \sum_{q=1}^{k} u_{t-q} + O(t^{-\theta-1}) \sum_{q=1}^{t-1} u_{t-q}$. Then, it follows from Minkowski's inequality that $E\sup_{|\theta|\leq 1/\log n} |\sum_{k=1}^{t-1} b_k(\theta)u_{t-k}|^2 \leq C(\theta^2 + t^{-1})$ for a finite constant C and for all t. The third term on the right of (45) is $O_p(n^{d_0-1/2}t^{-d_0}\mathbf{1}\{d_0 < 3/4\}) + O_p(t^{-d_0}\mathbf{1}\{d_0 > 1/2\})$, because $(1-L)^{\hat{d}} v_t = (\Gamma(1-\hat{d}))^{-1} t^{-\hat{d}}(1+O(t^{-1}))$ from (44) and $\hat{\mu}(\hat{d}) = O_p(n^{d_0-1/2}\mathbf{1}\{d_0 < 3/4\}) + O_p(\mathbf{1}\{d_0 > 1/2\})$. Therefore, the sample autocovariances of $\hat{u}_t$ are equal to those of u_t up to on $o_p(1)$ term, and $s^2(l) - \tilde{s}^2(l) = o_p(1)$ and (39) follow. □

8.6 Proof of Lemma 6

The stated result follows if

$$n^{-1/2}\sum_{t=1}^{[nr]}\left[(1-L)^{\hat{d}}(\dot{X}_t-\dot{\mu}(\hat{d}))\right] \;\Rightarrow\; \omega W_2(r;d_0), \tag{46}$$

$$s^2(l) \;\rightarrow\; {}_p\omega^2. \tag{47}$$

Assume $\mu_0=\beta_0=0$ without loss of generality. First, we derive the limit of $\tilde{\beta}$. For $\tilde{\beta}$, applying summation by parts and (7) gives

$$n^{-d_0-3/2}\sum_{t=1}^{n}(t-\bar{t})X_t = n^{-d_0-3/2}\left[-\sum_{t=1}^{n-1}\sum_{k=1}^{t}X_k+(n-\bar{t})\sum_{k=1}^{n}X_k\right]\rightarrow_d (1/12)\omega\xi(d_0).$$

Since $n^{-3}\sum_{t=1}^{n}(t-\bar{t})^2\rightarrow 1/12$, it follows that $n^{-d_0+3/2}\tilde{\beta}\rightarrow\omega\xi(d_0)$.

We proceed to show (46). Substituting the definition of $\dot{X}_t$ into $\dot{\mu}(\hat{d})$ gives

$$\begin{aligned}\dot{X}_t-\dot{\mu}(d) &= X_t-\tilde{\mu}-\tilde{\beta}t-w(d)(\bar{X}-\tilde{\mu}-\tilde{\beta}\bar{t})-(1-w(d))(X_1-\tilde{\mu}-\tilde{\beta})\\ &= X_t-\tilde{\beta}t-w(d)(\bar{X}-\tilde{\beta}\bar{t})-(1-w(d))(X_1-\tilde{\beta})\\ &= X_t-\tilde{\beta}(t-1)-w(d)[\bar{X}-\tilde{\beta}(\bar{t}-1)]-(1-w(d))X_1.\end{aligned}$$

Since $\sum_{t=1}^{[nr]}(1-L)^{\hat{d}}v_t=(1-L)^{\hat{d}-1}v_{[nr]}=(\Gamma(2-d_0))^{-1}(nr)^{1-d_0}(1+o_p(1))$ from (44) and $\sum_{t=1}^{[nr]}(1-L)^{\hat{d}}(t-1)=(1-L)^{\hat{d}-2}v_{[nr]-1}=(\Gamma(3-d_0))^{-1}(nr)^{2-d_0}(1+o_p(1))$, it follows from (41) and the limit of $\tilde{\beta}$ that

$$\begin{aligned}n^{-1/2}\sum_{t=1}^{[nr]}(1-L)^{\hat{d}}X_t &\Rightarrow \omega W(r),\\ n^{-1/2}\sum_{t=1}^{[nr]}\tilde{\beta}(1-L)^{\hat{d}}(t-1) &\Rightarrow \frac{\omega}{\Gamma(3-d_0)}r^{2-d_0}\xi(d_0),\\ n^{-1/2}\sum_{t=1}^{[nr]}(1-L)^{\hat{d}}w(\hat{d})[\bar{X}-\tilde{\beta}(\bar{t}-1)] &\Rightarrow \frac{\omega w(d_0)r^{1-d_0}}{\Gamma(2-d_0)}\left[\frac{W_{d_0+1}(1)}{\Gamma(d_0+1)}-\frac{\xi(d_0)}{2}\right],\end{aligned}$$

and $n^{-1/2}\sum_{t=1}^{[nr]}(1-L)^{\hat{d}}(1-w(d))X_1=O_p(n^{1/2-d_0}\mathbf{1}\{d_0>1/2\})=o_p(1)$. Therefore, (46) follows.

For (47), we obtain the limit of $\tilde{u}_t$ similarly to $\hat{u}_t$ as (note that $(1-L)^d t=(1-L)^{d-1}v_t$)

$$\begin{aligned}\tilde{u}_t &= (1-L)^{\hat{d}}X_t-\tilde{\beta}(1-L)^{\hat{d}}(t-1)-w(\hat{d})[\bar{X}-\tilde{\beta}(\bar{t}-1)](1-L)^{\hat{d}}v_t-(1-w(d))X_1(1-L)^{\hat{d}}v_t\\ &= u_t+O_p(\theta+t^{-1/2})+O_p(n^{d_0-3/2}t^{1-d_0})\\ &\quad +O_p(n^{d_0-1/2}t^{-d_0}\mathbf{1}\{d_0<3/4\})+O_p(t^{-d_0}\mathbf{1}\{d_0>1/2\}),\end{aligned}$$

and (47) follows. □

8.7 Auxiliary results

The following technical results that are used in the proof of Lemma 2.

Proposition 1 $\sum_{t=(a-1)n/b+1}^{an/b} E(z_t b^{1/2} z_t^{(a)} | F_{t-1}) = b^{1/2} \sum_{t=(a-1)n/b+1}^{an/b-1} \sum_{s=(a-1)n/b+1}^{an/b-t} c_s \tilde{c}_s + o_p(1)$ *for* $a = 1, \ldots, b$.

Proof The left hand side is equal to

$$\left(b^{1/2} \sum_{t=(a-1)n/b+1}^{an/b} \sum_{s=(a-1)n/b+2}^{t-1} \varepsilon_s^2 c_{t-s} \tilde{c}_{t-s} - 1 \right) + b^{1/2} \sum_{t=(a-1)n/b+1}^{an/b} \sum\sum_{r \neq s} \varepsilon_s \varepsilon_r c_{t-s} \tilde{c}_{t-r}, \quad (48)$$

where $\sum\sum_{r\neq s}$ in the second term sums over $r = 1, \ldots, t-1$ and $s = (a-1)n/b + 1, \ldots, t-1$ with $r \neq s$. The second term in (48) has mean zero and variance

$$\begin{aligned} & \sum_{t=(a-1)n/b+2}^{an/b} \sum_{u=(a-1)n/b+2}^{an/b} \sum_{r=1, r\neq s}^{\min(t-1,u-1)} c_{t-r} c_{u-r} \sum_{s=(a-1)n/b+1}^{\min(t-1,u-1)} \tilde{c}_{t-s} \tilde{c}_{u-s} \\ = & \sum_{t=(a-1)n/b+2}^{an/b} \sum_{r=1}^{t-1} \sum_{s=(a-1)n/b+1}^{t-1} c_{t-r}^2 \tilde{c}_{t-s}^2 \\ & +2 \sum_{t=(a-1)n/b+3}^{an/b} \sum_{u=(a-1)n/b+2}^{t-1} \sum_{r=1}^{u-1} \sum_{s=(a-1)n/b+1}^{u-1} c_{t-r} c_{u-r} \tilde{c}_{t-s} \tilde{c}_{u-s}. \end{aligned}$$

Note that $\tilde{c}_s$ satisfies $|\tilde{c}_s| = O(n^{-1} m^{-1/2} \log m)$, $|\tilde{c}_s| = O(m^{-1/2} s^{-1} \log m)$ for $1 \leq s \leq n/2b$ and $\sum_{s=1}^{n/b} \tilde{c}_s^2 = O(n^{-1} \log^2 m)$ (Robinson, 1995). In view of the order of c_s (Robinson, 1995) and $\tilde{c}_s$, the first term is $O(n^{-1} \log^4 m)$. The second term is bounded by

$$\begin{aligned} & O\left(\sum_{r=1}^{n} c_r^2 \right) 4 \sum_{t=(a-1)n/b+3}^{an/b} \sum_{u=(a-1)n/b+2}^{t-1} \left(\sum_{s=(a-1)n/b+1}^{u-1} \tilde{c}_{u-s}^2 \right)^{1/2} \left(\sum_{s=(a-1)n/b+1}^{u-1} \tilde{c}_{t-s}^2 \right)^{1/2} \\ = & O\left(n^{-3/2} (\log m)^3 \right) 4 \sum_{t=(a-1)n/b+3}^{an/b} \sum_{u=(a-1)n/b+2}^{t-1} \left(\sum_{r=t-u+1}^{t-(a-1)n/b+1} \tilde{c}_r^2 \right)^{1/2} \\ = & O\left(n^{-3/2} (\log m)^3 \sum_{t=(a-1)n/b+3}^{an/b} \sum_{u=(a-1)n/b+2}^{t-1} \frac{\log m}{m^{1/2} (t-u)^{1/2}} \right) = O(m^{-1/2} \log^4 m), \end{aligned}$$

and hence the second term in (48) is $o_p(1)$.

The first term in (48) is

$$\left(b^{1/2}\sum_{t=(a-1)n/b+1}^{an/b-1}(\varepsilon_t^2-1)\sum_{s=(a-1)n/b+1}^{an/b-t}c_s\tilde{c}_s\right)+b^{1/2}\sum_{t=(a-1)n/b+1}^{an/b-1}\sum_{s=(a-1)n/b+1}^{an/b-t}c_s\tilde{c}_s-1.$$

The first term is $o_p(1)$ by using the bound on c_s and $\tilde{c}_s$ and applying the argument of Robinson (1995, p. 1546), and the stated result follows. □

Proposition 2 $b^{1/2}\sum_{t=(a-1)n/b+1}^{an/b-1}\sum_{s=(a-1)n/b+1}^{an/b-t}c_s\tilde{c}_s \to 1$ *for* $a=1,\ldots,b$.

Proof Set $a=1$. The proof for the other values of a follow the same argument. Observe that

$$\begin{aligned}
b^{1/2}\sum_{t=1}^{n/b-1}\sum_{s=1}^{n/b-t}c_s\tilde{c}_s &= \frac{4b^2}{mn^2}\sum_{j=1}^{m}\sum_{k=1}^{m/b}\nu_j\tilde{\nu}_k\sum_{t=1}^{n/b-1}\sum_{s=1}^{n/b-t}\cos(s\lambda_j)\cos(s\tilde{\lambda}_k)\\
&= \frac{4b^2}{mn^2}\sum_{k=1}^{m/b}\nu_{bk}\tilde{\nu}_k\sum_{t=1}^{n/b-1}\sum_{s=1}^{n/b-t}\cos(s\lambda_{bk})\cos(s\tilde{\lambda}_k) \qquad (49)\\
&\quad +\frac{4b^2}{mn^2}\sum_{k=1}^{m/b}\sum_{j\neq bk}^{m}\nu_j\tilde{\nu}_k\sum_{t=1}^{n/b-1}\sum_{s=1}^{n/b-t}\cos(s\lambda_j)\cos(s\tilde{\lambda}_k) \qquad (50)
\end{aligned}$$

We show (49) $\to b^{-1}$ and (50) $\to 1-b^{-1}$. For (49), since $\nu_{bk}=\log(bk)-m^{-1}\sum_{j=1}^{m}\log j = \log k-\log m/b+1+o(\log^{-2}m)=\tilde{\nu}_k+o(\log^{-2}m)$, $\sum_{t=1}^{n/b-1}\sum_{s=1}^{n/b-t}\cos(s\tilde{\lambda}_k)^2=(n/b-1)^2/4$ (Robinson, 1995, p.1654), and $(b/m)^{-1}\sum_{k=1}^{m/b}\tilde{\nu}_k^2\to 1$, we have

$$\begin{aligned}
(49) &= \frac{4b^2}{mn^2}\sum_{k=1}^{m/b}(\tilde{\nu}_k+o((\log m)^{-2}))\tilde{\nu}_k\sum_{t=1}^{n/b-1}\sum_{s=1}^{n/b-t}\cos(s\tilde{\lambda}_k)^2\\
&= \frac{4b^{-1}}{(m/b)(n/b)^2}\sum_{k=1}^{m/b}\tilde{\nu}_k^2\left[\frac{n^2}{4b^2}+O(n)\right]+o(1)\to\frac{1}{b}.
\end{aligned}$$

For (50), from (4.18) of Robinson (1995), for $j\neq bk$,

$$\begin{aligned}
\sum_{t=1}^{n/b-1}\sum_{s=1}^{n/b-t}\cos(s\lambda_j)\cos(s\tilde{\lambda}_k) &= \frac{1}{2}\sum_{t=1}^{n/b-1}\sum_{s=1}^{n/b-t}\left[\cos\left\{s(\lambda_j+\tilde{\lambda}_k)\right\}+\cos\left\{s(\lambda_j-\tilde{\lambda}_k)\right\}\right]\\
&= A_{jk}+B_{jk},
\end{aligned}$$

where

$$
\begin{aligned}
A_{jk} &= -\frac{n}{2} + \frac{1}{2}\frac{1-\cos(n(\lambda_j+\tilde{\lambda}_k)/b)}{4\sin^2((\lambda_j+\tilde{\lambda}_k)/2)} = -\frac{n}{2} + O(n^2(j+kb)^{-2}), \\
B_{jk} &= \frac{1}{2}\frac{1-\cos(n(\lambda_j+\tilde{\lambda}_k)/b)}{4\sin^2((\lambda_j-\tilde{\lambda}_k)/2)} = \frac{1}{2}\frac{n^2}{(2\pi)^2}\frac{1-\cos(n(\lambda_j+\tilde{\lambda}_k)/b)}{(j-bk)^2} + O\left(1\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{4b^2}{mn^2}\sum_{k=1}^{m/b}\sum_{j\neq bk}^{m}\nu_j\tilde{\nu}_k A_{jk} &= O\left(\frac{\log^2 m}{mn^2}\sum_{k=1}^{m/b}\sum_{j=1}^{m}\left(n+n^2(j+kb)^{-2}\right)\right) = o(1), \\
\frac{4b^2}{mn^2}\sum_{k=1}^{m/b}\sum_{j\neq bk}^{m}\nu_j\tilde{\nu}_k B_{jk} &= \frac{2b^2}{m(2\pi)^2}\sum_{k=1}^{m/b}\sum_{j\neq bk}^{m}\nu_j\tilde{\nu}_k\frac{1-\cos(n(\lambda_j+\tilde{\lambda}_k)/b)}{(j-bk)^2} + o(1),
\end{aligned}
$$

and it follows that

$$
\begin{aligned}
(50) &= \frac{2b^2}{m(2\pi)^2}\sum_{k=1}^{m/b}\sum_{j\neq bk}\nu_j\tilde{\nu}_k\frac{1-\cos(n(\lambda_j+\tilde{\lambda}_k)/b)}{(j-bk)^2} + o(1) \\
&= \frac{2b^2}{m(2\pi)^2}\sum_{k=1}^{m/b}\sum_{j\geq b+1, j\neq bk}^{m}\nu_j\tilde{\nu}_k\frac{1-\cos(n(\lambda_j+\tilde{\lambda}_k)/b)}{(j-bk)^2} + o(1).
\end{aligned}
$$

Observe that

$$
\begin{aligned}
&\frac{2b^2}{m(2\pi)^2}\sum_{k=1}^{m/b}\sum_{j\geq b+1, j\neq bk}^{m}\frac{1-\cos(2\pi(j-bk)/b)}{(j-bk)^2} + o(1) \\
=\ &\frac{4b^2}{m(2\pi)^2}\sum_{h=1}^{m-b}\left(\frac{m}{b}-\left[\frac{h}{b}\right]\right)\frac{1-\cos(2\pi h/b)}{h^2} + o(1), \quad h = j-bk \\
=\ &\frac{4b}{(2\pi)^2}\left(\frac{\pi}{2}\frac{2\pi}{b} - \frac{1}{4}\frac{(2\pi)^2}{b^2}\right) + o(1) \to 1-\frac{1}{b},
\end{aligned}
$$

where $[h/b]$ denotes the integer part of h/b, and the fourth line follows from $\sum_{h=1}^{\infty}\cos(hx)/h^2 = \pi^2/6 - \pi x/2 + x^2/4$ (Gradshteyn and Ryzhik, 1994, 1.443.3).

Hence, (50)$\to 1-b^{-1}$ follows if

$$
\frac{1}{m}\sum_{k=1}^{m/b}\sum_{j\geq b+1, j\neq bk}^{m}(\nu_j\tilde{\nu}_k - 1)\frac{1-\cos(2\pi(j-bk)/b)}{(j-bk)^2} \to 0. \tag{51}
$$

Define $h = j-bk$ and $a_h = h^{-2}(1-\cos(2\pi h/b))$, and note that $0 \leq a_h \leq h^{-2}$. Rewrite the left hand side of (51) as

$$\frac{1}{m}\sum_{h=1}^{m-b}\sum_{k=1}^{m/b-[h/b]} a_h\left(\nu_{bk+h}\tilde{\nu}_k-1\right)+\frac{1}{m}\sum_{h=-m+b}^{-1}\sum_{k=[-h/b]+1}^{m/b} a_h\left(\nu_{bk+h}\tilde{\nu}_k-1\right) \tag{52}$$

The first term in (52) is equal to

$$\begin{aligned}
&\frac{1}{m}\sum_{h=1}^{m-b} a_h \sum_{k=1}^{m/b-[h/b]}\left(\tilde{\nu}_k\left\{\log(k+h/b)+\log b-\log m+1+O(m^{-1}\log m)\right\}-1\right)\\
=\;&\frac{1}{m}\sum_{h=1}^{m-b} a_h \sum_{k=1}^{m/b-[h/b]}\left(\tilde{\nu}_k\log(k+h/b)-1\right)+o(1),
\end{aligned}$$

because $\sum_{k=1}^{m/b-[h/b]}\tilde{\nu}_k=\sum_{k=1}^{[h/b]-1}\tilde{\nu}_k=O(h\log m)$. Since $\tilde{\nu}_k=\log k-\log(m/b)+1+O(m^{-1}\log m)$ and $\log(k+h/b)=\log k+O(k^{-1}h)$, we have

$$\begin{aligned}
&\tilde{\nu}_k\log(k+h/b)\\
=\;&\log^2 k-\log(m/b)\log k+\log k+O(k^{-1}h\log m)+O(m^{-1}\log^2 m).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\frac{1}{m}\sum_{h=1}^{m-b} a_h \sum_{k=1}^{m/b-[h/b]}\left(\tilde{\nu}_k\log(k+h/b)-1\right)\\
=\;&\frac{1}{m}\sum_{h=1}^{m-b} a_h \sum_{k=1}^{m/b-[h/b]}\left[\log^2 k-\log(m/b)\log k+\log k-1\right]+o(1)\\
=\;&\frac{1}{m}\sum_{h=1}^{m-b} a_h \sum_{k=1}^{m/b}\left[\log^2 k-\log(m/b)\log k+\log k-1\right]+o(1)\\
=\;&\frac{1}{m}\sum_{h=1}^{m-b} a_h O\left(\log^2 m\right)+o(1)=o(1),
\end{aligned}$$

where the second line follows from $\sum_{k=1}^{n}\log^2 k = n\log^2 n-2n\log n+2n+O(\log^2 n)$ and $\sum_{k=1}^{n}\log k=n\log n-n+O(\log n)$. The second term in (52) is $o(1)$ by a similar argument, giving (51) and the stated result follows. □

References

[1] Andersen, T. G. and T. Bollerslev (1997). “Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns.” *Journal of Finance* 52: 975-1005.

[2] Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys (2001a). “The distribution of realized exchange rate volatility.” Journal of the American Statistical Association 96: 42-55.

[3] Andersen, T. G., T. Bollerslev, F. X. Diebold and H. Ebens (2001b). “The distribution of realized stock return volatility.” Journal of Financial Economics 61: 43-76.

[4] Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys (2003). “Modeling and forecasting realized volatility.” *Econometrica* 71: 579-625.

[5] Andrews, D. W. K. and Y. Sun (2004). “Adaptive local polynomial Whittle estimation of long-range dependence.” *Econometrica* 72: 569-614.

[6] Baillie, R. T., T. Bollerslev and H. O. Mikkelsen (1996). “Fractionally integrated generalized autoregressive conditional heteroskedasticity.” *Journal of Econometrics* 74: 3-30.

[7] Banerjee, A. and G. Urga (2005). “Modelling structural breaks, long memory and stock market volatility: an overview.” *Journal of Econometrics* 129: 1-34.

[8] Beran, J. and N. Terrin (1994). ”Estimation of the long-memory parameter, based on a multivariate central limit theorem.” Journal of Time Series Analysis 15: 269-278.

[9] Berkes, I., L. Horváth, P. Kokoszka and Q.-M. Shao (2006). “On discriminating between long-range dependence and changes in mean.” *Annals of Statistics* 34, 1140-1165.

[10] Bhardwaj, G. and N. R. Swanson (2006). “An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series.” Journal of Econometrics 131: 539-578.

[11] Bollerslev, T. and J. H. Wright (2000). “Semiparametric Estimation of Long-Memory Volatility Dependencies: The Role of High-Frequency Data.” *Journal of Econometrics* 98: 81-106.

[12] Breidt, F. J., N. Crato and P. d. Lima (1998). “The Detection and Estimation of Long Memory in Stochastic Volatility.” *Journal of Econometrics* 83: 325-348.

[13] Chambers, M. (1998). “Long memory and aggregation in macroeconomic time series.” *International Economic Review* 39: 1053-1072.

[14] Choi, K. and E. Zivot (2005). Long memory and structural changes in the forward discount: an empirical investigation. Forthcoming in Journal of International Money and Finance.

[15] de Jong, R. M. and J. Davidson (2000). "Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices." Econometrica 68: 407-423.

[16] Deo, R., C. Hurvich and Y. Lu (2005). "Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment." Journal of Econometrics 131: 29-58.

[17] Diebold, F. X. and A. Inoue (2001). "Long memory and regime switching." *Journal of Econometrics* 105: 131-159.

[18] Ding, Z., C. W. J. Granger and R. F. Engle (1993). "A Long Memory Property of Stock Market Returns and a New Model." *Journal of Empirical Finance* 1: 83-106.

[19] Dolado, J. J., J. Gonzalo and L. Mayoral (2005). "What is what?: a simple time-domain test of long-memory vs. structural breaks." Mimeographed, Universidad Carlos III de Madrid.

[20] Engle, R. F. and A. D. Smith (1999). "Stochastic permanent breaks." *Review of Economics and Statistics* 81: 553-574.

[21] Fox, R. and M. S. Taqqu (1986). "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series." Annals of Statistics 14: 517-532.

[22] Giraitis, L. and D. Surgailis (1990). "A Central Limit Theorem for Quadratic Forms in Strongly Dependent Linear Variables and its Application to Asymptotical Normality of Whittle's Estimate." Probability Theory and Related Fields 86: 87-104.

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

[24] Gourieroux, C. and J. Jasiak (2001). "Memory and infrequent breaks." *Economics Letters* 70: 29-41.

[25] Gourieroux, C. and C. Y. Robert (2005). "Stochastic unit root models." Forthcoming in Econometric Theory.

[26] Gradshteyn, I. S. and I. M. Ryzhik (1965). *Tables of Integrals, Series and Products.* 5th edition. New York, Academic Press.

[27] Granger, C. W. J. and N. Hyung (2004). "Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns." *Journal of Empirical Finance* 11: 399-421.

[28] Hansen, B. E. (1992). "Consistent covariance matrix estimation for dependent heterogeneous processes." Econometrica 60(4): 967-972.

[29] Hosoya, Y. (2005). “Fractional invariance principle.” Journal of Time Series Analysis 26: 463-486.

[30] Hurvich, C.M., Chen, W.W. (2000). An efficient taper for potentially overdifferenced long-memory time series. Journal of Time Series Analysis 21, 155-180.

[31] Hurvich, C. M., E. Moulines and P. Soulier (2005). “Estimating long memory in volatility.” *Econometrica* 73: 1283-1328.

[32] Hurvich, C.M. and W.W. Chen (2000). An efficient taper for potentially overdifferenced long-memory time series. Journal of Time Series Analysis 21, 155-180.

[33] Hsu, C.-C. (2005). “Long memory or structural changes: an empirical examination on inflation rates.” Economics Letters 88: 289-294.

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

[35] Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin (1992). “Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How sure are we that economic time series have a unit root?” *Journal of Econometrics* 54: 159-178.

[36] Marinucci, D. and P. M. Robinson (2000). “Weak Convergence of Multivariate Fractional Processes.” Stochastic Processes and their Applications 86: 103-120.

[37] Mayoral, L. (2006). Testing for fractional integration versus short memory with trends and structural breaks. Mimeographed, Universidad Pompeu Fabra.

[38] Christensen, B. J. and M. ¯. Nielsen (2006). "Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting.” Journal of Econometrics 133: 343-371.

[39] Ohanissian, A., J. R. Russell and R. S. Tsay (2004). True of spurious long memory? A new test. Graduate School of Business, University of Chicago.

[40] Perron, P. and Z. Qu (2006). An analytical evaluation of the log-periodogram estimate in the presence of level shifts and its implications for stock returns volatility. Mimeographed, Boston University.

[41] Phillips, P. C. B. and P. Perron (1988). “Testing for a unit root in time series regression.” *Biometrika* 75: 335-46.

[42] Phillips, P. C. B. and K. Shimotsu (2004). "Local Whittle estimation in nonstationary and unit root cases." *Annals of Statistics* 32: 656-692.

[43] Phillips, P. C. B. and K. Shimotsu (2006). "Local Whittle estimation of fractional integration and some of its variants." *Journal of Econometrics* 103: 209-233.

[44] Rao, C. R. and S. K. Mitra (1971). *Generalized Inverse of Matrices and its Applications.* New York, Wiley.

[45] Robinson, P. M. (1995). "Gaussian semiparametric estimation of long range dependence." *Annals of Statistics* 23: 1630-1661.

[46] Shimotsu, K. (2006). Exact local Whittle estimation of fractional integration with unknown mean and time trend. Queen's Economics Department working paper No. 1061. Available at http://qed.econ.queensu.ca/pub/faculty/shimotsu/.

[47] Shimotsu, K. and Phillips, P. C. B. (2005). Exact local Whittle estimation of fractional integration. *Annals of Statistics* 33**:** 1890-1933.

[48] Velasco, C., 1999, Gaussian semiparametric estimation of non-stationary time series. Journal of Time Series Analysis 20, 87-127.

Table 1. Simulated critical values

	Z_t			KPSS		
d	10%	5%	1%	10%	5%	1%
0.0	-2.750	-3.025	-3.556	0.347	0.460	0.736
0.1	-2.710	-2.989	-3.532	0.344	0.460	0.737
0.2	-2.678	-2.960	-3.500	0.342	0.453	0.731
0.3	-2.640	-2.932	-3.469	0.337	0.446	0.715
0.4	-2.600	-2.893	-3.432	0.335	0.440	0.702
0.5	-2.558	-2.850	-3.398	0.334	0.435	0.699
0.6	-2.475	-2.767	-3.336	0.321	0.419	0.661
0.7	-2.550	-2.838	-3.430	0.340	0.451	0.721
0.8	-2.568	-2.855	-3.430	0.348	0.463	0.743
0.9	-2.563	-2.849	-3.428	0.347	0.462	0.736
1.0	-2.563	-2.849	-3.424	0.347	0.460	0.737
1.1	-2.564	-2.850	-3.425	0.347	0.460	0.735
1.2	-2.565	-2.851	-3.426	0.347	0.460	0.735
1.3	-2.564	-2.852	-3.427	0.346	0.460	0.736
1.4	-2.564	-2.852	-3.425	0.346	0.460	0.736

Table 2. Simulated critical values

	Z_t (detrended)			KPSS (detrended)		
d	10%	5%	1%	10%	5%	1%
0.0	-2.989	-3.260	-3.806	0.120	0.148	0.219
0.1	-3.036	-3.308	-3.854	0.121	0.150	0.221
0.2	-3.065	-3.343	-3.879	0.126	0.157	0.232
0.3	-3.087	-3.360	-3.905	0.133	0.168	0.252
0.4	-3.090	-3.362	-3.913	0.143	0.182	0.275
0.5	-3.083	-3.359	-3.892	0.155	0.199	0.305
0.6	-2.959	-3.217	-3.734	0.186	0.238	0.371
0.7	-3.021	-3.237	-3.670	0.289	0.380	0.604
0.8	-3.181	-3.429	-3.897	0.330	0.439	0.703
0.9	-3.211	-3.489	-4.027	0.342	0.454	0.730
1.0	-3.128	-3.405	-3.954	0.347	0.460	0.737
1.1	-2.995	-3.264	-3.799	0.344	0.458	0.733
1.2	-2.870	-3.144	-3.672	0.340	0.446	0.717
1.3	-2.770	-3.052	-3.596	0.332	0.435	0.693
1.4	-2.692	-2.979	-3.523	0.325	0.424	0.669

Table 3. Simulation results with $I(d)$ processes: $n = 5,000$

m	mean($\hat{d}$)	mean($\bar{d}$)			rejection freq. (W)			rejection freq. (W_c)		
		$b=2$	$b=4$	$b=8$	$b=2$	$b=4$	$b=8$	$b=2$	$b=4$	$b=8$
				$X_t = (1-L)^{-0.4}u_t$						
200	0.401	0.401	0.399	0.394	0.082	0.137	0.310	0.060	0.070	0.097
400	0.401	0.401	0.401	0.399	0.068	0.105	0.194	0.056	0.066	0.081
600	0.399	0.400	0.400	0.399	0.062	0.083	0.149	0.053	0.057	0.075
800	0.397	0.397	0.397	0.397	0.063	0.079	0.134	0.055	0.057	0.071
				$X_t = (1-L)^{-0.2}(1-0.6L)u_t$						
200	0.225	0.225	0.224	0.222	0.079	0.133	0.301	0.060	0.069	0.095
400	0.285	0.287	0.290	0.294	0.068	0.103	0.191	0.056	0.064	0.079
600	0.354	0.357	0.361	0.367	0.070	0.093	0.164	0.059	0.069	0.081
800	0.420	0.423	0.427	0.434	0.080	0.105	0.170	0.070	0.080	0.096
				$X_t = (1-L)^{-1}u_t\mathbf{1}\{t \geq 1\}$						
200	0.999	0.997	0.993	0.987	0.073	0.129	0.272	0.053	0.067	0.076
400	1.000	1.000	0.998	0.994	0.063	0.093	0.176	0.052	0.057	0.070
600	1.004	1.003	1.002	1.000	0.061	0.088	0.149	0.054	0.062	0.071
800	1.007	1.007	1.006	1.005	0.059	0.077	0.122	0.054	0.057	0.065

Table 4. Simulation results with spurious $I(d)$ processes: $n = 5,000$

m		mean($\hat{d}$)	mean($\bar{d}$)			rejection freq. (W_c)		
			$b=2$	$b=4$	$b=8$	$b=2$	$b=4$	$b=8$
	p			$X_t \sim$ mean-plus-noise model, $\sigma_\varepsilon^2 = 2$				
200	0.002	0.424	0.359	0.268	0.167	0.456	0.748	0.796
400	0.004	0.397	0.350	0.285	0.201	0.445	0.725	0.881
600	0.006	0.379	0.338	0.283	0.211	0.482	0.725	0.895
800	0.010	0.388	0.352	0.304	0.242	0.493	0.728	0.878
	γ			$X_t \sim$ STOPBREAK model				
200	180	0.423	0.377	0.316	0.239	0.265	0.286	0.242
400	120	0.398	0.358	0.307	0.243	0.351	0.444	0.408
600	90	0.392	0.356	0.309	0.251	0.414	0.544	0.556
800	70	0.395	0.362	0.319	0.266	0.459	0.620	0.656
	p_{00}, p_{11}			$X_t \sim$ Markov-switching model				
200	0.93	0.389	0.381	0.370	0.362	0.479	0.603	0.620
400	0.86	0.414	0.406	0.395	0.384	0.596	0.774	0.839
600	0.75	0.390	0.384	0.374	0.362	0.622	0.813	0.884
800	0.66	0.414	0.408	0.401	0.391	0.670	0.861	0.933
	α			$X_t \sim$ stochastic unit root model				
200	0.96	0.394	0.374	0.344	0.304	0.540	0.800	0.886
400	0.92	0.377	0.365	0.347	0.321	0.566	0.843	0.955
600	0.90	0.395	0.386	0.372	0.351	0.581	0.867	0.977
800	0.88	0.391	0.384	0.372	0.355	0.580	0.871	0.980

Table 5. Simulation results with $I(d)$ processes: $n = 5,000$, $u_t \sim iidN(0,1)$

m	mean($\hat{d}$)	rejection freq. (Z_t)			rejection freq. ($\hat{\eta}_\mu$)		
		10%	5%	1%	10%	5%	1%
			$X_t = (1-L)^{-0.4} u_t \mathbf{1}\{t \geq 1\}$				
200	0.400	0.079	0.032	0.004	0.077	0.031	0.002
400	0.400	0.088	0.040	0.005	0.092	0.039	0.004
600	0.399	0.090	0.042	0.007	0.098	0.045	0.007
800	0.397	0.087	0.041	0.007	0.103	0.050	0.009
			$X_t = (1-L)^{-0.2} \varepsilon_t \mathbf{1}\{t \geq 1\}$, $(1-0.6L)\varepsilon_t = u_t$				
200	0.225	0.064	0.024	0.002	0.089	0.041	0.005
400	0.285	0.289	0.170	0.052	0.017	0.004	0.000
600	0.354	0.661	0.516	0.277	0.001	0.000	0.000
800	0.420	0.909	0.827	0.623	0.000	0.000	0.000
			$X_t = (1-L)^{-1} u_t \mathbf{1}\{t \geq 1\}$				
200	0.999	0.091	0.043	0.007	0.090	0.041	0.005
400	1.000	0.095	0.046	0.009	0.094	0.043	0.006
600	1.004	0.102	0.049	0.008	0.091	0.042	0.006
800	1.007	0.112	0.057	0.011	0.082	0.038	0.006

Table 6. Simulation results with spurious $I(d)$ processes: $n = 5,000$

m	mean($\hat{d}$)	rejection freq. (Z_t)			rejection freq. ($\hat{\eta}_\mu$)		
		10%	5%	1%	10%	5%	1%
		$X_t \sim$ mean-plus-noise model, $p = 0.002$					
200	0.424	0.001	0.000	0.000	0.713	0.563	0.252
400	0.324	0.000	0.000	0.000	0.973	0.948	0.872
600	0.276	0.000	0.000	0.000	0.989	0.977	0.936
800	0.247	0.000	0.000	0.000	0.993	0.985	0.953
		$X_t \sim$ STOPBREAK model, $\gamma = 180$					
200	0.423	0.000	0.000	0.000	0.739	0.581	0.234
400	0.323	0.000	0.000	0.000	0.977	0.955	0.870
600	0.275	0.000	0.000	0.000	0.992	0.981	0.941
800	0.246	0.000	0.000	0.000	0.996	0.988	0.960
		$X_t \sim$ Markov-switching model, $p_{00} = p_{11} = 0.93$					
200	0.388	0.766	0.650	0.432	0.005	0.003	0.001
400	0.609	0.993	0.988	0.969	0.003	0.002	0.000
600	0.704	0.997	0.996	0.991	0.002	0.001	0.000
800	0.748	0.997	0.997	0.995	0.002	0.001	0.000
		$X_t \sim$ stochastic unit root model, $\alpha = 0.96$					
200	0.394	0.821	0.721	0.486	0.002	0.001	0.000
400	0.543	0.997	0.994	0.979	0.000	0.000	0.000
600	0.594	0.999	0.999	0.994	0.000	0.000	0.000
800	0.612	0.999	0.999	0.995	0.000	0.000	0.000

Table 7. Simulation results with $I(d)$ processes: $n = 240$, $u_t \sim iidN(0,1)$

m	mean($\hat{d}$)	rej. freq. (W_c)		rej. freq. (Z_t)			rej. freq. ($\hat{\eta}_\mu$)		
		$b=2$	$b=4$	10%	5%	1%	10%	5%	1%
		$X_t = (1-L)^{-0.4}u_t\mathbf{1}\{t \geq 1\}$							
20	0.387	0.079	0.050	0.073	0.040	0.013	0.068	0.025	0.003
40	0.389	0.070	0.103	0.062	0.024	0.002	0.072	0.025	0.002
60	0.385	0.065	0.092	0.074	0.031	0.004	0.093	0.037	0.004
		$X_t = (1-L)^{-0.2}\varepsilon_t\mathbf{1}\{t \geq 1\}, \quad (1-0.6L)\varepsilon_t = u_t$							
20	0.307	0.077	0.050	0.019	0.005	0.001	0.074	0.029	0.003
40	0.458	0.077	0.094	0.242	0.114	0.013	0.007	0.002	0.000
60	0.563	0.080	0.085	0.536	0.357	0.106	0.001	0.000	0.000
		$X_t = (1-L)^{-1}u_t\mathbf{1}\{t \geq 1\}$							
20	0.987	0.062	0.117	0.111	0.060	0.017	0.062	0.020	0.003
40	1.001	0.064	0.077	0.102	0.049	0.008	0.069	0.025	0.001
60	1.018	0.070	0.086	0.121	0.063	0.012	0.071	0.028	0.002

Table 8. Simulation results with spurious $I(d)$ processes: $n = 240$

m		mean($\hat{d}$)	rej. freq. (W_c)		rej. freq. (Z_t)			rej. freq. ($\hat{\eta}_\mu$)		
			$b=2$	$b=4$	10%	5%	1%	10%	5%	1%
	p		$X_t \sim$ mean-plus-noise model, $\sigma_\varepsilon^2 = 2$							
20	0.03	0.388	0.141	0.054	0.104	0.069	0.036	0.141	0.065	0.007
40	0.07	0.397	0.152	0.169	0.004	0.001	0.000	0.499	0.345	0.100
60	0.12	0.403	0.161	0.178	0.001	0.000	0.000	0.687	0.548	0.287
	γ		$X_t \sim$ STOPBREAK model							
20	40	0.405	0.100	0.043	0.103	0.069	0.033	0.137	0.053	0.005
40	25	0.397	0.120	0.124	0.003	0.001	0.000	0.503	0.341	0.092
60	20	0.377	0.145	0.142	0.001	0.000	0.000	0.707	0.574	0.312
	p_{00}, p_{11}		$X_t \sim$ Markov-switching model							
20	0.85	0.364	0.107	0.026	0.053	0.023	0.005	0.063	0.040	0.017
40	0.60	0.363	0.176	0.146	0.348	0.179	0.038	0.021	0.013	0.002
60	0.50	0.470	0.263	0.280	0.696	0.522	0.205	0.011	0.006	0.000
	α		$X_t \sim$ stochastic unit root model							
20	0.96	0.386	0.253	0.101	0.119	0.062	0.020	0.059	0.029	0.004
40	0.92	0.393	0.351	0.426	0.333	0.182	0.035	0.021	0.009	0.001
60	0.90	0.396	0.405	0.537	0.431	0.266	0.067	0.015	0.006	0.001

Table 9. Estimation and test results with S&P 500 log realized standard deviation

m	$\hat{d}$	$\bar{d}$		W_c		Z_t	$\hat{\eta}_\mu$
		$b=2$	$b=4$	$b=2$	$b=4$		
200	0.525	0.528	0.560	0.849	2.012	-1.443	0.126
300	0.488	0.489	0.525	1.372	7.664	-1.171	0.199
400	0.466	0.466	0.489	0.729	3.029	-1.017	0.256
500	0.452	0.451	0.474	0.217	6.296	-0.926	0.300
600	0.429	0.428	0.448	0.214	7.848*	-0.802	0.376
700	0.409	0.409	0.421	0.423	3.790	-0.702	0.456*
800	0.391	0.389	0.402	0.441	4.837	-0.624	0.534*

Note: * indicates rejection of the null at the 5% level. $\chi^2_{0.95}(1) = 3.84, \chi^2_{0.95}(3) = 7.82$.

Table 10. Estimation and test results with S&P 500 log realized standard deviation

m	$\hat{d}$	$\bar{d}$		W_c		Z_t	$\hat{\eta}_\mu$
		$b=2$	$b=4$	$b=2$	$b=4$		
				subperiod 1: 1/2/1985 – 12/14/1988			
40	0.549	0.531	0.340	0.068	3.825	-1.496	0.101
100	0.562	0.541	0.483	3.525	8.628*	-1.499	0.096
160	0.449	0.419	0.366	2.520	8.907*	-1.029	0.258
				subperiod 2: 12/15/1988 – 11/27/1992			
40	0.410	0.443	0.255	0.757	1.890	-1.379	0.200
100	0.358	0.375	0.304	0.015	1.709	-0.997	0.300
160	0.308	0.324	0.269	0.131	3.234	-0.775	0.404
				subperiod 3: 11/30/1992 – 11/11/1996			
40	0.367	0.390	0.364	0.207	1.536	-1.873	0.133
100	0.302	0.307	0.261	0.000	4.580	-1.327	0.246
160	0.274	0.269	0.236	0.010	8.101*	-1.162	0.305
				subperiod 4: 11/12/1996 – 10/27/2000			
40	0.485	0.539	0.376	0.054	6.976	-2.572	0.058
100	0.386	0.391	0.321	0.213	12.770*	-1.631	0.128
160	0.341	0.338	0.302	0.935	9.807*	-1.347	0.176
				subperiod 5: 10/30/2000 – 10/25/2004			
40	0.656	0.611	0.561	0.013	3.804	-1.761	0.051
100	0.597	0.571	0.527	0.469	2.168	-1.089	0.142
160	0.480	0.465	0.435	0.052	1.873	-0.226	0.517*

Note: * indicates rejection of the null at the 5% level. $\chi^2_{0.95}(1) = 3.84, \chi^2_{0.95}(3) = 7.82$.

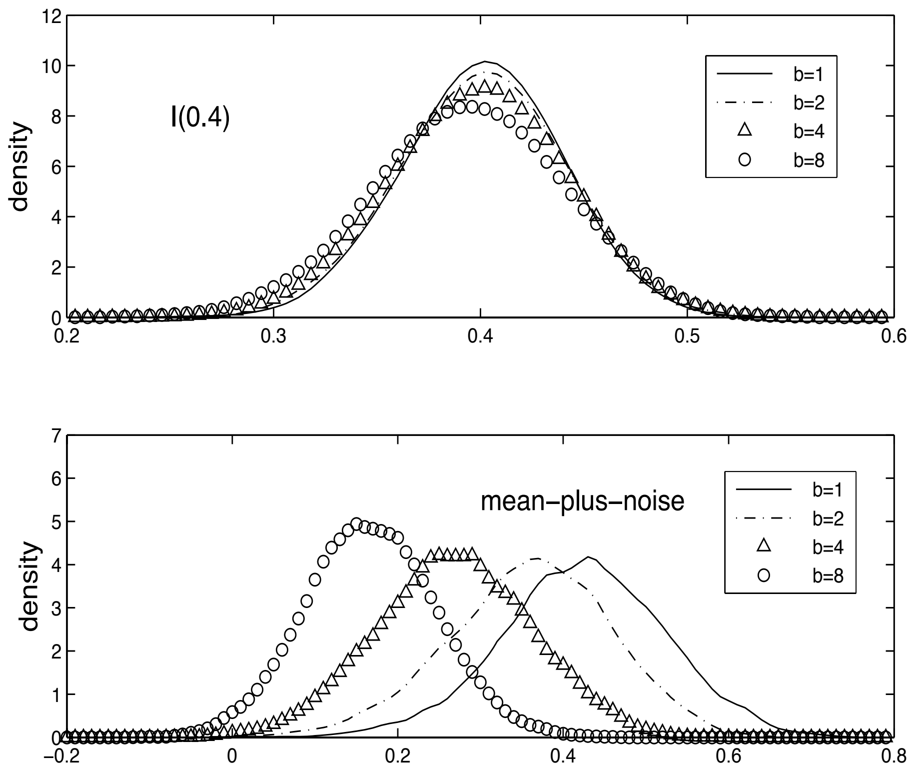

Figure 1: density of the estimators

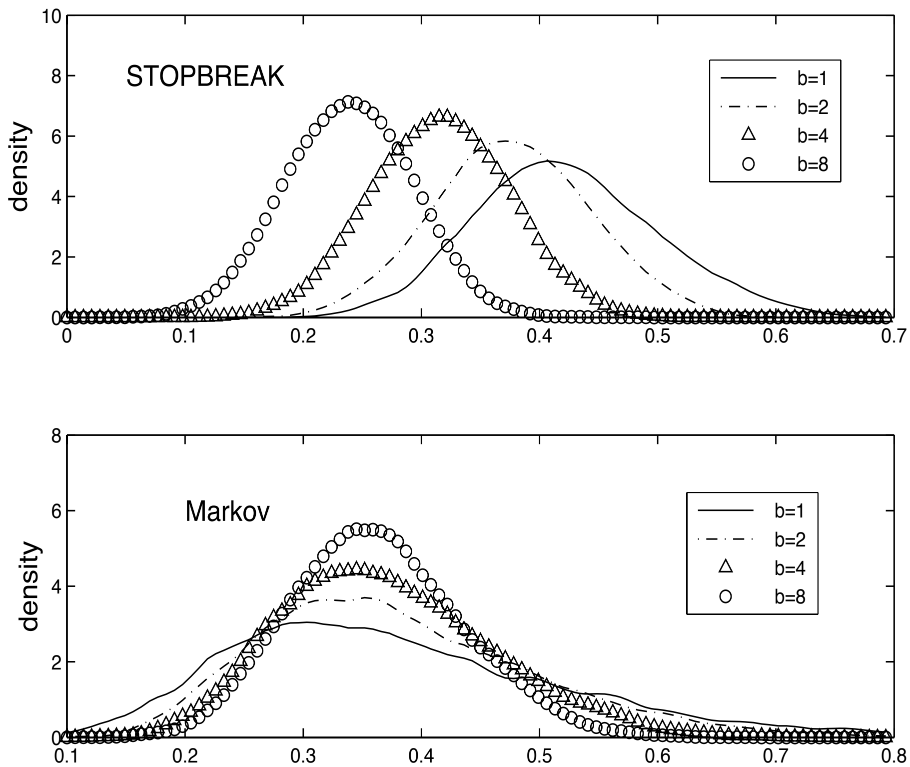

Figure 2: density of the estimators

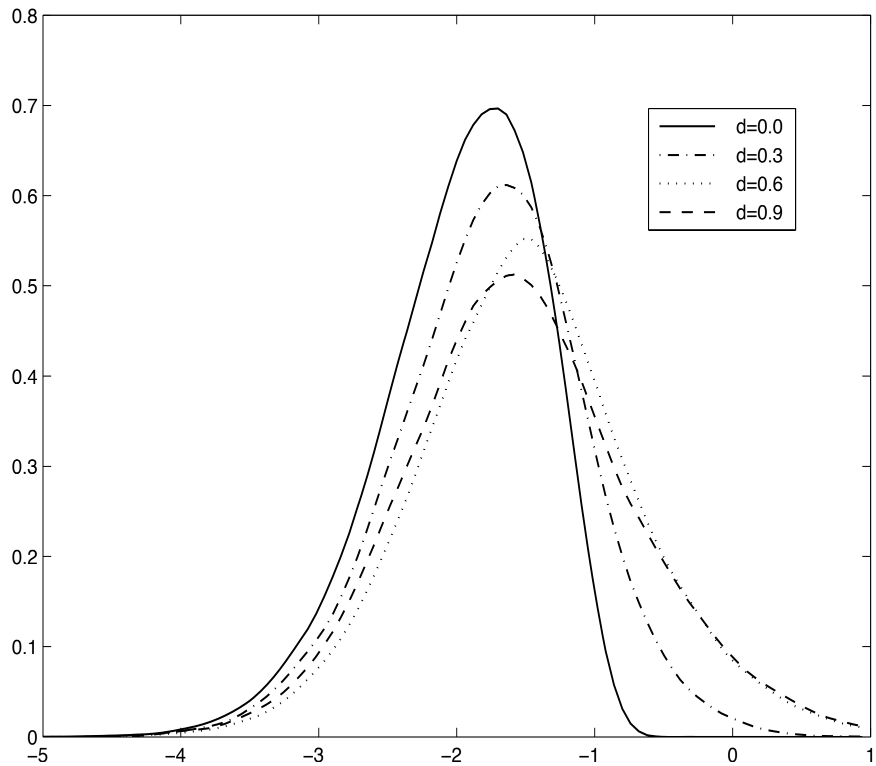

Figure 3: density of the Z_t statistic