Semiparametric Testing for Changes in Memory of Otherwise Stationary Time Series[*]

Adam McCloskey[†]

Original: March 2009; This Version: April 2009

Abstract

Many economic and financial time series are thought to exhibit long-memory behavior while nevertheless remaining covariance stationary. Changes in persistence have been widely documented though little formal analysis has been undertaken in the case of otherwise covariance stationary series. Minimal work has been done with regard to detecting change in the memory parameter d (or the Hurst parameter $H = d + 1/2$) of such series while the potential presence of such change has important implications for inference, forecasting and model building. I propose here a semiparametric test for change in d, which I dub the Range-Ratio Test (RRT). It detects changes in d when d remains in a region of stationarity $[0, 1/2)$, rather than testing against $I(0)$ or $I(1)$ alternatives. This new test's main advantage over the few existing tests for similar change in this persistence parameter is that it does not require specification of parameters affecting the spectral density at frequencies distant from zero. Asymptotic results show the RRT to be consistent with a simple null limiting distribution that is free of nuisance parameters for a wide range of null and alternative hypotheses. Monte Carlo simulations show that it performs well in moderately sized samples though care should be taken when interpreting the test statistic for initial estimates of d near the null hypothesis boundary of stationarity. The simulations also shed light on the trimming parameter that should be used for each sample size/d estimate pair. Finally, a short empirical application of the RRT is conducted providing evidence that the S&P 500 stock market volatility series exhibits rather frequent changes in memory.

JEL Classification Numbers: C12, C14, C22

Keywords: changes in persistence, hypothesis testing, long-memory processes, fractional integration, volatility, rescaled range statistics, structural change

1 Introduction

Many economic and financial time series are thought to exhibit long-memory behavior while nevertheless remaining covariance stationary. That is, although stationary, they exhibit a higher level of persistence than predicted by standard linear time series models. This phenomenon has been documented widely in volatility series and other series composed of powers of absolute returns.

In the time domain, a stationary long-memory process is characterized by a hyperbolically decaying autocorrelation function. For the covariance stationary process X_t, this can be described in the time domain as

$$\gamma_X(h) = \text{Cov}(X_t, X_{t+h}) \approx c_X(h)h^{2d-1} \quad \text{as } h \to \infty,$$

where $\approx$ denotes approximate equality, $c_X(\cdot)$ is some slowly-varying function for large values of its argument and $d \in (-1/2, 1/2)$. For $d \in (0, 1/2)$, this condition implies that the autocorrelations of X_t are not summable and, given mild conditions on $c_X(\cdot)$, the spectral density function of X_t follows

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_X(h)e^{-i\lambda h} \approx G_X|\lambda|^{-2d} \quad \text{as } \lambda \to 0, \tag{1}$$

where G_X is some strictly positive constant. Note that short-memory processes $(d = 0)$ with summable autocorrelations and finite spectral densities at zero are nested in these descriptions.

Many authors have proposed techniques to estimate the memory parameter (d) of a stationary process. These techniques fall under two broad categories: fully parametric and semiparametric. Fully parametric estimates of d require full specification of model parameters, including those that affect the spectral density at frequencies distant from zero. See Fox and Taqqu (1986) and Dahhaus (1989) for examples of parametric estimation techniques. Semiparametric techniques for estimating d have been more influential in recent years because they are robust to misspecification of parameters that affect the spectral density at frequencies distant from zero. See Geweke and Porter-Hudak (1983), Robinson (1995a) and Robinson (1995b) for examples and distributional properties of popular semiparametric estimation techniques.

In recent years, a large amount of effort has been devoted to analyzing the properties of short-memory processes with occasional breaks in mean and showing that they exhibit many of the same features as long-memory processes (e.g., see Diebold and Inoue, 2001

and Granger and Hyung, 2004). That is, such processes exhibit a hyperbolically decaying autocorrelation function and a spectral density function with a pole at the zero frequency that is still consistent with covariance stationarity (e.g., not a unit root process). In a current working paper, Perron and Qu (2008) argued that many financial time series that have been previously characterized as long-memory processes are better modeled by short-memory processes contaminated by level shifts. It has also been noted that tests for structural change in mean spuriously detect change when the underlying process has long-memory rather than short-memory with occasional breaks (e.g., Granger and Hyung, 2004). However, scant attention has been paid to the properties of long-memory processes that exhibit change *in* their memory parameter or whether economic time series exhibit such behavior.[1]

Relatedly, minimal work has been done with regards to detecting a change in d when allowing d to take on any value between zero and $1/2$ under the null hypothesis. Most tests for changes in persistence have focused on changes between trend-stationarity and difference stationarity, assuming the time series is generated by an $I(0)$ or $I(1)$ process under the null and alternative hypotheses (e.g., Kim, 2000 and Leybourne et al., 2003). To my knowledge, there are only three published works providing tests for this more general type of change in memory: Beran and Terrin (1996), Horváth and Shao (1999) and Horváth (2001).[2] However, Beran and Terrin (1996) incorrectly specified the null limiting distribution of their test statistic and were corrected by Horváth and Shao (1999) who proposed the same test statistic while obtaining the correct limiting distribution. Moreover, Horváth's (2001) test statistic is quite closely related to this same statistic. Both of the two existing published test statistics are based on fully parametric Whittle's estimates, subjecting them to the criticism of being sensitive to misspecification, discussed above. Another shortcoming of existing work on this topic is the total lack of investigation into the finite sample properties of these tests.

As shown in Section 5, certain financial time series indeed appear to exhibit changes in their memory parameters. The dearth of literature on this subject thus poses a major problem for econometric analysis. Given that many economically relevant time series exhibit long-memory behavior and/or structural change, determining whether there is a change in the memory parameter of a given series is both practically important and intuitively appeal-

---

[1]Beran and Terrin (1996) have noted that visual observation indicates that a Nile River flood level time series exhibits changes in its memory parameter.

[2]In a current working paper, Bardet and Kammoun (2008) also describe a testing procedure based on wavelet analysis. However, their approach is limited by the fact that it only applies to continuous time Gaussian processes. The majority of economic or financial series (typically volatility series) to which long-memory applies are characterized by high excess kurtosis.

ing. The potential occurrence of such change is important for inference, forecasting, model building and empirical verification of economic theory. In fact, Bollerslev and Mikkelsen (1996) have shown that the long-memory properties of stock market volatility have important implications for asset pricing.

In this paper, I propose a semiparametric test for change in the memory parameter d. The test has power against change in d when d remains in the region of stationarity $[0, 1/2)$ under both the null and alternative hypotheses rather than assuming d is $I(0)$ or $I(1)$ under the null or alternative. This new test's main advantage over the two existing tests mentioned above is that it does not require specification of parameters affecting only the short-run dynamics of a process, making it applicable in much more general settings. The test has a simple null limiting distribution free of nuisance parameters, requiring no bootstrap procedure. It can be applied to a very wide range of Gaussian or non-Gaussian processes.

The remainder of this paper is structured as follows: Section 2 is devoted to summarizing the preliminaries needed for the construction of a consistent test statistic; Section 3 develops the statistic, its null limiting distribution and its consistency property along with an interesting theorem describing the asymptotic behavior of the memory parameter estimate when the memory parameter changes; Section 4 provides an overview of the finite sample behavior of the statistic; Section 5 provides a brief empirical application of the RRT to stock market volatility data, providing evidence for changes in the memory parameter; Section 6 concludes with a discussion of possible extensions and future work related to this research; an appendix contains technical derivations including proofs of the main results; and the final pages of this paper are reserved for tables and graphs.

2 Preliminary Results

The test statistic I propose makes heavy use of a result that applies to a large class of processes satisfying (1). Let the process $\{X_t\}$ satisfy the following set of conditions:[3] $\varepsilon_i \sim i.i.d.(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon^2 < \infty$ and for all k,

$$X_k - EX_1 = \sum_{i=-\infty}^{k} c_{k-i}\varepsilon_i, \quad \text{where } c_k \approx L_d k^{d-1} \text{ as } k \to \infty, \tag{2}$$

where L_d is a constant that depends on the persistence parameter. A process satisfying (2) will exhibit long-memory as defined by (1). Almost all long-memory models used in

---

[3]There are many other sets of regularity conditions leading to the same result, see Davydov, 1970, Taqqu, 1975, Chan and Terrin, 1995 and Csőrgo and Mielniczuk, 1995 for examples.

econometric analysis satisfy this condition. Among these, the popular fractionally integrated autoregressive moving average (ARFIMA), fractionally integrated generalized autoregressive conditional heteroskedasticity (FIGARCH) and fractionally integrated stochastic volatility (FISV) processes satisfy (2) when they are stationary.

Now, define $S_{\lfloor Tr \rfloor}$ as $S_{\lfloor Tr \rfloor} = \sum_{t=1}^{\lfloor Tr \rfloor}(X_t - EX_1)$ and σ_T^2 as $\sigma_T^2 = \text{Var}(S_T)$, where T is the sample size and $0 \leq r \leq 1$. The result of Avram and Taqqu (1987) on which the test statistic crucially depends is given by the following functional central limit theorem (FCLT):

$$(1/\sigma_T)S_{\lfloor Tr \rfloor} \Rightarrow B_d(r), \tag{3}$$

where above and hereafter "$\Rightarrow$" denotes weak convergence in distribution under the Skorohod topology and $B_d(\cdot)$ is fractional Brownian motion on the unit interval with persistence parameter d. All convergence results in this paper are taken to be as $T \to \infty$ unless otherwise stated. Fractional Brownian motion on the unit interval with persistence parameter d is defined in terms of standard Brownian motion on the unit interval, denoted by $B(\cdot)$, by Samorodnitsky and Taqqu (1994)[4] as

$$B_d(r) = \frac{1}{A(d)} \int_0^r (r-s)^d \, \mathrm{d}B(s), \quad \text{where } A(d) = \left\{ \frac{1}{2d+1} + \int_0^\infty \left[(1+x)^d - x^d \right]^2 \mathrm{d}x \right\}^{\frac{1}{2}}.$$

It can be regarded as the approximate d^{th} fractional derivative of standard Brownian motion. Avram and Taqqu (1987) have also also established that $\sigma_T = O_p(T^{d+\frac{1}{2}})$ so that partial sums $S_{\lfloor Tr \rfloor}$ of the above form are uniformly $O_p(T^{d+\frac{1}{2}})$ variates.

The test statistic proposed in this paper is based upon the popular rescaled range procedure for testing the null hypothesis of short-memory ($d = 0$) versus the alternative of long-memory ($d > 0$), which examines the following quantity:

$$RS = \left[\max_{0 \leq n \leq T} S_n^* - \min_{0 \leq n \leq T} S_n^* \right], \quad \text{where } S_n^* = \sum_{t=1}^n (y_t - \bar{y})$$

and $\bar{y}$ is the sample mean of the process y_t. Statistics based upon this quantity and variants of it have enjoyed a long period of prevalence in the literature and they include the R/S-type tests of Hurst (1951), Mandelbrot and Taqqu (1979) and Lo (1991); the KPSS test introduced by Kwiatkowski et al. (1992) and the relatively new V/S test of Giraitis et al. (2003).

---

[4]Note: there are discrepancies in the literature concerning the definition of fractional Brownian motion. Marinucci and Robinson (1999) have dealt with this issue extensively. Here I make use of the proper definition for the fractional Brownian motion for this context in (3).

Before introducing the semiparamteric test statistic for change in memory, some quantities must be defined and I introduce some preliminary results to impart the intuition behind the final statistic onto the reader. Let $\varepsilon \in (0, 1/2)$, the trimming parameter of the statistic. Consider $0 < p = \lfloor \delta T \rfloor \leq T$ and $0 < n = \lfloor rT \rfloor \leq T$, where $\delta, r \in (0, 1]$ and p and n vary with the sample size T accordingly. Define the partial sum $S_n = \sum_{t=1}^{n} x_t$ for some discrete process x_t, $t = 1, \ldots, T$. Finally, let $\Lambda_{\varepsilon,T} \equiv \{n \in \mathbb{N} | \lfloor \varepsilon T \rfloor + 1 < n \leq T - \lfloor \varepsilon T \rfloor - 1\}$ and $\Lambda_{\varepsilon} \equiv \{r | \varepsilon < r \leq 1 - \varepsilon\}$, where $\mathbb{N}$ denotes the natural numbers.

Assume for now that the underlying process being tested $\{x_t\}$ satisfies (2) with $Ex_1 = 0$, where its memory parameter d may or may not change somewhere in the observed sample.[5] I will explicitly impose a further set of very mild assumptions under the null and alternative hypotheses in the next section in order to obtain the null limiting distribution and consistency of the proposed test statistic. The relevant null and alternative hypotheses for this test are given by

$H_0 : d$ is constant for all $t = 1, \ldots, T$, vs.

$H_a : d = d_1$ for $t = 1, \ldots, T_b$ and $d = d_2$ for $t = T_b + 1, \ldots, T$, for some $d_1 \neq d_2$,

where T_b is some integer in $[1, T]$. Now let

$$RS(n) \equiv \frac{\max_{p \in \mathbb{Z} \cap [n - \lfloor \varepsilon T \rfloor, n]} S_p - \min_{p \in \mathbb{Z} \cap [n - \lfloor \varepsilon T \rfloor, n]} S_p}{\max_{p \in \mathbb{Z} \cap [n+1, n + \lfloor \varepsilon T \rfloor + 1]} S_p - \min_{p \in \mathbb{Z} \cap [n+1, n + \lfloor \varepsilon T \rfloor + 1]} S_p}$$

and

$$LRS(r) \equiv \frac{\max_{\delta \in [r-\varepsilon, r]} B_d(\delta) - \min_{\delta \in [r-\varepsilon, r]} B_d(\delta)}{\max_{\delta \in [r, r+\varepsilon]} B_d(\delta) - \min_{\delta \in [r, r+\varepsilon]} B_d(\delta)},$$

where $\mathbb{Z}$ denotes the set of integers.

To gain intuition about the test statistic below, note that if $\{x_t\}$, satisfies (2) and H_0, the FCLT (3) and continuous mapping theorem (CMT) immediately give use the following result:

$$RS(n) = \frac{\max_{p \in [n - \lfloor \varepsilon T \rfloor, n]} \frac{1}{\sigma_T} S_p - \min_{p \in [n - \lfloor \varepsilon T \rfloor, n]} \frac{1}{\sigma_T} S_p}{\max_{p \in [n+1, n + \lfloor \varepsilon T \rfloor + 1]} \frac{1}{\sigma_T} S_p - \min_{p \in [n+1, n + \lfloor \varepsilon T \rfloor + 1]} \frac{1}{\sigma_T} S_p} \Rightarrow LRS(r).$$

And by another application of the CMT, we have

$$\sup_{n \in \Lambda_{\varepsilon,T}} \max\{RS(n), RS(n)^{-1}\} \Rightarrow \sup_{r \in \Lambda_{\varepsilon}} \max\{LRS(r), LRS(r)^{-1}\}.$$

---

[5]If one wishes to apply these arguments to a non-zero mean process, the true mean of the process must be estimated and the function $RS(n)$ in what follows must be slightly modified so that the partial sums S_p are replaced by S_p^*, defined above. In this case, the limiting null distribution will be the same with the fractional Brownian bridge $B_d(\delta) - \delta B_d(1)$ replacing the fractional Brownian motion $B_d(\delta)$. The test will remain consistent.

Hence, if d were known under H_0, i.e., $H_0 : d_0 = d$ for all $t = 1, \ldots, T$, we would already have a null limiting distribution. For a researcher willing to specify the value of the memory parameter under the null hypothesis, the above provides a fully *nonparametric* test. The critical values are easy to simulate for each value of d. Tests requiring the value of d under the null hypothesis, alternative hypothesis or both are common in the literature on testing for fractional integration (e.g., Mayoral, 2006). However, the case of more practical relavence and more generality is when the researcher does not specify the values of d under H_0 or H_a.

Similarly, the above test would be consistent. Suppose now that $\{x_t\}$, satisfies (2) and H_a, without loss of generality (WLOG) that $d_2 > d_1$ and $T_b/T \to \tau \in (0,1)$. That is, $\{x_t; t = 1, \ldots, T_b\}$ satisfies (2) with $d = d_1$, $\{x_t; t = T_b + 1, \ldots, T\}$ satisfies (2) with $d = d_2 > d_1$ and the time of persistence change grows at the same rate as the sample size. By the facts that S_l is uniformly $O_p(T^{d_1 + \frac{1}{2}})$ for all $l = 1, \ldots, T_b$ and S_h is uniformly $O_p(T^{d_2 + \frac{1}{2}})$ for all $h = T_b + 1, \ldots, T$, we have,

$$RS(T_b) \equiv \frac{\max_{p \in [T_b - \lfloor \varepsilon T \rfloor, T_b]} S_p - \min_{p \in [T_b - \lfloor \varepsilon T \rfloor, T_b]} S_p}{\max_{p \in [T_b + 1, T_b + \lfloor \varepsilon T \rfloor + 1]} S_p - \min_{p \in [T_b + 1, T_b + \lfloor \varepsilon T \rfloor + 1]} S_p}$$

$$= \frac{O_p(T^{d_1 + \frac{1}{2}}) - O_p(T^{d_1 + \frac{1}{2}})}{O_p(T^{d_2 + \frac{1}{2}}) - O_p(T^{d_2 + \frac{1}{2}})} = \frac{O_p(T^{d_1 + \frac{1}{2}})}{O_p(T^{d_2 + \frac{1}{2}})} = O_p(T^{d_1 - d_2}).$$

Hence, $0 \leq \max\{RS(T_b), RS(T_b)^{-1}\} = O_p\left(T^{d_2 - d_1}\right)$. If $T_b \in \Lambda_{\varepsilon, T}$,

$$\sup_{n \in \Lambda_{\varepsilon, T}} \max\{RS(n), RS(n)^{-1}\} \geq \max\{RS(T_b), RS(T_b)^{-1}\} \xrightarrow{p} \infty,$$

since $T_b/T \to \tau \in (0, 1)$.

3 The Test Statistic

If d were known under H_0, we would have a consistent *nonparametric* test statistic, namely, $\sup_{n \in \Lambda_{\varepsilon, T}} \max\{RS(n), RS(n)^{-1}\}$. However, I make no assumptions on d under H_0 and it should effectively be regarded as a nuisance parameter in this setting. In order to construct a test statistic whose null limiting distribution is free of this nuisance parameter, we must obtain a consistent estimate of it. I make use of the semiparametric local Whittle's (LW) estimate of d, originally proposed by Künsch (1987) in the construction of the statistic for a couple reasons: (i) it is robust to misspecification of model parameters that affect the spectral density away from the zero frequency, giving the test a distinct advantage over fully parametric forms; (ii) because it has been shown by Robinson (1995a) to be the most efficient of such estimates. Nevertheless, one could use a different semiparametric estimate,

such as the log-periodogram regression estimate proposed by Geweke and Porter-Hudak (1983), although Theorem 1, given below, would not directly apply.

I briefly describe the construction of the LW estimate and refer interested readers to Robinson (1995a) for details and asymptotic theory. The LW estimate is based directly on the behavior of the spectral density of a long-memory process at frequency zero, given by (1). First, define the discrete Fourier transform and periodogram of the process $\{x_t\}$ at frequency λ as

$$w(\lambda) = \frac{1}{(2\pi T)^{1/2}} \sum_{t=1}^{T} x_t e^{it\lambda} \quad \text{and} \quad I(\lambda) = |w(\lambda)|^2.$$

Since we are only concerned here with the spectral behavior of the process at the zero frequency, the estimate will only involve the computation of $I(\lambda)$ at frequencies $\lambda_j = 2\pi j/T$ for $j = 1, \ldots, m$, where m is typically small relative to the sample size T (to be made more precise later). The estimate is based on the approximate Gaussian likelihood function given by (see Künsch, 1987 for details)

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

where $I_j \equiv I(\lambda_j)$. Letting $\Theta = [\Delta_1, \Delta_2]$ with $-1/2 < \Delta_1 < \Delta_2 < 1/2$ define the compact set of admissible estimates of the true memory parameter d_0 joint minimization of $Q(G, d)$ in G and d leads to the following LW estimate of d_0:

$$\hat{d} = \operatorname{argmin}_{d \in \Theta} R(d),$$

where $R(\cdot)$ is the concentrated likelihood function given by

$$R(d) = \log \hat{G}(d) - 2d\frac{1}{m} \sum_{j=1}^{m} \log \lambda_j \quad \text{with} \quad \hat{G}(d) = \frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2d} I_j.$$

In our case, we are only concerned with the stationary yet persistent region for the memory parameter. So from here onward, regard $\Theta \equiv [0, \Delta)$, where $\Delta < 1/2$.

Before giving the expression for the test statistic, a very important function must first be defined: let

$$\widetilde{f}_d(t, u) \equiv du^{-d} \int_u^t s^{d-1}(s - u)^{-d}\mathrm{d}s - \left(\frac{t}{u}\right)^d (t - u)^{-d}. \tag{4}$$

This function is involved in transforming fractional Brownian motion on the unit interval into standard Brownian motion on the unit interval. Let the following partial sum of this

7

function be denoted as follows:

$$\widetilde{F}_d(p) = \sum_{i=2}^{p} \widetilde{f}_d\left(\frac{p}{T}, \frac{i-1}{T}\right).$$

Also define the following averages of the time series being tested as

$$\bar{x}_-(n) = \frac{1}{n-1}\sum_{i=2}^{n} x_i \quad \text{and} \quad \bar{x}_+(n) = \frac{1}{T-n}\sum_{i=n+1}^{T} x_i.$$

Define the range of a function $g(\cdot)$ over the compact set Γ as

$$\mathrm{RG}_{x\in\Gamma}\{g(x)\} = \max_{x\in\Gamma} g(x) - \min_{x\in\Gamma} g(x).$$

Now let

$$RRS(n,\hat{d}) \equiv \frac{\mathrm{RG}_{p\in\mathbb{Z}\cap[n-\lfloor\varepsilon T\rfloor,n]}\left\{\sum_{i=2}^{p} \widetilde{f}_{\hat{d}}\left(\frac{p}{T}, \frac{i-1}{T}\right) x_i - \bar{x}_-(n)\widetilde{F}_{\hat{d}}(p)\right\}}{\mathrm{RG}_{p\in\mathbb{Z}\cap[n+1,n+\lfloor\varepsilon T\rfloor+1]}\left\{\sum_{i=2}^{p} \widetilde{f}_{\hat{d}}\left(\frac{p}{T}, \frac{i-1}{T}\right) x_i - \bar{x}_+(n)\widetilde{F}_{\hat{d}}(p)\right\}}.$$

The test statistic, given below, is a simple transformation of this random function. $RRS(n,\hat{d})$ is a ratio of ranges of weighted partial sums of the underlying time series. The extra terms $\bar{x}_-(n)\widetilde{F}_{\hat{d}}(p)$ and $\bar{x}_+(n)\widetilde{F}_{\hat{d}}(p)$ are subtracted in order to correct for the effect of the mean on the weighted sums in finite samples. This may be seen more clearly in the proof of Theorem 3. These terms involve local, rather than full sample, averages of the process in order to help bring the nominal size of the test closer to its asymptotic value. The use of weighted partial sums, with weightings given by various values of $\widetilde{f}_{\hat{d}}$, will allow the test statistic to converge to a functional of standard, rather than fractional, Brownian motion. This eliminates the problem of d being a nuisance parameter in the null limiting distribution. Finally, we arrive at the test statistic for testing against a change in the memory parameter: $\sup_{n\in\Lambda_{\varepsilon,T}} \max\{RRS(n,\hat{d}), RRS(n,\hat{d})^{-1}\}$. I coin this test the "Ratio-Range Test" (RRT) for obvious reasons.

Again, I assume that the process being tested $\{x_t\}$ satisfies (2), where d does not change in the observed sample if H_0 is satisfied but changes if H_a is true. I impose one additional weak assumption on the spectral density function of the process under the null. Let $f(\cdot)$ denote the spectral density function of $\{x_t\}$.

Assumption 1. For some $\gamma > 0$, $f(\lambda)$ is differentiable for all $\lambda \in (0, \gamma)$ and

$$\frac{d}{d\lambda} \log f(\lambda) = O\left(\lambda^{-1}\right) \quad \text{as} \quad \lambda \to 0_+.$$

8

This assumption is identical to Assumption A2 of Robinson (1995a) and is used to obtain consistency of $\hat{d}$ under H_0.

An assumption on the growth of the bandwidth parameter m used in the construction of the LW estimate must also be imposed to obtain consistency of $\hat{d}$.

Assumption 2. As $T \to \infty$,

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

This is the minimal assumption on m (A4 in Robinson, 1995a) that is required for consistency since the number of periodogram ordinates used in the estimation of d must grow with the sample while the estimates of the spectral density must remain local to the zero frequency. Typical choices for m are in the range of $T^{1/2}$ to $T^{4/5}$.

The following theorem establishes the limiting distribution of the test statistic under the null hypothesis.

Theorem 1. *If $\{x_t\}$, satisfies H_0, (2) and Assumption 1 while Assumption 2 and $d_0 \in \Theta$ hold, then*

$$\sup_{n \in \Lambda_{\varepsilon,T}} \max\{RRS(n, \hat{d}), RRS(n, \hat{d})^{-1}\} \Rightarrow \sup_{r \in \Lambda_\varepsilon} \max\{LRRS(r), LRRS(r)^{-1}\},$$

where d_0 is the true value of d under H_0 and

$$LRRS(r) \equiv \frac{\max_{\delta \in [r-\varepsilon,r]} B(\delta) - \min_{\delta \in [r-\varepsilon,r]} B(\delta)}{\max_{\delta \in [r,r+\varepsilon]} B(\delta) - \min_{\delta \in [r,r+\varepsilon]} B(\delta)}.$$

Note that the null limiting distribution is a simple function of Brownian motion, free of nuisance parameters.

Now working under H_a, before establishing consistency of the RRT, I must modify Assumption 1 to suitably fit the framework of a break in the memory parameter. Now let $f_1(\cdot)$ and $f_2(\cdot)$ denote the spectral density functions of the processes $\{x_t, t = 1, \dots, T_b\}$ and $\{x_t, t = T_b + 1, \dots, T\}$, respectively.

Assumption 1*. For some $\gamma > 0$, $f_1(\lambda)$ and $f_2(\lambda)$ are differentiable for all $\lambda \in (0, \gamma)$ and for $i = 1, 2$,

$$\frac{d}{d\lambda} \log f_i(\lambda) = O(\lambda^{-1}) \quad \text{as} \quad \lambda \to 0_+.$$

I must also make an assumption that is standard in the structural break literature regarding the time of change in the memory parameter under the alternative hypothesis.

Assumption 3. As $T \to \infty$,

$$\frac{T_b}{T} \to \tau \in (0,1).$$

With these two new assumptions, I can now introduce an important result, the break case counterpart to Robinson's (1995a) consistency result. It will be crucial to establishing consistency of the RRT.

Theorem 2. *If $\{x_t\}$ satisfies H_a and Assumption 1*, where $\{x_t, t = 1, \ldots, T_b\}$ satisfies (2) with $d = d_1$ while $\{x_t, t = T_b + 1, \ldots, T\}$ satisfies (2) with $d = d_2$ and Assumptions 2-3 hold, then $\hat{d} \xrightarrow{p} d_1$ if $d_1 > d_2$ and $d_1 \in \Theta$ and $\hat{d} \xrightarrow{p} d_2$ if $d_2 > d_1$ and $d_2 \in \Theta$.*

Theorem 2 is quite an interesting result. It indicates that perhaps some of the estimates of the memory parameter provided in the literature are actually estimates of the largest value of the memory parameter for processes that change persistence in this manner. This is interesting because it would lead one to believe that many processes are exhibiting higher persistence, in terms of the memory parameter, than they actually are over the observed sample. I must note, however, that unreported simulation evidence indicates that the convergence to the higher of the two parameters is rather slow. Nonetheless, even in small samples, simulations show that the average $\hat{d}$ is closer to the larger of the two parameters.

I now present the theorem establishing consistency of the RRT.

Theorem 3. *If $\{x_t\}$ satisfies H_a and Assumption 1*, where $\{x_t, t = 1, \ldots, T_b\}$ satisfies (2) with $d = d_1$ while $\{x_t, t = T_b + 1, \ldots, T\}$ satisfies (2) with $d = d_2 > d_1$ and Assumptions 2-3 hold, then $\sup_{n \in \Lambda_{\varepsilon,T}} \max\{RRS(n, \hat{d}), RRS(n, \hat{d})^{-1}\} \xrightarrow{p} \infty$ for $T_b \in \Lambda_{\varepsilon,T}$.*

Since in practice it is unknown whether or not $d_2 > d_1$ under H_a, practical implementation takes two steps. First, one must test against H_a for which $d_2 > d_1$ by constructing the RRT statistic exactly as described previously. Second, one should reverse the order of the data in the observed sample and perform the exact same test. This second step will test against the alternative for which $d_1 > d_2$ and Theorem 3 will apply to this reversed data.

Please note that the proofs of the above two theorems can be straightforwardly extended to include alternative hypotheses entailing multiple changes in the memory parameter d. They are only omitted for brevity's sake. The only requirement for such an extension to hold is that the break dates grow at the same rate as the sample size T. Hence, $\hat{d}$ is biased toward the largest of memory parameters in the observed sample and the RRT is actually consistent against any number of changes in persistence.

I conjecture, without proof, that Theorem 2 also holds for other semiparametric estimates of the persistence parameter, such as the log-periodogram regression estimate of Geweke and Porter-Hudak (1983). This can be seen from the periodogram decomposition of Lemma A.1 in the appendix. Because Theorem 3 applies the result of Theorem 2, the RRT will likely also be consistent when these alternative estimates are used in the construction of the test statistic. One must make sure, however, to use a semiparametric estimate in order for the test to remain robust to the misspecification discussed above. The practitioner should also take note that in order for the RRT statistic to asymptotically detect a change in memory, it is necessary for the break date to lie in the set $\Lambda_{\varepsilon,T}$. Although, as we will see, a larger a trimming ε helps control the size properties of the RRT in smaller samples, too large a trimming may exclude potential changes from being detected.

4 Finite Sample Properties of the Test

Upon implementing the RRT, one will encounter the problem that the integral inside of $\widetilde{f}_d(t, u)$ has no analytical solution and the integrand exhibits a singularity at the lower bound of the integral. This singularity makes standard numerical approximation of the integral difficult. Nevertheless, the integral can be approximated up to arbitrary accuracy via the use of the following application of integration by parts:

$$\int_u^t s^{d-1}(s-u)^{-d}ds = \frac{1}{1-d}(t-u)^{1-d}t^{d-1} + \int_u^t s^{d-2}(s-u)^{1-d}ds$$
$$= \sum_{i=1}^{\infty} \frac{1}{i-d}(t-u)^{i-d}t^{d-i}.$$

Depending upon the values of t and u (which depend on the sample size), different truncations for the above infinite sum should be used to obtain accurate approximations. Large truncations are required for accuracy when the value of u is very small but one may use quite small truncations and obtain accurate approximations when u is not close to zero.

To assess the finite sample properties of the RRT, I conducted Monte Carlo simulations for three different sample sizes $T = 1,000, 2,000$ and $4,000$ and three different values of the trimming parameter $\varepsilon = 0.2$, 0.25 and 0.3. Note that the sample sizes analyzed range from moderate to large. I analyze these sizes because (i) the RRT is semiparametric in nature, requiring somewhat larger sample sizes and (ii) the vast majority of applications of the RRT will use data recorded daily or more often as data exhibiting long-memory features are typically recorded at higher frequencies. The DGPs used to assess size are ARFIMA$(0, d_0, 0)$

(or fractional white noise) processes. The DGPs used to assess power are ARFIMA$(0, d, 0)$ processes for which a change in d from d_1 to d_2 occurs at mid-sample $T_b = T/2$. The nominal size and power values are recorded from $1,000$ replications, using a 5% critical value. All size and power values are recorded for the tests against H_a for which $d_2 > d_1$ only. Although the finite sample distribution of the RRT is quite simple to simulate, I have provided critical values for various values of the trimming parameter in Table 1. These are based on $10,000$ replications.

Beginning with the relatively smaller sample size of $1,000$ in Table 2, note the large liberal size distortions occurring at higher values of d_0. On the other hand, note that for d_0 below 0.25, the size of the test is close to its nominal value. These features are indicative of a boundary issue in size. The closer the memory parameter is to the boundary of stationarity 0.5, the larger the size distortions. This problem is akin to boundary issues arising in unit root testing when the sum of the autoregressive parameters of a process is local to unity. Notice that we do not encounter this problem for d_0 closer to zero because the ARFIMA$(0, d_0, 0)$ process is stationary for all $d_0 \in (-1/2, 1/2)$. Some of these size distortions for the smaller samples can be dealt with by using a large trimming. Moreover, the construction of the test statistic requires an initial LW estimate of d. So the researcher will know how much of an issue size distortions may be prior to computing the RRT statistic. For example, if $\hat{d}$ is around 0.15, there is little need for concern but if $\hat{d}$ is around 0.45, the researcher should use a large trimming and carefully interpret results. That is, for small samples and large values of $\hat{d}$, the researcher should use a large trimming and look for large values of the RRT before drawing the conclusion that there is significant evidence for rejecting the null of no change.

As can be seen from Tables 3 and 4, the size distortions decrease monotonically as the sample size increases. For sample sizes around $4,000$, size distortions are no longer much of an issue unless the memory parameter is quite close to the boundary of nonstationarity. But even for $d_0 = 0.45$, in larger samples, size distortions become only a slight concern. I only report higher values of d_0 in Table 4 because, for all other values, the nominal sizes are very close to their exact levels. A final feature to note concerning size is that as sample sizes grow, the trimming parameter chosen has much less of an effect on size. However, as we will see in the following, larger trimmings generally lead to higher power.

Tables 5 and 6 show that the RRT performs very well in detecting large breaks in the memory parameter. Even for the smaller sample sizes, a change from 0 to 0.45 is detected over 90% of the time for most trimmings recorded. Unreported results show that these large breaks are detected nearly 100% of the time for sample sizes of $4,000$ or larger. As for

smaller breaks and the smallest sample size, Table 5 indicates that the RRT distinguishes a break in the memory parameter from a small to moderate size about as well as a break from a moderate to large size. Moreover, these smaller breaks are detected about 40% of the time and depend little on the trimming parameter used, with a higher trimming giving higher power. However, as the sample size increases, Tables 6 and 7 show that the RRT becomes somewhat better at detecting smaller moderate value to large value breaks than small value to moderate value breaks. Furthermore, the trimming parameter used begins to matter more for power as the sample size grows. This provides an interesting dichotomy with the results for size which indicate that the trimming parameter matters less as the sample size grows. Hence, we do not have the typical size-power tradeoff but something similar.

Overall, since the trimming parameter does not have much of an effect on size for large samples but has a somewhat significant effect on power, one may wish to experiment with different trimmings for larger samples. Though a larger trimming generally gives greater power, the researcher must note that Theorem 3 tells us that too large of a trimming will prevent a break from being detected. That is, the RRT only diverges when $T_b \in \Lambda_{\varepsilon,T}$.

5   Empirical Application of the Range-Ratio Test

I applied the RRT to an S&P 500 stock market volatility series composed of daily observations from October 1, 1928 to March 23, 2004, comprising a total of 20,000 observations. The daily volatility series was constructed in the standard way from daily returns. Let P_t denote the price index at date t. Then the returns are constructed as $r_t = \log(P_t) - \log(P_{t-1})$ and the volatilities are taken here to be the squared returns r_t^2. Again, I only tested against the $d_2 > d_1$ alternative. Hence, also applying the test to the time-reversed data may strengthen the evidence for the presence of memory changes in the data but will not weaken it.[6]

I obtained the LW estimate of the memory parameter as well as the value of the RRT statistic for 20 non-overlapping, size $T = 1,000$ subsamples of the data, each recorded for two different values of the bandwidth parameter, $m = T^{1/2}$ and $m = T^{4/5}$, and the trimming $\varepsilon = 0.3$. I set the trimming parameter to be large in order to weaken the potential for size distortions mentioned in the previous section. I recorded the estimated values of d to assess the relevance of potential size distortions at this small sample size. The estimates, statistics and dates corresponding to the subsamples are recorded in Table 8 with an asterisk marking

---

[6]In finite samples, the RRT statistic achieves high values when $d_1 > d_2$ even for non-reversed data. However, without reversing the data, the test is not consistent against H_a for which $d_1 > d_2$, but rather converges to a limiting distribution with larger critical values.

those values of the test statistic that reject the null at the 1% significance level. The first thing to note is the very high frequency at which significant levels of the test statistic are achieved, indicating that the memory parameter changes rather frequently. A few of these high values may be due to size distortions. However, the four largest values of the RRT statistic are achieved when size distortions should not play much of a role: from October 1, 1928 to February 15, 1932, $\hat{d}$ takes values 0.34 and 0.19; from June 29, 1935 to October 26, 1938, $\hat{d}$ takes values 0.29 and 0.21; from July 11, 1960 to June 30, 1964, $\hat{d}$ takes values 0.32 and 0.21; from May 31, 1984 to May 13, 1988, $\hat{d}$ takes values 0.03 and 0.17.

One possibly disconcerting feature of Table 8 is that the values of $\hat{d}$ often differ between the two frequencies they are recorded for. Perron and Qu (2008) and others have shown this to be a feature of a short-memory process contaminated by level shifts. Although this may provide some evidence that structural changes in levels may also be present, level shifts cannot explain the whole picture. That is, level shifts cannot be the sole reason for the RRT's significant values. For example, from June 29, 1935 to October 26, 1938, the two estimates of $\hat{d}$ do not differ significantly, yet the values of the RRT statistic are very high. The same goes for the periods January 15, 1949 to July 28, 1952 and July 29, 1952 to July 19, 1956. Change in the memory parameter appears to be a large piece of the puzzle.

Finally, I recorded the same values for two larger subsamples of 4,000 observations each. I selected one of these two subsamples to contain the oil crisis of the early 1970's (July 1, 1964 to June 16, 1980) and the other to contain the stock market crash of October 1987 (May 16, 1978 to March 19, 1994). I again used a trimming parameter of $\varepsilon = 0.3$ to increase power as I was assured the dates most likely to exhibit structural change lay within the middle 40% of the sample. The results are recorded in Table 9, with interpretations analogous to those of Table 8 but now with an asterisk if they are significant at the 5% level as size distortions are much less of an issue at this sample size. Again, we see evidence of change in the memory parameter. For the reasons previously discussed, neither size distortions nor level shifts seem to provide a solid contending argument for the apparent change in memory occurring between May 16, 1978 and March 9, 1994.

6   Extensions and Future Work

Although I have not yet conducted a thorough simulation study on the size properties of the RRT under different null DGPs, initial evidence shows that an ARFIMA process under which the autoregressive coefficients sum close to unity causes larger liberal size distortions than the pure FWN processes explored in Section 4. This problem could be due to two

things: (i) the well known fact that estimates based upon the sample periodogram, such as the semiparametric estimates of d, have poor finite sample properties in the presence of strong serial correlation; (ii) when the sum of the autoregressive coefficients is close to one, the process nears the boundary of stationarity. For reason (ii), I suspect that such size distortions will be more pronounced when d_0 is close to 0.5. It remains to be shown if this is the case and how severe a size problem this would entail. Qu (2008) has indicated that the autoregressive and moving average parameters of estimated ARFIMA models are generally small implying that these potential size distortions may not be much of an issue in practice.

If the size problem is unmanageable for the test in its current form, I would suggest a filtering procedure akin to "pre-whitening" prior to application of the RRT statistic. More specifically, let $\widetilde{x}_t = x_t - \bar{x}$, where $\bar{x}$ is just the sample mean. Estimate an ARFIMA$(p, d, 0)$ "autoregression" of the form

$$\widetilde{x}_t = \sum_{j=1}^{p} \hat{a}_j \widetilde{x}_{t-j} + \hat{v}_t$$

and apply the RRT to the residuals $\{\hat{v}_t\}$.

Just as pre-whitening procedures do not themselves assume anything about the parametric structure of the underlying DGP, the above procedure does not assume that the DGP follows an ARFIMA$(p, d, 0)$. It is only meant to remove strong serial correlation that is produced by short-term, rather than long-memory, dynamics. So the test will remain semiparametric after the above procedure is conducted and the procedure should remove a large portion of size distortions induced by short-term dynamics. Hence, an exploration of the size (and power) properties of the RRT with and without involving such a procedure is of interest if one believes that some types of economic data may both have long-memory and autoregressive coefficients that sum close to unity. For example, Breidt et al. (1998) estimate a FISV model that would exhibit such behavior if correctly specified. Nevertheless, most long-memory parametric approximations do not indicate such behavior.

Other extensions of this work include different test statistics based upon ratios of functions of partial sums. For example, one may take inspiration from the KPSS or the V/S tests mentioned in Section 2. It may also be possible to form a test statistic that exploits the frequency domain properties of a stationary process that changes persistence given in Lemma A.1 of the appendix. It would also be interesting to explore estimation extensions of the RRT statistic. The quantity

$$\text{argsup}_{n \in \Lambda_{\varepsilon,T}} \max \left\{ RRS\left(n, \hat{d}\right), RRS\left(n, \hat{d}\right)^{-1} \right\}$$

likely yields a consistent estimate of the break date in the single break case (when $T_b \in \Lambda_{\varepsilon,T}$ and $d_2 > d_1$) while multiple breaks may be estimated by maximizing the RRT statistic over various permissible grids in the time dimension. The limiting properties of such estimates would be quite useful for inference and forecasting.

Appendix 1: Technical Derivations

Proof of Theorem 1 : First note that if x_t satisfies H_0 and (2) and $d_0 \in [0, \Delta)$, then Assumptions A1 and A3 of Robinson (1995a) are satisfied. With Assumptions 1 and 2, we can immediately apply the result of Theorem 1 in Robinson (1995a): $\hat{d} \xrightarrow{p} d_0$. Second, working with the weighted partial sum $\sum_{i=2}^p \widetilde{f}_{\hat{d}}(\frac{p}{T}, \frac{i-1}{T}) x_i$, let $\mu_x = E x_1$ and observe that, by the CMT, $\hat{d} \xrightarrow{p} d_0$ implies

$$\frac{1}{T} \sum_{i=2}^{\lfloor T\delta \rfloor} \widetilde{f}_{\hat{d}}\left(\frac{\lfloor T\delta \rfloor}{T}, \frac{i-1}{T}\right) \rightarrow \int_0^\delta \widetilde{f}_{d_0}(\delta, u)\mathrm{d}u \tag{A.1}$$

which is finite since $\widetilde{f}_d(\delta, \cdot)$ is integrable over $(0, \delta)$ because it is integrable with respect to (fractional) Brownian motion, as established by Pipiras and Taqqu (2002). This integrability holds for all values of δ between zero and one. Hence, for all $\delta \in [0, 1]$,

$$\frac{1}{T\sigma_T} \sum_{i=2}^{p} \widetilde{f}_{\hat{d}}\left(\frac{p}{T}, \frac{i-1}{T}\right) x_i = \frac{1}{T} \sum_{i=2}^{\lfloor T\delta \rfloor} \widetilde{f}_{\hat{d}}\left(\frac{\lfloor \delta T \rfloor}{T}, \frac{i-1}{T}\right) \left[\frac{1}{\sigma_T} S_i - \frac{1}{\sigma_T} S_{i-1}\right]$$

$$+ \frac{\mu_x}{\sigma_T} \frac{1}{T} \sum_{i=2}^{\lfloor T\delta \rfloor} \widetilde{f}_{\hat{d}}\left(\frac{\lfloor T\delta \rfloor}{T}, \frac{i-1}{T}\right)$$

$$\Rightarrow \int_0^\delta \widetilde{f}_{d_0}(\delta, u) dB_{d_0}(u) \stackrel{d}{=} B(\delta),$$

where the weak convergence in distribution occurs by (3) under (2) and H_0, the CMT, the properties of stochastic integrals and the fact that $\hat{d} \xrightarrow{p} d_0$. The equality in distribution was established by Pipiras and Taqqu (2002) (their deconvolution formula for fractional Brownian motion defined over an interval). Third, $\bar{x}_-(n) = \frac{1}{\lfloor Tr \rfloor} S_{\lfloor Tr \rfloor} + \mu_x = O_p(1)$ for all n so that for all relevant p and n,

$$\frac{1}{T\sigma_T} \bar{x}_-(n) \widetilde{F}_{\hat{d}}(p) = \frac{1}{\sigma_T} O_p(1) = o_p(1),$$

using (A.1). The exact same result holds for the quantity $\frac{1}{T\sigma_T} \bar{x}_+(n) \widetilde{F}_{\hat{d}}(p)$. Finally, note that

$$RRS(n, \hat{d}) \equiv \frac{\mathrm{RG}_{p \in \mathbb{Z} \cap [n - \lfloor \varepsilon T \rfloor, n]} \left\{ \frac{1}{T\sigma_T} \sum_{i=2}^p \widetilde{f}_{\hat{d}}\left(\frac{p}{T}, \frac{i-1}{T}\right) x_i - \frac{1}{T\sigma_T} \bar{x}_-(n) \widetilde{F}_{\hat{d}}(p) \right\}}{\mathrm{RG}_{p \in \mathbb{Z} \cap [n+1, n + \lfloor \varepsilon T \rfloor + 1]} \left\{ \frac{1}{T\sigma_T} \sum_{i=2}^p \widetilde{f}_{\hat{d}}\left(\frac{p}{T}, \frac{i-1}{T}\right) x_i - \frac{1}{T\sigma_T} \bar{x}_+(n) \widetilde{F}_{\hat{d}}(p) \right\}}$$

so that another application of the CMT establishes the theorem's claim. ∎

Before proceeding to the proof of Theorem 2, we must first state and prove a lemma which may be of interest in its own right. But before introducing this lemma, we lay out a condition

necessary for it to hold. This condition is merely the break case analog to Assumption A1 of Robinson (1995a), imposed to establish consistency of the LW estimator.

Condition 1. As $\lambda \to 0_+$,

$$f_1(\lambda) \sim G_1 \lambda^{-2d_1} \qquad \text{and} \qquad f_2(\lambda) \sim G_2 \lambda^{-2d_2},$$

where $G_1, G_2 \in (0, \infty)$ and $0 \le d_1, d_2 < \Delta$ for some $\Delta \in (0, 1/2)$.

Lemma A.1. *Under Assumptions 1*-3 and Condition 1, for any $j = 1, \ldots, m$ as $T \to \infty$,*

$$E \frac{I(\lambda_j)}{G_1 \lambda_j^{-2d_1}} = \tau \left[1 + O\left(\frac{\log j}{j}\right)\right] \qquad \text{if } d_1 > d_2$$

and

$$E \frac{I(\lambda_j)}{G_2 \lambda_j^{-2d_2}} = (1 - \tau) \left[1 + O\left(\frac{\log j}{j}\right)\right] \qquad \text{if } d_2 > d_1.$$

Proof of Lemma A.1 : Since exactly analogous results apply to either case, we only provide the proof for $d_2 > d_1$. Let $w_t = x_t$, for $t = 1, \ldots, T_b$ and $z_t = x_{t+T_b}$, for $t = 1, \ldots, T - T_b$ and consider the following decomposition of the periodogram of x_t:

$$I(\lambda_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T_b} w_t \exp(i\lambda_j t) + \sum_{t=1}^{T-T_b} z_t \exp(i\lambda_j (t + T_b)) \right|^2$$

$$= \frac{1}{2\pi T} \left| \sum_{t=1}^{T_b} w_t \exp(i\lambda_j t) \right|^2 + \frac{1}{2\pi T} \left| \sum_{t=1}^{T-T_b} z_t \exp(i\lambda_j (t + T_b)) \right|^2$$

$$+ \frac{1}{\pi T} \left| \sum_{t=1}^{T_b} w_t \exp(i\lambda_j t) \right| \left| \sum_{t=1}^{T-T_b} z_t \exp(i\lambda_j (t + T_b)) \right|$$

$$= \frac{T_b}{T} I_{w,T_b}(\lambda_j) + \frac{1}{2\pi T} |\exp(i\lambda_j T_b)|^2 \left| \sum_{t=1}^{T-T_b} z_t \exp(i\lambda_j t) \right|^2$$

$$+ \frac{1}{\pi T} |\exp(i\lambda_j T_b)| \left| \sum_{t=1}^{T_b} w_t \exp(i\lambda_j t) \right| \left| \sum_{t=1}^{T-T_b} z_t \exp(i\lambda_j t) \right|$$

$$= \frac{T_b}{T} I_{w,T_b}(\lambda_j) + \frac{T - T_b}{T} I_{z,T-T_b}(\lambda_j)$$

$$+ \frac{2\sqrt{T_b(T - T_b)}}{T} \left| \frac{1}{\sqrt{2\pi T_b}} \sum_{t=1}^{T_b} w_t \exp(i\lambda_j t) \right| \left| \frac{1}{\sqrt{2\pi (T - T_b)}} \sum_{t=1}^{T-T_b} z_t \exp(i\lambda_j t) \right|, \quad \text{(A.2)}$$

where $I_{w,T_b}(\lambda) \equiv \frac{1}{2\pi T_b} |\sum_{t=1}^{T_b} w_t \exp(i\lambda t)|^2$ and $I_{z,T-T_b}(\lambda) \equiv \frac{1}{2\pi (T-T_b)} |\sum_{t=1}^{T-T_b} z_t \exp(i\lambda t)|^2$.

Now, by Assumptions 1* and 2, Condition 1 and Theorem 2 of Robinson (1995b),

$$E\frac{I_{w,T_b}(\lambda_j)}{G_1\lambda_j^{-2d_1}} = 1 + O\left(\frac{\log j}{j}\right) \quad \text{and} \tag{A.3}$$

$$E\frac{I_{z,T-T_b}(\lambda_j)}{G_2\lambda_j^{-2d_2}} = 1 + O\left(\frac{\log j}{j}\right) \tag{A.4}$$

uniformly in $j = 1, \ldots, m$ as $T \to \infty$. As for the final contribution to (A.2),

$$\frac{1}{\sqrt{G_1 G_2}\lambda_j^{-d_1-d_2}} E\left|\frac{1}{\sqrt{2\pi T_b}}\sum_{t=1}^{T_b} w_t \exp(i\lambda_j t)\right|\left|\frac{1}{\sqrt{2\pi(T-T_b)}}\sum_{t=1}^{T-T_b} z_t \exp(i\lambda_j t)\right|$$

$$\leq \sqrt{\frac{EI_{w,T_b}}{G_1\lambda_j^{-2d_1}}\frac{EI_{z,T-T_b}}{G_2\lambda_j^{-2d_2}}} = 1 + O\left(\frac{\log j}{j}\right), \tag{A.5}$$

uniformly in $j = 1, \ldots, m$, where the inequality follows from the Cauchy-Schwartz inequality and the equality results from (A.3) and (A.4). Along with Assumption 3, (A.2)-(A.5) imply the lemma's result since $d_2 > d_1$. ∎

We now have the necessary tools to provide the proof of Theorem 2 which relies heavily on results derived by Robinson (1995a).

Proof of Theorem 2 : Again, without loss of generality, we only give the proof for when $d_2 > d_1$. The conditions required for Lemma 1 to hold are weaker than those imposed by Theorem 2. In fact, if $\{x_t, t = 1, \ldots, T\}$ satisfies H_a where $\{x_t, t = 1, \ldots, T_b\}$ satisfies (2) with $d = d_1$ while $\{x_t, t = T_b + 1, \ldots, T\}$ satisfies (2) with $d = d_2$, Condition 1 holds. This also means Assumption A3 of Robinson (1995a) is satisfied and we can write a Wold decomposition of $\{x_t, t = 1, \ldots, T\}$ as follows:

$$x_t - Ex_1 = \sum_{j=0}^{\infty} a_j\varepsilon_{t-j} \text{ for } t = 1, \ldots, T_b \quad \text{and} \quad x_t - Ex_1 = \sum_{j=0}^{\infty} b_j\varepsilon_{t-j} \text{ for } t = T_b + 1, \ldots, T.$$

Robinson (1995a) has shown that in this context, for which $0 < d_2 < \Delta < 1/2$, for any $\alpha \in (0, 1/2)$

$$P\left(\left|\hat{d} - d_2\right| \geq \alpha\right) \leq P\left(\sup_{\Theta}|T(d)| \geq \frac{1}{2}\alpha^2\right),$$

where

$$T(d) = \log\left\{\frac{\hat{G}(d_2)}{G_2}\right\} - \log\left\{\frac{\hat{G}(d)}{G(d)}\right\} - \log\left\{\frac{2(d-d_2)+1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{2(d-d_2)}\right\}$$

$$+ 2(d-d_2)\left\{\frac{1}{m}\sum_{j=1}^{m}\log j - (\log m - 1)\right\} \quad \text{and}$$

$$G(d) = G_2\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2(d-d_2)}.$$

19

Hence, the proof reduces to showing that $\sup_\Theta T(d) \xrightarrow{p} 0$. Moreover, Robinson (1995a) has also established that both

$$\sup_\Theta \left| \log \left\{ \frac{2(d - d_2) + 1}{m} \sum_{j=1}^m \left(\frac{j}{m} \right)^{2(d-d_2)} \right\} \right|$$

and

$$\sup_\Theta \left| 2(d - d_2) \left\{ \frac{1}{m} \sum_{j=1}^m \log j - (\log m - 1) \right\} \right|$$

are $o(1)$, so that the proof is complete if

$$\sup_\Theta \left| \log \left\{ \frac{\hat{G}(d_2)}{G_2} \right\} - \log \left\{ \frac{\hat{G}(d)}{G(d)} \right\} \right| \tag{A.6}$$

is $o_p(1)$.

To this end, note that

$$\left| \frac{\hat{G}(d_2)}{G_2} + \tau 2\pi \frac{1}{m} \sum_{j=1}^m I_{\varepsilon j} - 1 \right| = \left| \frac{1}{m} \sum_{j=1}^m \left(\frac{I_j}{g_j} + \tau 2\pi I_{\varepsilon j} - 1 \right) \right|, \tag{A.7}$$

where $g_j = G_2 \lambda_j^{-2d_2}$, $I_{\varepsilon j} = |w_\varepsilon(\lambda_j)|^2$ and $w_\varepsilon(\lambda) = (2\pi T)^{-1/2} \sum_{t=1}^T \varepsilon_t e^{it\lambda}$. Now, using a variant of Robinson's (1995a) decomposition,

$$\frac{I_j}{g_j} + \tau 2\pi I_{\varepsilon j} - 1 = \left(1 - \frac{g_j}{f_j} \right) \frac{I_j}{g_j} + \frac{1}{f_j}(I_j - (1 - \tau)|\beta_j|^2 I_{\varepsilon j}) + (2\pi I_{\varepsilon j} - 1), \tag{A.8}$$

where $f_j = f_2(\lambda_j)$ and $\beta_j = \sum_{l=0}^\infty b_l e^{il\lambda_j}$. In light of Lemma 1, Robinson's (1995a) results immediately imply

$$\left| \frac{1}{m} \sum_{j=1}^m \left(1 - \frac{g_j}{f_j} \right) \frac{I_j}{g_j} \right| = o_p(1) \tag{A.9}$$

and

$$\left| \frac{1}{m} \sum_{j=1}^m (2\pi I_{\varepsilon j} - 1) \right| = o_p(1) \tag{A.10}$$

since $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$. Turning to the other relevant quantity of the decomposition, Robinson (1995a) establishes

$$E|I_j - (1 - \tau)|\beta_j|^2 I_{\varepsilon j}|$$
$$\leq \left(EI_j - \sqrt{1 - \tau}\beta_j Ew_\varepsilon(\lambda_j)\bar{w}(\lambda_j) - \sqrt{1 - \tau}\bar{\beta}_j E\bar{w}_\varepsilon(\lambda_j)w(\lambda_j) + (1 - \tau)|\beta_j|^2 EI_{\varepsilon j} \right)^{1/2}$$
$$\times \left(EI_j + \sqrt{1 - \tau}\beta_j Ew_\varepsilon(\lambda_j)\bar{w}(\lambda_j) + \sqrt{1 - \tau}\bar{\beta}_j E\bar{w}_\varepsilon(\lambda_j)w(\lambda_j) + (1 - \tau)|\beta_j|^2 EI_{\varepsilon j} \right)^{1/2}$$
$$\tag{A.11}$$

20

via use of the Cauchy-Schwartz inequality. Lemma 1 and Theorem 2 of Robinson (1995b) imply that

$$EI_j = f_j \left[(1 - \tau) + O\left(\frac{\log j}{j} \right) \right],$$

$$Ew(\lambda_j)\bar{w}_\varepsilon(\lambda_j) = \frac{\beta_j}{2\pi}\sqrt{1-\tau} + O\left(\frac{\log j}{j}\lambda_j^{-d_2} \right) \quad \text{and}$$

$$EI_{\varepsilon j} = \frac{1}{2\pi} + O\left(\frac{\log j}{j} \right)$$

uniformly in $j = 1, \ldots, m$. Hence, (A.11) is $O(f_j(\log j/j)^{1/2})$ so that for some constant C,

$$E \left| \frac{1}{m} \sum_{j=1}^m \frac{1}{f_j} \left(I_j - (1-\tau)|\beta_j|^2 I_{\varepsilon j} \right) \right| \leq \frac{C}{m} \sum_{j=1}^m \left(\frac{\log j}{j} \right)^{1/2}$$

$$\leq \frac{C}{m^{1/4}} \sum_{j=1}^m \frac{(\log j)^{1/2}}{j^{5/4}} = o(1). \qquad (A.12)$$

With (A.9) and (A.10), (A.12) implies that (A.7) is $o_p(1)$.

Applying similar techniques, let

$$\hat{H}(d) = \frac{2\pi \sum_{j=1}^m \left(\frac{j}{m} \right)^{2(d-d_2)} I_{\varepsilon j}}{\sum_{j=1}^m \left(\frac{j}{m} \right)^{2(d-d_2)}}$$

and note that

$$\frac{\hat{G}(d)}{G(d)} + \tau\hat{H}(d) - 1 = \frac{A(d)}{B(d)}, \qquad (A.13)$$

where

$$A(d) = \frac{2(d - d_2) + 1}{m} \sum_{j=1}^m \left(\frac{j}{m} \right)^{2(d-d_2)} \left(\frac{I_j}{g_j} + \tau 2\pi I_{\varepsilon j} - 1 \right) \quad \text{and}$$

$$B(d) = \frac{2(d - d_2) + 1}{m} \sum_{j=1}^m \left(\frac{j}{m} \right)^{2(d-d_2)}.$$

Robinson (1995a) has shown

$$\inf_{\Theta} B(d) \geq \frac{1}{2} \qquad (A.14)$$

and using his same arguments, the supremum of $|A(d)|$ on Θ can be shown to be bounded by

$$6 \sum_{r=1}^{m-1} (\frac{r}{m})^{-2d_2+1} \frac{1}{r^2} \left| \sum_{j=1}^r \left(\frac{I_j}{g_j} + \tau 2\pi I_{\varepsilon j} - 1 \right) \right| + \frac{3}{m} \left| \sum_{j=1}^m \left(\frac{I_j}{g_j} + \tau 2\pi I_{\varepsilon j} - 1 \right) \right|, \qquad (A.15)$$

21

so we can again apply the decomposition (A.8). Given Assumption 1 and that (A.11) is $O(f_j(\log j/j)^{1/2})$, direct appeal to Robinson's (1995a) arguments establish the first term of (A.15) is $o_p(1)$. Further, the second term of (A.15) is also $o_p(1)$ because (A.7) is.

Since (A.7) and (A.15) are both $o_p(1)$, with (A.14), we have established

$$\left| \frac{\hat{G}(d_2)}{G_2} + \tau 2\pi \frac{1}{m} \sum_{j=1}^{m} I_{\varepsilon j} - 1 \right| \xrightarrow{p} 0 \tag{A.16}$$

and

$$\sup_{\Theta} \left| \frac{\hat{G}(d)}{G(d)} + \tau \hat{H}(d) - 1 \right| \xrightarrow{p} 0. \tag{A.17}$$

Now, compare the structure of $|\hat{H}(d)-1|$ to $\frac{A(d)}{B(d)}$ in (A.13) to note that, via the decomposition (A.8), $|\hat{H}(d)-1|$ is merely one of the terms that bounds the supremum of $|A(d)|$ on Θ. Again, Robinson (1995a) has shown this term to be $o_p(1)$ so that $\hat{H}(d) \xrightarrow{p} 1$ for all $d \in \bar{\Theta}$, the closure of the set Θ. Hence, (A.17) implies

$$\frac{\hat{G}(d)}{G(d)} \xrightarrow{p} 1 - \tau \tag{A.18}$$

for all $d \in \bar{\Theta}$. Similarly, (A.10) establishes $2\pi \frac{1}{m} \sum_{j=1}^{m} I_{\varepsilon j} \xrightarrow{p} 1$ so that (A.16) implies

$$\frac{\hat{G}(d_2)}{G_2} \xrightarrow{p} 1 - \tau. \tag{A.19}$$

With the aid of Slutsky's Theorem, (A.18) and (A.19) imply

$$\left| \log \left\{ \frac{\hat{G}(d_2)}{G_2} \right\} - \log \left\{ \frac{\hat{G}(d)}{G(d)} \right\} \right| \xrightarrow{p} 0$$

for all $d \in \bar{\Theta}$ which finally yields $\sup_{\Theta} T(d) \xrightarrow{p} 0$ by (A.6), completing the proof. ∎

Proof of Theorem 3 : To establish this result, one needs to find the orders of the partial sums in $RRS(n, \hat{d})$ to find the order of $\sup_{n \in \Lambda_{\varepsilon,T}} \max\{RRS(n, \hat{d}), RRS(n, \hat{d})^{-1}\}$. In this vain, I examine the behavior of the function $\widetilde{f}_{\hat{d}}(\delta, \cdot)$. As is apparent from its functional form (4), for fixed $d \in (0, 1)$ and $\delta \in (0, 1]$, $\widetilde{f}_d\left(\delta, \frac{i-1}{T}\right)$ exhibits asymptotes at $\frac{i-1}{T} = 0$, where it tends toward ∞, and $\frac{i-1}{T} = \delta$, where it tends toward $-\infty$. Working with a term in the function $\widetilde{f}_d(\delta, \cdot)$, assume for now that $d \in (0, 1)$ is fixed. Then for all $\delta \in (0, 1]$ and $u \in (0, \delta)$, we have

$$0 \leq \int_u^\delta s^{\hat{d}-1}(s-u)^{-d}\mathrm{d}s \leq \int_u^\delta u^{d-1}(s-u)^{-d}\mathrm{d}s = u^{d-1}\int_u^\delta (s-u)^{-d}\mathrm{d}s, \tag{A.20}$$

where the second inequality follows from the assumption $d \in (0, 1)$.

22

Examining the asymptote of $\widetilde{f}_d(\delta, \cdot)$ at zero,

$$\lim_{u \to 0^+} \left\{ \widetilde{f}_d(\delta, u) u \right\} = \lim_{u \to 0^+} \left\{ \frac{du^{-d} \int_u^\delta s^{d-1}(s-u)^{-\hat{d}} \mathrm{d}s - \left(\frac{\delta}{u} \right)^d (\delta-u)^{-d}}{u^{-d} u^{d-1}} \right\}$$

$$= \lim_{u \to 0^+} \left\{ \frac{d \int_u^\delta s^{d-1}(s-u)^{-d} \mathrm{d}s - \delta^d (\delta-u)^{-d}}{u^{d-1}} \right\}$$

$$= \lim_{u \to 0^+} \left\{ \frac{d \int_u^\delta s^{d-1}(s-u)^{-d} \mathrm{d}s}{u^{d-1}} \right\}$$

$$\leq \lim_{u \to 0^+} \left\{ \frac{du^{d-1} \int_u^\delta (s-u)^{-d} \mathrm{d}s}{u^{d-1}} \right\}$$

$$= d \int_0^\delta s^{-d} \mathrm{d}s = \frac{d\delta^{1-d}}{1-d}, \tag{A.21}$$

where the third equality uses $\lim_{u \to 0^+} \left\{ \frac{\delta^d (\delta-u)^{-d}}{u^{d-1}} \right\} = 0$ and the inequality follows from (A.20).

Similarly, examining the asymptote of $\widetilde{f}_d(\delta, \cdot)$ at δ, now note that

$$u^{d-1} \int_u^\delta (s-u)^{-d} \mathrm{d}s = \frac{u^{d-1}}{1-d}(\delta-u)^{1-d}, \tag{A.22}$$

so we have

$$0 \leq \lim_{u \to \delta^-} \left\{ du^{-d} \int_u^\delta s^{d-1}(s-u)^{-d} \mathrm{d}s \right\} \leq \lim_{u \to \delta^-} \left\{ du^{-d} \left[u^{d-1} \int_u^\delta (s-u)^{-d} \mathrm{d}s \right] \right\}$$

$$= \lim_{u \to \delta^-} \left\{ \frac{du^{d-2}(\delta-u)^{1-d}}{1-d} \right\} = 0, \tag{A.23}$$

where the second inequality again follows from (A.20) and the first equality is a result of (A.22). Then,

$$\lim_{u \to \delta^-} \left\{ \widetilde{f}_d(\delta, u)(\delta-u)^d \right\} = \lim_{u \to \delta^-} \left\{ \frac{du^{-d} \int_u^\delta s^{d-1}(s-u)^{-d} \mathrm{d}s - \left(\frac{\delta}{u} \right)^d (\delta-u)^{-d}}{(\delta-u)^{-d}} \right\}$$

$$= \lim_{u \to \delta^-} \left\{ -\left(\frac{\delta}{u} \right)^d \right\} = -1. \tag{A.24}$$

where the second equality holds by (A.23). In the case that $d = 0$, $\widetilde{f}_d(\delta, u) = -1$ for all $\delta \in (0, 1]$ and $u \in (0, \delta)$, so that

$$\lim_{u \to 0^+} \left\{ \widetilde{f}_d(\delta, u) \right\} = \lim_{u \to \delta^-} \left\{ \widetilde{f}_d(\delta, u) \right\} = -1 \tag{A.25}$$

23

for all $\delta \in (0, 1]$.

Results (A.21) and (A.24) rely on the assumption that $d \in (0, 1)$, while (A.25) holds for $d = 0$. A subset of these values contain the relevant cases for the test statistic. In fact, $d \in [0, \frac{1}{2})$ are the relevant values here because, by the assumption $d_2 > d_1$ and Theorem 2, $\hat{d}$ converges in probability to $d_2 \in [0, \frac{1}{2})$ and $\hat{d}$ is restricted to be in the set $\Theta = [0, \Delta) \subset [0, 1/2]$. In the following arguments, results (A.21) and (A.24) will be applied to examining the limiting behavior of $\widetilde{f}_{\hat{d}}\left(\frac{p}{T}, \frac{i-1}{T}\right)$ as $T \longrightarrow \infty$.

Given that $\hat{d} \xrightarrow{p} d_2$ if $d_2 \in \Theta^7$, for any $\hat{d} \in \Theta$ and $\delta \in (0, 1]$,

$$
\text{plim}_{T\to\infty} \left\{ T^{-1} \max_{\frac{i-1}{T} \in \left[\frac{1}{T}, \frac{p}{T} - \frac{1}{T}\right]} \widetilde{f}_{\hat{d}}\left(\frac{p}{T}, \frac{i-1}{T}\right) \right\} = \text{plim}_{T\to\infty} \left\{ T^{-1} \widetilde{f}_{\hat{d}}\left(\delta, \frac{1}{T}\right) \right\}
$$

$$
= \lim_{T\to\infty} \left\{ T^{-1} \widetilde{f}_{d_2}\left(\delta, \frac{1}{T}\right) \right\} = \lim_{u\to 0^+} \left\{ \widetilde{f}_{d_2}(\delta, u) u \right\} \leq \frac{d_2 \delta^{1-d_2}}{1 - d_2} \tag{A.26}
$$

by the CMT and (A.21), implying that

$$
\max_{\frac{i-1}{T} \in \left[\frac{1}{T}, \frac{p}{T} - \frac{1}{T}\right]} \widetilde{f}_{\hat{d}}\left(\frac{p}{T}, \frac{i-1}{T}\right) = O_p(T).
$$

Similarly,

$$
\text{plim}_{T\to\infty} \left\{ T^{-d_2} \min_{\frac{i-1}{T} \in \left[\frac{1}{T}, \frac{p}{T} - \frac{1}{T}\right]} \widetilde{f}_{\hat{d}}\left(\frac{p}{T}, \frac{i-1}{T}\right) \right\} = \text{plim}_{T\to\infty} \left\{ T^{-d_2} \widetilde{f}_{\hat{d}}\left(\delta, \delta - \frac{1}{T}\right) \right\}
$$

$$
= \lim_{T\to\infty} \left\{ T^{-d_2} \widetilde{f}_{d_2}\left(\delta, \delta - \frac{1}{T}\right) \right\} = \lim_{u\to\delta^-} \left\{ \widetilde{f}_{d_2}(\delta, u)(\delta - u)^{d_2} \right\} = -1 \tag{A.27}
$$

by the CMT and (A.24), implying that

$$
\min_{\frac{i-1}{T} \in \left[\frac{1}{T}, \frac{p}{T} - \frac{1}{T}\right]} \widetilde{f}_{\hat{d}}\left(\frac{p}{T}, \frac{i-1}{T}\right) = O_p\left(T^{d_2}\right).
$$

Thus one can conclude

$$
\widetilde{f}_{\hat{d}}\left(\frac{p}{T}, \frac{i-1}{T}\right) = O_p(T) \tag{A.28}
$$

over the ranges of $\frac{i-1}{T}$ considered in the test statistic.

Before proceeding to establish the orders of the weighted partial sums $\sum_{i=2}^{p} \widetilde{f}_{\hat{d}}\left(\frac{p}{T}, \frac{i-1}{T}\right) x_i$, it should be noted that

$$
\sup_{d_2 \in \left[0, \frac{1}{2}\right), \delta \in (0, 1]} \left\{ \frac{d_2 \delta^{1-d_2}}{1 - d_2} \right\} = \sup_{d_2 \in \left[0, \frac{1}{2}\right)} \left\{ \frac{d_2}{1 - d_2} \right\} = 1.
$$

---

[7]The case for which $d_2 \notin \Theta$ does not pose a problem since the LW estimate will converge to Δ, the upper bound of Θ. In this case, just replace d_2 with Δ.

Hence, by (A.26), (A.27) and (A.28),

$$\frac{1}{T} \sum_{i=2}^{p} \widetilde{f}_{\hat{d}} \left(\frac{p}{T}, \frac{i-1}{T} \right) x_i = \sum_{i=2}^{p} c_i x_i, \tag{A.29}$$

where $\Pr(|c_i| \leq 1) \to 1$ as $T \to \infty$ for all $i = 2, \ldots, p$, because $\hat{d} \xrightarrow{p} d_2$.

$$\frac{\sum_{i=2}^{p} c_i x_i}{\sum_{i=2}^{p} x_i} = O_p(1). \tag{A.30}$$

Putting (A.29) and (A.30) together, we have

$$\frac{1}{T} \sum_{i=2}^{p} \widetilde{f}_{\hat{d}} \left(\frac{p}{T}, \frac{i-1}{T} \right) x_i = \frac{\sum_{i=2}^{p} c_i x_i}{\sum_{i=2}^{p} x_i} \sum_{i=2}^{p} x_i = O_p(1) \sum_{i=2}^{p} x_i$$

so that by the FCLT (3),

$$\frac{1}{T} \sum_{i=2}^{p} \widetilde{f}_{\hat{d}} \left(\frac{p}{T}, \frac{i-1}{T} \right) x_i = O_p \left(T^{d_1 + \frac{1}{2}} \right) \text{ for all } p \leq T_b \text{ and}$$

$$\frac{1}{T} \sum_{i=2}^{p} \widetilde{f}_{\hat{d}} \left(\frac{p}{T}, \frac{i-1}{T} \right) x_i = O_p \left(T^{d_2 + \frac{1}{2}} \right) \text{ for all } p > T_b \tag{A.31}$$

uniformly in p.

As shown in the proof of Theorem 1, $\frac{1}{T} \sum_{i=2}^{p} \widetilde{f}_{\hat{d}} \left(\frac{p}{T}, \frac{i-1}{T} \right)$, $\bar{x}_{-}(T_b)$ and $\bar{x}_{+}(T_b)$ are $O_p(1)$ for all relevant values of p. Hence with (A.31), we have

$$\begin{aligned} RRS \left(T_b, \hat{d} \right) &\equiv \frac{\text{RG}_{p \in \mathbb{Z} \cap [T_b - \lfloor \varepsilon T \rfloor, T_b]} \left\{ \frac{1}{T} \sum_{i=2}^{p} \widetilde{f}_{\hat{d}} \left(\frac{p}{T}, \frac{i-1}{T} \right) x_i - \frac{1}{T} \bar{x}_{-}(T_b) \widetilde{F}_{\hat{d}}(p) \right\}}{\text{RG}_{p \in \mathbb{Z} \cap [T_b + 1, T_b + \lfloor \varepsilon T \rfloor + 1]} \left\{ \frac{1}{T} \sum_{i=2}^{p} \widetilde{f}_{\hat{d}} \left(\frac{p}{T}, \frac{i-1}{T} \right) x_i - \frac{1}{T} \bar{x}_{+}(T_b) \widetilde{F}_{\hat{d}}(p) \right\}} \\ &= \frac{\text{RG}_{p \in \mathbb{Z} \cap [T_b - \lfloor \varepsilon T \rfloor, T_b]} \left\{ O_p \left(T^{d_1 + \frac{1}{2}} \right) + O_p(1) \right\}}{\text{RG}_{p \in \mathbb{Z} \cap [T_b + 1, T_b + \lfloor \varepsilon T \rfloor + 1]} \left\{ O_p \left(T^{d_2 + \frac{1}{2}} \right) + O_p(1) \right\}} \\ &= \frac{O_p \left(T^{d_1 + \frac{1}{2}} \right)}{O_p \left(T^{d_2 + \frac{1}{2}} \right)} = O_p \left(T^{d_1 - d_2} \right), \end{aligned}$$

providing us with the final result:

$$\sup_{n \in \Lambda_{\varepsilon, T}} \max \left\{ RRS \left(n, \hat{d} \right), RRS \left(n, \hat{d} \right)^{-1} \right\} \geq \max \left\{ RRS \left(T_b, \hat{d} \right), RRS \left(T_b, \hat{d} \right)^{-1} \right\} \xrightarrow{p} \infty$$

when $T_b \in \Lambda_{\varepsilon, T}$. ∎

References

Avram, F. and Taqqu, M. (1987). Noncentral limit theorems and appell polynomials. *Annals of Probability*, 15:767–775.

Bardet, J.-M. and Kammoun, I. (2008). Detecting changes in the fluctuations of a gaussian process and an application to heartbeat time series. Working Paper.

Beran, J. and Terrin, N. (1996). Testing for change of the long-memory parameter. *Biometrika*, 83:627–638.

Bollerslev, T. and Mikkelsen, H. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73:151–184.

Breidt, F., Crato, N., and de Lima, P. (1998). The detection and estimation of long memory in stochastic volatility models. *Journal of Econometrics*, 83:325–348.

Chan, N. and Terrin, N. (1995). Inference for unstable long-memory processes with applications to fractional unit autoregressions. *Annals of Statistics*, 23:1662–1683.

Csőrgo, S. and Mielniczuk, J. (1995). Distant long-range dependent sums and regression estimation. *Stochastic Processes and Applications*, 59:143–155.

Dahhaus, R. (1989). Efficient parameter estimation for self similar processes. *Annals of Statistics*, 17:1749–1766.

Davydov, Y. (1970). The invariance principle for stationary processes. *Theory of Probability and its Applications*, 15:487–498.

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

Fox, R. and Taqqu, M. (1986). Large sample properties of parameter estimates for strongly dependent stationary gaussian time series. *Annals of Statistics*, 14:517–532.

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

Giraitis, L., Kokoszka, P., Leipus, R., and Teyssière, G. (2003). Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics*, 112:265–294.

Granger, C. W. J. and Hyung, N. (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.

Horváth, L. (2001). Change point detection in long-memory processes. *Journal of Multivariate Analysis*, 78:218–234.

Horváth, L. and Shao, Q.-M. (1999). Limit theorems for quadratic forms with applications to whittle's estimate. *The Annals of Applied Probability*, 9:146–187.

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

Kim, J.-Y. (2000). Detection of change in persistence of a linear time series. *Journal of Econometrics*, 95:97–116.

Künsch (1987). Statistical aspects of self-similar processes. In Prohorov, Y. and Sazarov, V., editors, *Proceedings of the First World Congress of the Bernoulli Society*, volume 1, pages 67–74. VNU Science Press, Utrecht.

Kwiatkowski, D., Phillips, P., Schmidt, T., and Shin, Y. (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.

Leybourne, S., Kim, T.-H., Smith, V., and Newbold, P. (2003). Tests for a change in persistence against the null of difference-stationarity. *Econometrics Journal*, 6:291–311.

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

Mandelbrot, B. and Taqqu, M. (1979). Robust r/s analysis of long run serial correlation. *Bulletin of International Statistical Institute*, 48(book 2):59–104.

Marinucci, D. and Robinson, P. (1999). Alternative forms of fractional brownian motion. *Journal of Statistical Planning and Inference*, 80:111–122.

Mayoral, L. (2006). Testing for fractional integration versus short memory with trends and structural breaks. Universidad Pompeu Fabra Working Paper.

Perron, P. and Qu, Z. (2008). Long-memory and level shifts in the volatility of stock market return indices. Working Paper.

Pipiras, V. and Taqqu, M. (2002). Deconvolution of fractional brownian motion. *Journal of Time Series Analysis*, 23:487–501.

Qu, Z. (2008). A test against spurious long memory. Boston University Working Paper.

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

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

Samorodnitsky, G. and Taqqu, M. (1994). *Stable non-Gaussian random processes.* Chapman and Hall, New York.

Taqqu, M. (1975). Weak convergence to fractional brownian motion and to the rosenblatt process. *Z. Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31:287–302.

Table 1. Critical Values of the RRT Statistic

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$\varepsilon = 0.05$	4.6001	3.9264	3.6088
$\varepsilon = 0.1$	4.2094	3.5164	3.2198
$\varepsilon = 0.15$	3.9473	3.2714	2.9869
$\varepsilon = 0.2$	3.7435	3.0866	2.7896
$\varepsilon = 0.25$	3.5228	2.9073	2.6217
$\varepsilon = 0.3$	3.4299	2.7928	2.4836

Table 2. Size of the RRT when T = 1,000 (nominal size is 5%)

	$d_0 = 0$	$d_0 = 0.15$	$d_0 = 0.25$	$d_0 = 0.35$	$d_0 = 0.45$
$\varepsilon = 0.2$	6.3%	9.4%	10.9%	13.3%	20.6%
$\varepsilon = 0.25$	6.1%	8.0%	9.4%	13.4%	17.7%
$\varepsilon = 0.3$	4.6%	5.5%	7.1%	10.4%	14.5%

Table 3. Size of the RRT when T = 2,000 (nominal size is 5%)

	$d_0 = 0$	$d_0 = 0.25$	$d_0 = 0.35$	$d_0 = 0.45$
$\varepsilon = 0.2$	5.4%	8.2%	10.2%	14.0%
$\varepsilon = 0.25$	5.5%	4.2%	9.1%	14.1%
$\varepsilon = 0.3$	3.7%	4.7%	7.0%	15.0%

Table 4. Size of the RRT when T = 4,000 (nominal size is 5%)

	$d_0 = 0.35$	$d_0 = 0.45$
$\varepsilon = 0.2$	7.5%	11.1%
$\varepsilon = 0.25$	7.5%	11.8%
$\varepsilon = 0.3$	4.8%	9.7%

Table 5. Power of the RRT when T $= 1,000$ (nominal size is 5%)

$(\mathbf{d_1}, \mathbf{d_2}) =$	$(\mathbf{0}, \mathbf{0.25})$	$(\mathbf{0.25}, \mathbf{0.45})$	$(\mathbf{0}, \mathbf{0.45})$
$\varepsilon = \mathbf{0.2}$	36.2%	37.7%	84.6%
$\varepsilon = \mathbf{0.25}$	41.9%	40.8%	90.1%
$\varepsilon = \mathbf{0.3}$	44.4%	43.2%	93.5%

Table 6. Power of the RRT when T $= 2,000$ (nominal size is 5%)

$(\mathbf{d_1}, \mathbf{d_2}) =$	$(\mathbf{0}, \mathbf{0.25})$	$(\mathbf{0.25}, \mathbf{0.45})$	$(\mathbf{0}, \mathbf{0.45})$
$\varepsilon = \mathbf{0.2}$	33.1%	38.0%	93.4%
$\varepsilon = \mathbf{0.25}$	37.2%	47.1%	96.5%
$\varepsilon = \mathbf{0.3}$	44.3%	49.8%	98.2%

Table 7. Power of the RRT when T $= 4,000$ (nominal size is 5%)

$(\mathbf{d_1}, \mathbf{d_2}) =$	$(\mathbf{0}, \mathbf{0.25})$	$(\mathbf{0.25}, \mathbf{0.45})$
$\varepsilon = \mathbf{0.2}$	41.9%	50.2%
$\varepsilon = \mathbf{0.25}$	51.7%	53.5%
$\varepsilon = \mathbf{0.3}$	56.5%	59.4%

Table 8. Values of $\hat{d}$ and the RRT Statistic for S&P 500 Volatility for T $= 1,000$

Dates	RRT(T$^{1/2}$)	$\hat{d}$(T$^{1/2}$)	RRT(T$^{4/5}$)	$\hat{d}$(T$^{4/5}$)
10/1/1928-2/15/1932	12.8200*	0.34	9.7554*	0.19
2/16/1932-6/28/1935	6.5935*	0.03	3.5549*	0.11
6/29/1935-10/26/1938	9.0647*	0.29	8.9498*	0.21
10/27/1938-2/24/1942	2.8416	0.21	3.1623	0.17
2/25/1942-6/23/1945	2.8489	0.08	2.7992	0
6/25/1945-1/14/1949	3.0083	0.14	2.9412	0.22
1/15/1949-7/28/1952	4.0444*	0.22	4.2651*	0.18
7/29/1952-7/19/1956	5.8307*	0.00	5.7454*	0.04
7/20/1956-7/8/1960	4.5257*	0.49	4.1343*	0.18
7/11/1960-6/30/1964	7.9113*	0.32	8.0254*	0.21
7/1/1964-6/24/1968	2.8664	0.28	2.8886	0.27
6/25/1968-7/13/1972	3.5897*	0.35	3.6939*	0.27
7/14/1972-6/30/1976	3.7664*	0.49	4.2533*	0.15
7/1/1976-6/16/1980	4.5176*	0.41	3.1935	0.08
6/17/1980-5/30/1984	3.4458*	0.31	4.2587*	0.13
5/31/1984-5/13/1988	18.191*	0.03	19.676*	0.17
5/16/1988-4/28/1992	3.0299	0.09	3.7956*	0.04
4/29/1992-4/11/1996	1.6123	0.28	1.8207	0.05
4/12/1996-3/28/2000	3.5304*	0.24	3.5738*	0.12
3/29/00-3/23/04	2.4241	0.36	2.6462	0.27

Table 9. Values of $\hat{d}$ and the RRT Statistic for S&P 500 Volatility for T $= 4,000$

Dates	RRT(T$^{1/2}$)	$\hat{d}$(T$^{1/2}$)	RRT(T$^{4/5}$)	$\hat{d}$(T$^{4/5}$)
7/1/1964-6/16/1980	3.4297*	0.46	4.2660*	0.26
5/16/1978-3/9/1994	5.7243*	0.13	6.3188*	0.19