2022

Robust testing of time trend and mean with unknown integration order errors

S.Y. Chang, P. Perron, J. Xu. 2022. "Robust testing of time trend and mean with unknown integration order errors" Journal of Statistical Computation and Simulation, Volume 92, Issue 17, pp.3561-3582. https://doi.org/10.1080/00949655.2022.2074420

https://hdl.handle.net/2144/48981

Robust Testing of Time Trend and Mean with Unknown Integration Order Errors*

Seong Yeon Chang† **Pierre Perron**‡ **Jiawen Xu**§

Soongsil University **Boston University** **USST**

January 28, 2022

Abstract

We provide tests to perform inference on the coefficient of a linear trend assuming the noise to be a fractionally integrated process with memory parameter $d \in (-0.5, 1.5)$ excluding the boundary case 0.5 by applying a quasi-generalized least squares procedure using d-differences of the data. Doing so, the asymptotic distribution of the ordinary least squares estimators applied to quasi-differenced data and their t-statistics are unaffected by the value of d and have a normal limiting distribution. To have feasible tests, we use the exact local whittle estimator, valid for processes with a linear trend. The small sample properties of the tests are investigated via simulations and we provide comparisons with existing tests valid for a short-memory stationary, $I(0)$, or an autoregressive unit root, $I(1)$, noise. The results are encouraging in that our test is valid under more general conditions, yet has power approaching to the Perron and Yabu [Estimating deterministic trends with an integrated or stationary noise component. J. of Econometrics. 2009;151;56-69] tests that apply to the dichotomous cases with d either being 0 or 1. We also use our method of proof to show that the main result of Iacone, Leybourne and Taylor [Testing for a break in trend when the order of integration is unknown. J. of Econometrics. 2013;176:30-45] dealing with a test for a break in the slope of a trend function with a fractionally integrated noise is valid for $d \in (-0.5, 0.5) \cup (0.5, 1.5)$.

JEL Classification: C12, C14

Keywords: confidence intervals; fractional integration; inference; linear time trend; long memory; quasi-GLS procedure

---

*We are grateful to Zhongjun Qu for useful comments. Jiawen Xu acknowledges the support from the National Natural Science Foundation of China (Grant No. 72003117).

†Department of Economics, Soongsil University, 369 Sangdo Rd., Seoul, Korea 06987 (sychang@ssu.ac.kr).

‡Corresponding author: Department of Economics, Boston University, 270 Bay State Rd., Boston MA 02215 (perron@bu.edu).

§Business School, University of Shanghai for Science and Technology, 334 Jungong Rd., Shanghai, 200093 China (xu_jiawen@usst.edu.cn).

1 Introduction

Many time series are well captured by a deterministic linear trend. With a logarithmic transformation, the slope of the trend function is the average growth rate, a quantity of interest. To be more precise, consider the following model for the time series process y_t:

$$y_t = \beta_1 + \beta_2 t + u_t, \quad (t = 1, \dots, T) \tag{1}$$

where u_t are the deviations from the trend. The parameter β_2 is of primary interest. If $\beta_2 = 0$, then tests about β_1 pertain to the mean of the time series. Hypothesis testing on the slope of the trend function is important for many reasons. First, assessing whether a trend is present is of direct interest in many applications. Second, the correct specification of the trend function is important in other testing problems such as assessing the nature of the noise component u_t (see, e.g., Perron, 1988). Third, tests for hypotheses about the values of β_1 and β_2 allow for the construction of confidence intervals via inversions. There is extensive literature on issues pertaining to inference about the slope of a linear trend function, most related to the case in which the noise component is stationary, i.e., integrated of order zero, $I(0)$. The classic result of Grenander and Rosenblatt (1957) states that the estimate of β_2 obtained from the ordinary least squares (OLS) regression applied to (1) is asymptotically as efficient as the generalized least squares (GLS) estimate when the process for u_t is correctly specified. However, when u_t has an autoregressive unit root, i.e., integrated of order one, $I(1)$, the estimate of the mean of the first-differenced series is efficient in large samples.

Several papers have tackled the issue of constructing tests and confidence intervals for the parameter β_2 when it is not known a priori if u_t is $I(1)$ or $I(0)$. Sun and Pantula (1999) proposed a pre-test method which first applies a unit root test and then chooses the critical value to be used according to the outcome of the test. Since the probability of using the critical values from the $I(0)$ case does not converge to zero when the errors are $I(1)$, the simulations reported accordingly show that substantial size distortions remain. Canjels and Watson (1997) considered various feasible GLS methods. Their analysis is, however, restricted to cases in which u_t is either $I(1)$ or the autoregressive root is local to one. They do not allow $I(0)$ processes and, moreover, their method yields confidence intervals that are substantially conservative with common sample sizes. Roy et al. (2004) considered a test based on a one-step Gauss-Newton regression, but its limit distribution is not the same in the $I(1)$ and $I(0)$ cases (see Perron and Yabu, 2012). Vogelsang (1998), Bunzel and Vogelsang (2005), and Harvey et al. (2007) proposed tests valid with either $I(1)$ or $I(0)$ errors. Their approach, however, uses randomly scaled versions of tests for trends, so that in finite samples the good properties of such tests are lost, at least to some extent. Perron and Yabu (2009) considered a feasible quasi-GLS approach that uses a super-efficient estimate of the sum of the autoregressive parameters α when $\alpha = 1$. The estimate of α is based on the OLS obtained from an autoregression applied to detrended data and is truncated to take a value 1 when the estimate is in a $T^{-\delta}$ neighborhood of 1. This makes the estimate "super-efficient" when $\alpha = 1$ and implies that inference on the slope parameter can be performed using the standard normal distribution whether $\alpha = 1$ or $|\alpha| < 1$.

Much of the literature focused on u_t being $I(0)$ or $I(1)$, special cases of fractionally integrated, $I(d)$, processes with memory parameter d. Since d can take any real value (within some interval), a fractional process extends the classical dichotomy of $I(0)$ and $I(1)$ processes. Our aim is to provide tests to perform inference on the coefficient of a linear trend function, assuming the noise component to be an $I(d)$ process with $d \in (-0.5, 1.5)$ excluding the boundary case 0.5. We apply a quasi-GLS procedure using d-differences of the data. The error term is then short memory, and the asymptotic distribution of the OLS estimators of (β_1, β_2) and their t-statistics are unaffected by the value of d and standard methods can be applied with the normal limiting distribution. No truncation or pre-test is needed, given the continuity with respect to d. To make our procedure feasible, we need an estimator of d, valid with a fitted linear time trend and for a wide range of d. After experimenting with various possible estimators, we opted to use the exact local whittle (ELW) estimator of Shimotsu (2010), who extended Shimotsu and Phillips (2005) to cover processes with a linear trend. It is valid for values of d in the range of $(-0.5, 1.5)$ and yields tests with good finite sample properties. Of related interest, Abadir et al. (2011) considered an $I(d)$ model with trends and cycles and derived the asymptotic distribution of the OLS estimate of the parameter of the slope of the trend. A related paper is Iacone et al. (2013), who proposed a test for a break in the slope of a linear time trend when the order of integration is unknown, whose methodology is similar to ours. We use our method of proof to show that their result is valid for $d \in (-0.5, 0.5) \cup (0.5, 1.5)$.

This paper is organized as follows. Section 2 describes the model and the test statistics, and Section 3 discusses the estimate of d used to have feasible tests. Section 4 presents simulation results about the size and power of the tests in finite samples and a comparison with the tests of Perron and Yabu (2009) (henceforth, PY), which is valid when u_t is either $I(0)$ or $I(1)$. The results are encouraging in that our test, valid under much more general conditions, has slightly lower power than the PY tests when d is either 0 or 1. Section 5 illustrates their usefulness via applications to the U.S. equity indices. Section 6 considers generalizing the main result of Iacone et al. (2013). Section 7 provides brief conclusions, and technical derivations are collected in an appendix.

2 Model and Test Statistics

The data-generating process (DGP) is assumed to be, for $t = 1, ..., T$:

$$y_t = \beta_1 + \beta_2 t + u_t \tag{2}$$

with u_t being a fractionally integrated process satisfying the following assumptions.

Assumption 1. *The process u_t is generated by $\Delta^d u_t = (1-L)^d u_t = \varepsilon_t \mathbf{1}\{t \geq 1\}$ with $d \in (-0.5, 0.5) \cup (0.5, 1.5)$, where Δ^d is the fractional difference operator, and $\mathbf{1}\{A\}$ is the indicator function of the event A.*[1] *Also, ε_t is a short memory process generated by $\varepsilon_t = A(L)v_t = \sum_{j=0}^{\infty} A_j v_{t-j}$ with $A(1)^2 > 0$, $\sum_{l=0}^{\infty} l|A_l| < \infty$, $v_t \sim$ i.i.d. $(0, \sigma_v^2)$ and $\mathbf{E}|v_t|^q < \infty$ with $q > \max\{4, 2/(3-2d)\}$. The long-run variance of ε_t is $\sigma^2 := \sum_{k=-\infty}^{\infty} \mathbf{E}[\varepsilon_t \varepsilon_{t-k}]$.*

The spectral density of ε_t is defined as $f_\varepsilon(\lambda) = (1/2\pi) \sum_{j=-\infty}^{\infty} \gamma_j e^{-ij\lambda}$, where $i = \sqrt{-1}$ and $\{\gamma_j\}_{j=-\infty}^{\infty}$ is the sequence of autocovariances of ε_t, satisfying $f_\varepsilon(\lambda) \sim G$ as $\lambda \to 0_+$, where "$\sim$" means that the ratio of left- and right-hand sides tends to 1, G is positive and finite.

Assumption 2. *(i) $f_\varepsilon(\lambda) \sim G_0 \in (0, \infty)$ and, for some $\theta \in (0, 2]$, $f_\varepsilon(\lambda) = G_0(1 + O(\lambda^\theta))$ as $\lambda \to 0_+$; (ii) In a neighborhood $(0, \delta)$ of the origin, $A(e^{i\lambda})$ is differentiable and $(d/d\lambda)A(e^{i\lambda}) = O(\lambda^{-1})$ as $\lambda \to 0_+$; (iii) $f_\varepsilon(\lambda)$ is bounded for $\lambda \in [0, \pi]$.*

Assumptions 1 and 2 are mainly from Iacone et al. (2013) and Shimotsu (2010) and allow the estimate of d to be consistent and asymptotically normally distributed. They follow Marinucci and Robinson (2000) and allow for a functional central limit theorem for the partial sums of u_t. Applying a d-differencing transformation, the DGP is:

$$y_t^d := \Delta^d y_t = \beta_1 \Delta^d \mathbf{1}\{t \geq 1\} + \beta_2 \Delta^d t \mathbf{1}\{t \geq 1\} + \Delta^d u_t \mathbf{1}\{t \geq 1\}, \qquad (t = 1, ..., T).$$

Note that $\Delta^d u_t = \varepsilon_t$ and $\Delta^d y_1 = y_1$. For any matrix A, let A' denote its transpose. Let $\beta = (\beta_1, \beta_2)'$, $X_t = [1, t]'$ and $X_t^d := \Delta^d X_t = [\Delta^d \mathbf{1}\{t \geq 1\}, \Delta^d t \mathbf{1}\{t \geq 1\}]'$ with $\Delta^d X_1 = [1, 1]'$. The GLS transformed regression is then $y_t^d = X_t^{d\prime} \beta + \varepsilon_t$ $(t = 1, ..., T)$. To obtain a feasible regression, we replace d by some consistent estimate $\hat{d}$ to be discussed in the next section. The tests will then be based on the regression

$$y_t^{\hat{d}} = X_t^{\hat{d}\prime} \beta + u_t^{\hat{d}}, \qquad (t = 1, ..., T) \tag{3}$$

[1]The restriction that $d \neq 0.5$ is pervasive in the long memory literature because the case with $d = 0.5$ needs to be treated separately from the case with $d \neq 0.5$ (see, e.g., Tanaka, 1999; Iacone et al., 2013).

where $u_t^{\hat{d}} = \Delta^{\hat{d}} u_t \mathbf{1}\{t \geq 1\}$. Let $\hat{\beta} = (X^{\hat{d}\prime} X^{\hat{d}})^{-1} X^{\hat{d}\prime} y^{\hat{d}}$ denote the OLS estimator of β, where $X^{\hat{d}} = [X_1^{\hat{d}}, ..., X_T^{\hat{d}}]'$ and $y^{\hat{d}} = [y_1^{\hat{d}}, ..., y_T^{\hat{d}}]'$. The test statistic on the time trend coefficient β_2 for $H_0 : \beta_2 = \beta_2^0$ against $H_1 : \beta_2 \neq \beta_2^0$ is constructed as the usual t-statistic:

$$t_{\hat{\beta}_2}(\hat{d}, \hat{\sigma}^2) = R(\hat{\beta} - \beta^0)/[\hat{\sigma}^2 R(X^{\hat{d}\prime} X^{\hat{d}})^{-1} R']^{1/2}$$

where $R = [0 \quad 1]$, $\beta^0 = [\beta_1^0, \beta_2^0]'$ and $\hat{\sigma}^2$ is a consistent estimator of the long-run variance $\sigma^2 = \sum_{j=-\infty}^{\infty} \Gamma(j)$, where $\Gamma(j) = \mathbf{E}[\varepsilon_t \varepsilon_{t-j}]$. Similarly, the test statistic on the constant term β_1 for $H_0 : \beta_1 = \beta_1^0$ against $H_1 : \beta_1 \neq \beta_1^0$ can also be constructed as usual with:

$$t_{\hat{\beta}_1}(\hat{d}, \hat{\sigma}^2) = R_1(\hat{\beta} - \beta^0)/[\hat{\sigma}^2 R_1(X^{\hat{d}\prime} X^{\hat{d}})^{-1} R_1']^{1/2}$$

where $R_1 = [1 \quad 0]$. The next theorem provides the limit distribution of the test statistics.

Theorem 1. *Let $\{y_t\}$ be generated by (2) under Assumptions 1-2 and $\xrightarrow{d}$ denote convergence in distribution. With estimates such that $\hat{d} - d = O_p(T^{-\kappa})$ for any $\kappa > 0$ and $\hat{\sigma}^2 - \sigma^2 = o_p(1)$: (i) Under H_0: $\beta_2 = \beta_2^0$, $t_{\hat{\beta}_2}(\hat{d}, \hat{\sigma}^2) \xrightarrow{d} N(0, 1)$ for any $d \in (-0.5, 0.5) \cup (0.5, 1.5)$; (ii) Under H_0: $\beta_1 = \beta_1^0$, $t_{\hat{\beta}_1}(\hat{d}, \hat{\sigma}^2) \xrightarrow{d} N(0, 1)$ for any $d \in (-0.5, 0.5)$.*

Iacone et al. (2013) considered the fully extended local whittle (FELW) estimator of Abadir et al. (2007) to construct the test statistics based on $\hat{d}_{FELW}$-differences of the data. To establish the limiting distribution of $\hat{d}_{FELW}$, it is required that the bandwidth parameter $m = [c_1 T^n]$ with $0 < n < 0.8$ for some $c_1 > 0$ (Abadir et al., 2007, Corollary 2.1), where $[x]$ denotes the integer part of x. Note that the infeasible and feasible statistics share the same limit distribution if $n > \max\{0, 2(d - 1)\}$ (Iacone et al., 2013, Theorem 2). It imposes $d < 1.33$ with $n = 0.65$ and $d < 1.40$ with $n = 0.79$. Here, we only require $\hat{d}$ to be consistent at any polynomial rate for all $d \in (-0.5, 0.5) \cup (0.5, 1.5)$. Using our strategy to prove the result, it is easy to modify their proof, so that their Theorem 2 holds under the same condition as our Theorem 1; see Section 5.

Theorem 2 (Consistency)**.** *Let $\{y_t\}$ be generated by (2) under Assumptions 1-2: (i) Under H_1: $\beta_2 = \beta_2^1 \neq \beta_2^0$, $t_{\hat{\beta}_2}(\hat{d}, \hat{\sigma}^2)$ is consistent at rate $O_p(T^{3/2-d})$ for all $d \in (-0.5, 0.5) \cup (0.5, 1.5)$, (ii) Under H_1: $\beta_1 = \beta_1^1 \neq \beta_1^0$, $t_{\hat{\beta}_1}(\hat{d}, \hat{\sigma}^2)$ is consistent at rate $O_p(T^{1/2-d})$ for $d \in (-0.5, 0.5)$, while it is inconsistent for $d \in (0.5, 1.5)$.*

Theorem 2 states that the t-test on the time trend coefficient, $t_{\hat{\beta}_2}(\hat{d}, \hat{\sigma}^2)$, is consistent with $I(d)$ errors with $d \in (-0.5, 0.5) \cup (0.5, 1.5)$. The result for the t-test on the intercept, $t_{\hat{\beta}_1}(\hat{d}, \hat{\sigma}^2)$, shows that consistency depends on d. It is consistent with $d \in (-0.5, 0.5)$, i.e., stationary errors. For $d \in (0.5, 1.5)$, i.e., non-stationary errors, it is inconsistent as expected. Intuitively, $t_{\hat{\beta}_1}(\hat{d}, \hat{\sigma}^2)$

is inconsistent for $d \in (0.5, 1.5)$ because the non-stationary noise component u_t dominates the deviation from the null hypothesis, $\beta_1^1 - \beta_1^0$. It is not the case for $t_{\hat{\beta}_2}(\hat{d}, \hat{\sigma}^2)$.

A consistent estimate of σ^2 is readily available. Popular estimates are weighted sums of autocovariances of the form $\hat{\sigma}^2 = \hat{\Gamma}(0) + 2\sum_{j=1}^{T-1} h(j,l)\hat{\Gamma}(j)$, where $\hat{\Gamma}(j) = T^{-1}\sum_{t=j+1}^{T} u_t^{\hat{d}} u_{t-j}^{\hat{d}}$ with $u_t^{\hat{d}}$ the OLS residuals from the regression (3) and $h(\cdot, l)$ a kernel function with bandwidth l. In the simulations below, we use the Bartlett kernel and Andrews' (1991) data dependent method for selecting the bandwidth based on an AR(1) approximation. The choice of an appropriate estimate of d is more delicate and discussed in the next section.

3 Estimate of d

The exact local whittle (ELW) estimation procedure for d was studied by Shimotsu and Phillips (2005) and extended by Shimotsu (2010) for cases with an unknown trend function, which is required in our context. It is valid for a wide range of values for d, including values greater than 1. Accordingly, we shall adopt it as the estimator of d when constructing our test statistics. Let the discrete Fourier transform and the periodogram of y_t evaluated at the fundamental frequencies as $\omega_y(\lambda_j) = (2\pi T)^{-1/2}\sum_{t=1}^{T} y_t \exp(it\lambda_j)$ and $I_y(\lambda_j) = |\omega_y(\lambda_j)|^2$ for $\lambda_j = (2\pi j/T)$, $j = 1, ..., T$. The ELW estimator of d is the minimizer of

$$Q_m(G, d) = m^{-1}\sum_{j=1}^{m}[\log(G\lambda_j^{-2d}) + G^{-1}I_{\Delta^d y}(\lambda_j)].$$

Concentrating $Q_m(G, d)$ with respect to G, the objective function is $R(d) = \log \hat{G}(d) - 2d(m^{-1})\sum_{j=1}^{m}\log(\lambda_j)$ where $\hat{G}(d) = m^{-1}\sum_{j=1}^{m} I_{\Delta^d y}(\lambda_j)$ and, within a pre-specified range to be defined below, the ELW estimator is $\widetilde{d} = \arg\min_{d\in[\Delta_1,\Delta_2]} R(d)$. Shimotsu (2010) extended the ELW estimation procedure to cover an unknown linear time trend via a two-step procedure applied to detrended data. The first step detrends the data via OLS regression of y_t on $(1, t)$, with residuals denoted by $\hat{y}_t$. The modified objective function is:

$$R_F(d) = \log \hat{G}_F(d) - 2dm^{-1}\sum_{j=1}^{m}\log(\lambda_j), \quad \hat{G}_F(d) = m^{-1}\sum_{j=1}^{m} I_{\Delta^d(\hat{y}-\varphi(d))}(\lambda_j),$$

where $\varphi(d) = (1 - w(d))\hat{y}_1$ with $w(d)$ a twice continuous differentiable weight function such that $w(d) = 1$ for $d \leq 1/2$ and $w(d) = 0$ for $d \geq 3/4$. As recommended by Shimotsu (2010), $w(d) = (1/2)[1 + \cos(4\pi d)]$ for $d \in [1/2, 3/4]$. A two-step procedure is applied to ensure the global consistency of the estimate. In the first step, one uses the tapered local Whittle estimator of Velasco (1999), denoted $\hat{d}_T$, which is $\sqrt{m}$-consistent and invariant to a linear trend for $d \in (-1/2, 5/2)$. The second step estimator involves the following modification:

$$\hat{d}_{ELW}^{*} = \hat{d}_T - R_F'(\hat{d}_T)/R_F''(\hat{d}_T) \qquad (4)$$

where $R_F'(\hat{d}_T)$ and $R_F''(\hat{d}_T)$ are the 1^{st} and 2^{nd} derivatives of $R_F(d)$. As in Shimotsu (2010), we use $\max\{R_F''(\hat{d}_T), 2\}$ to improve the finite sample properties. The final estimator, $\hat{d}_{ELW}$, is obtained by iterating (4). To obtain the limiting distribution, we also need the following additional assumptions: (i) as $T \to \infty$, $m^{-1} + m^{1+2\theta}(\log m)^2 T^{-2\theta} + m^{-\gamma} \log T \to 0$ for any $\gamma > 0$; (ii) $-1/2 < \Delta_1 < \Delta_2 \leq 7/4$. Then, $\sqrt{m}(\hat{d}_{ELW} - d) \xrightarrow{d} N(0, 1/4)$ (Shimotsu, 2010, Theorem 4). Hence, with our test statistics constructed using $\hat{d}_{ELW}$, Theorem 1 continues to hold provided that $m = [c_1 T^n]$ for any $n > 0$ and some constant $c_1 > 0$. For all values of $d \in (-0.5, 0.5) \cup (0.5, 1.5)$, we can use a bandwidth that satisfies $m = [T^{0.65}]$.

4 Simulation Results

We now consider the size and power of the test $t_{\hat{\beta}_2}$ for the slope of the trend via simulations, using 1000 replications throughout.[2] The data are generated by (2), with u_t an autoregressive fractionally integrated moving average process, ARFIMA(p, d, q), of the form $(1-L)^d u_t = \varepsilon_t \mathbf{1}\{t \geq 1\}$ with $A(L)\varepsilon_t = B(L)e_t$, where $A(L) = 1 - a_1 L - ... - a_p L^p$ and $B(L) = 1 + b_1 L + ... + b_q L^q$ are the autoregressive and moving average lag polynomials, and $e_t \sim i.i.d.\, N(0, 1)$. Assumptions 1 and 2 are satisfied if the roots of $A(L) = 0$ and $B(L) = 0$ are outside the unit circle. In all cases, we set $\beta_1 = \beta_2 = 0$ under the null hypothesis without loss of generality. Also, the estimate $\hat{d}_{ELW}$ is constructed with $m = [T^{0.65}]$. We consider two-sided tests at the 5% significance level and for $d = 0$ or 1, the results are compared to those obtained with the two versions of the Perron and Yabu (2009) tests (PY), $t_\beta^{FS}(MU)$ or $t_\beta^{FS}(UB)$, which use different autoregressive estimates before applying the truncation (MU stands for Median Unbiased and UB for Upper Biased).

We start with the case of pure fractional processes with $A(L) = B(L) = 1$. We consider the range $d \in [-0.4, 1.4]$ and $T = 500$, 1000 and 2000. The results, presented in Table 1, show that the empirical size of $t_{\hat{\beta}_2}$ is close to the nominal size in all cases. On the other hand, $t_\beta^{FS}(MU)$ and $t_\beta^{FS}(UB)$ show substantial size distortions unless $d = 0$ or 1. When d is negative, these tests are very conservative, while for $0 < d < 1$, they have liberal size distortions, which are especially pronounced when $d = 1.4$. The power functions for a two-sided test of $\beta_2 = 0$ are presented in Figure 1 for $T = 500$. Given the size distortions of the PY tests when d is not 0 or 1, we include them only for $d = 1$ (we return below to the case $d = 0$). When $d = 1$, $t_\beta^{FS}(MU)$ and $t_\beta^{FS}(UB)$ have higher power, as expected, since they apply a truncation to 1 when the autoregressive parameter is in the neighborhood of 1, leading to a smaller bias when $d = 1$. The maximum power difference between $t_{\hat{\beta}_2}$ and the PY tests is 0.259. As expected, the power

[2] The results for the test $t_{\hat{\beta}_1}$ for the mean are qualitatively similar for the range $d \in (-0.5, 0.5)$, and hence can be omitted for brevity.

of $t_{\hat{\beta}_2}$ is highest when d is small, with power decreasing as d increases (note the different scaling).

Table 2 presents results for the size of the tests for processes with short-run dynamics of the autoregressive form with an $AR(1)$, so that $A(L) = 1 - aL$ with $d = 0$, cases for which the PY tests were designed. We consider values of a ranging from 0 to 0.95. The results show that the empirical size remains close to the nominal 5% level, unless a is close to 1, in which case the empirical size of $t_{\hat{\beta}_2}$ is below the nominal size. It is well known that in the presence of a short-run component with strong correlation, most estimates of d are biased. Accordingly, it is of some comfort to see that our test retains decent size and exhibits no liberal size distortions. The power functions for a two-sided test of $\beta_2 = 0$ are presented in Figure 2 for $T = 500$. When $a = 0$, 0.3 or 0.5, all tests have essentially the same power. When $a = 0.7$ or 0.9, the PY tests are more powerful and the maximum power differences are 0.382 and 0.548, respectively. When $a = 0.95$, $t_{\hat{\beta}_2}$ has much higher power, despite being conservative, unless the alternative is close to the null value. The maximum power difference between $t_{\hat{\beta}_2}$ and the PY tests is 0.359.

We next consider the size and power of the tests using five different DGPs used in Qu (2011), which were motivated by financial applications of interest. The following DGPs are used. **DGP-1:** ARFIMA$(1, d, 0)$, $(1 - a_1L)(1 - L)^{0.4}\varepsilon_t = e_t$, where $a_1 = 0.4$ and -0.4; **DGP-2**: ARFIMA$(0, d, 1)$, $(1-L)^{0.4}\varepsilon_t = (1+b_1L)e_t$, where $b_1 = 0.4$ and -0.4; **DGP-3:** ARFIMA$(2, d, 0)$, $(1 - a_1L)(1 - a_2L)(1 - L)^{0.4}\varepsilon_t = e_t$, with $a_1 = 0.3$, $a_2 = 0.5$; **DGP-4**: $\varepsilon_t = z_t + \eta_t$, where $(1 - L)^{0.4}z_t = e_t$ and $\eta_t \sim i.i.d\ N(0, var(z_t))$; **DGP-5:** $(1 - L)^{0.4}\varepsilon_t = \eta_t$ with $\eta_t = \sigma_t e_t$, $\sigma_t^2 = 1+0.1\eta_{t-1}^2+0.85\sigma_{t-1}^2$. In all cases, $e_t \sim i.i.d.\, N(0, 1)$. **DGP-6:** ARFIMA$(0, d, 0)$, $(1-L)^{0.4}\varepsilon_t = e_t$ with $e_t \sim i.i.d.\, t(5)$. DGPs 1-3 are different cases of ARFIMA processes, DGP-4 is a fractionally integrated process with measurement errors, DGP-5 is a generalized autoregressive conditional heteroskedasticity (GARCH) process, and DGP-6 is a heavy-tailed process. Note that DGPs 4-6 do not satisfy the conditions of Assumptions 1-2. We nevertheless include them to assess the robustness of the results given that measurement errors, conditional heteroskedasticity, and heavy-tailed process are prevalent features of many time series. Given the size distortions of the PY tests when d is different from 0 or 1, we only present results for the test $t_{\hat{\beta}_2}$. Table 3 shows that the empirical size is near 5%, except for DGPs 2 and 5, for which the test has slight liberal size distortions when $T = 500$, which decrease as T increases. The power functions of the test for $T = 500$ are presented in Figure 3. In all cases, power increases rapidly to 1 as β_2 deviates from 0, except, perhaps, with GARCH errors. Comparing across DGPs, power decreases when additional short-run dynamics are present. The effect of measurement errors and heavy-tailed process on the power is minor.

To further confirm the consistency result in Theorem 2, we simulated the power of one-sided tests $t_{\hat{\beta}_1}$ and $t_{\hat{\beta}_2}$ using a DGP with u_t following a pure fractional process ($A(L) = B(L) = 1$). The (unreported) results show the power of our tests increasing with T, though the increase is slower as d gets closer to 1.5 for $t_{\hat{\beta}_2}$ (slope) and close to 0.5 for $t_{\hat{\beta}_1}$ (intercept).

5 Empirical Application

To illustrate the usefulness of our tests, we apply the proposed test statistics to the U.S. equity indices. The daily closing price of four equity indices in the U.S. stock market, S&P 500 index, NASDAQ composite index, Dow Jones Industrial Average, Wilshire 5000 Total Market Full Cap Index, were obtained from Federal Reserve Economic Data (FRED) database and a logarithmic transformation was applied. The sample period is from January 17, 2012 to January 13, 2022 for all indices. Table 4 presents test results for the null hypothesis of no trend, that is, $\mathbb{H}_0 : \beta_2 = 0$. We also report the associated estimate of the order of fractional integration d following the estimation procedure in Section 3, denoted as $\hat{d}_{ELW}$. Figure 4 plots the series and the fitted deterministic components, while Figure 5 plots the detrended series.

The $t_{\hat{\beta}_2}$, $t_{\beta}^{FS}(MU)$, and $t_{\beta}^{FS}(UB)$ tests detect a linear trend for each of four equity indices, which is consistent with small sample power results reported in Figures 1-2. Furthermore, examining the values of $\hat{d}_{ELW}$ in Table 4, we can infer that the residuals from the detrended series are compatible with processes which are $I(d)$ and non-stationary. As a further indication regard to the integration properties of the data, the augmented Dickey-Fuller test (Dickey and Fuller, 1979; Said and Dickey, 1984), denoted as "ADF", is also reported in Table 4. The unit root null hypothesis is rejected for the S&P 500, Dow Jones Industrial Average, and Wilshire 5000 Total Market Full Cap Index at the 5%, 1%, and 10% significance level, respectively. For NASDAQ composite index, the unit root null is not rejected at conventional significance level. It is well known that standard unit root tests often reject the unit root null hypothesis when the true process is fractionally integrated with $d \in (0.5, 1)$. This can lead to the misleading conclusion that the process of interest is stationary. Then, we need procedures for trend detection that are robust to the order of fractional integration of the data being considered. The proposed testing procedures in this paper achieve such robustness asymptotically, and offer useful complements to existing tests. We may conclude that deterministic trends exist in the U.S. equity indices without deciding whether the noise component is either $I(0)$ or $I(1)$.

6 Extension

Iacone et al. (2013) considered the problem of testing for a break in trend when the noise component is a fractional process, i.e., testing whether $\beta_3 = 0$ in the model:

$$y_t = \beta_1 + \beta_2\, t + \beta_3\, DT_t(\tau_0) + u_t, \quad (t = 1, \ldots, T) \tag{5}$$

with $DT_t(\tau_0) := (t - [\tau_0 T])\mathbf{1}\{t > [\tau_0 T]\}$, τ_0 unknown and u_t satisfying Assumption 1. The basic method is to take a quasi-difference of the data and trend regressors using an estimate $\hat{d}$ of the order of integration $d \in [0, 0.5) \cup (0.5, 1.5)$. A sup-Wald test $\mathcal{SW}(\hat{d}, \hat{\sigma}^2)$ is then applied to the transformed regression. When the rate of convergence of the estimate of d is $T^{n/2}$, that is, $T^{n/2}(\hat{d} - d) = O_p(1)$, they showed that the feasible version of the sup-Wald test $\mathcal{SW}(\hat{d}, \hat{\sigma}^2)$ shares the same limit distribution as $\mathcal{SW}(d, \sigma^2)$ provided $n > \max\{0, 2(d-1)\}$. This result imposes restrictions on implementing the sup-Wald test $\mathcal{SW}(\hat{d}, \hat{\sigma}^2)$. The rate of convergence for the estimate of d should be faster for $d > 1$, while for a given rate of convergence, the distributional equivalence between $\mathcal{SW}(d, \sigma^2)$ and $\mathcal{SW}(\hat{d}, \hat{\sigma}^2)$ does not hold for all values of d in the range considered. We show that a simple modification of their proof, using our approach to prove Theorem 1, makes the results hold for any $n > 0$.

Theorem 3 (Theorem 2 of Iacone et al., 2013)**.** *Let $\{y_t\}$ be generated according to (5), and let Assumption 1 hold. Also, let $m := [c_1 T^n]$, for some constants $c_1 > 0$ and $n > 0$. Then, under a local alternative of the form $H_1^{\kappa,d} : \beta_3 = \kappa T^{d-3/2}$, uniformly in τ, $\mathcal{W}(\hat{d}, \tau, \hat{\sigma}^2) - \mathcal{W}(d, \tau, \sigma^2) = o_p(1)$, $\mathcal{SW}(\hat{d}, \hat{\sigma}^2) - \mathcal{SW}(d, \sigma^2) = o_p(1)$.*

7 Conclusion

We provided tests to perform inference on the coefficients of a linear trend function, assuming the noise to be a fractionally integrated process with memory parameter in the interval $(-0.5, 1.5)$, excluding the boundary case 0.5. The results are encouraging in that our test is valid under much more general conditions, yet has power approaching to that of the Perron and Yabu (2009) tests designed only for the dichotomous cases with d either being 0 or 1. When d is different from 0 or 1, its empirical size is close to the nominal size, and power is good. Our procedure provides a useful tool for inference about the coefficient of a linear trend under general conditions about the noise component. Though we assumed the errors to follow a Type II long memory process, we conjecture that our results remain valid with a Type I process, as defined by Marinucci and Robinson (1999). First, as Shimotsu (2010) argued, his results remain valid for both types of processes. Also, the conditions for a functional central limit theorem for Type I processes are very similar (see, e.g., Wang et al., 2003) and could be slightly modified accordingly. We used our method of proof to show that the main result of Iacone et al. (2013) is valid for the full range $d \in (-0.5, 1.5)$, excluding $d = 0.5$.

References

Abadir, K. M., Distaso, W. & Giraitis, L. (2007) Nonstationarity-extended local Whittle estimation. *Journal of Econometrics* **141**: 1353-1384.

Abadir, K. M., Distaso, W. & Giraitis, L. (2011) An I(d) model with trend and cycles. *Journal of Econometrics* **163**: 186-199.

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

Brockwell, P. J. & Davis, R. A. (1991) *Time Series: Theory and Methods* (2nd ed.). New York: Springer-Verlag.

Bunzel, H. & Vogelsang, T. J. (2005) Powerful trend function tests that are robust to strong serial correlation with an application to the Prebish-Singer hypothesis. *Journal of Business and Economic Statistics* **23**: 381-394.

Canjels, E. & Watson, M. W. (1997) Estimating deterministic trends in the presence of serially correlated errors. *Review of Economics and Statistics* **79**: 184-200.

Dickey, D.A. & Fuller, W.A. (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of American Statistical Association* **74**: 427–431.

Granger, C.W.J. (1980) Long memory relationship and the aggregation of dynamic models. *Journal of Econometrics* **14**: 227-238.

Grenander, U. & Rosenblatt, M. (1957) *Statistical Analysis of Stationary Time Series*, New York: John Wiley.

Harvey, D. I., Leybourne, S. J. & Taylor, A. M. R. (2007) A simple, robust and powerful test of the trend hypothesis. *Journal of Econometrics* **141**: 1302-1330.

Iacone, F., Leybourne, S. J. & Taylor, A. M. R. (2013) Testing for a break in trend when the order of integration is unknown. *Journal of Econometrics* **176**: 30-45.

Krantz, S. G. (1999) "T he Gamma and Beta Functions." 13.1 in Handbook of Complex Variables. Boston, MA: Birkhauser, 155-158.

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

Marinucci, D. & Robinson, P. M. (2000) Weak convergence of multivariate fractional Brownian motion. *Stochastic Processes and their Applications* **86**: 103-120.

Perron, P. (1988) Trends and random walks in macroeconomic time series: Further evidence from a new approach. *Journal of Economic Dynamics and Control* **12**: 297-332.

Perron, P. & Yabu, T. (2009) Estimating deterministic trends with an integrated or stationary noise component. *Journal of Econometrics* **151**: 56-69.

Perron, P. & Yabu, T. (2012) Testing for trend in the presence of autoregressive error: a comment. *Journal of the American Statistical Association* **107**: 844.

Qu, Z. (2011) A test against spurious long memory. *Journal of Business and Economic Statistics* **29**: 423-438.

Robinson, P. M. (2005) Efficiency improvements in inference on stationary and nonstationary fractional time series. *Annals of Statistics* **33**: 1800-1842.

Robinson, P. M. & Iacone, F. (2005) Cointegration in fractional systems with deterministic trends. *Journal of Econometrics* **129**: 263-298.

Roy, A., Falk, B. & Fuller, W. A. (2004) Testing for trend in the presence of autoregressive errors. *Journal of the American Statistical Association* **99**: 1082-1091.

Said, S.E. & Dickey, D.A. (1984) Testing for unit roots in autoregressive moving average models with unknown order *Biometrika* **71**: 599-607.

Shimotsu, K. (2010) Exact local Whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory* **26**: 501-540.

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

Sun, H. & Pantula, S. G. (1999) Testing for trends in correlated data. *Statistics and Probability Letters* **41**: 87-95.

Tanaka, K. (1999) The nonstationary fractional unit root. *Econometric Theory* **15**: 549-582.

Velasco, C. (1999) Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* **20**: 87-127.

Vogelsang, T. J. (1998) Trend function hypothesis testing in the presence of serial correlation. *Econometrica* **66**: 123-148.

Wang, Q., Lin, Y.-X. & Gulati, C. M. (2003) Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory* **19**: 143-164.

Appendix

As a matter of notation, let $X^d = [X_1^d, ..., X_T^d]'$, $X_t^d = [\mu_{0,t}\ \mu_{1,t}]'$, $\mu_{i,t} = \Delta^d t^i \mathbf{1}\{t \geq 1\}$ for $i = \{0, 1\}$, and $\varepsilon = [\varepsilon_1, ..., \varepsilon_T]'$. Also, $W(r)$ is a standard Brownian motion with $\mathbf{E}[W(r)^2] = r$. Throughout the appendix, $\mathcal{C}$ denotes a finite generic constant whose specific value is not crucial. We define a fractionally integrated process following the notation of Wang et al. (2003) and Robinson (2005). Define the difference operator $\Delta^{-d} = (1-L)^{-d}$ as

$$\Delta^{-d} = \sum_{k=0}^{\infty} \pi_{k,d} L^k,\ \pi_{k,d} = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)},$$

where L denotes the lag operator, $\Delta = 1 - L$ is the difference operator, and Γ is the Gamma function with $\Gamma(d) = \infty$ for $d = 0, -1, \ldots$, and $\Gamma(0)/\Gamma(0) = 1$. Let $\{\varepsilon_t, t = 0, \pm 1, \ldots\}$ be a zero-mean short-memory covariance stationary process, with spectral density that is bounded and bounded away from zero. For $d \in (-0.5, 0.5)$, $\zeta_t = \Delta^{-d}\varepsilon_t$, $(t = 0, \pm 1, \ldots)$, is covariance stationary and invertible for $d > -0.5$. The truncated version of ζ_t is defined as $\zeta_t^{\#} = \zeta_t \mathbf{1}\{t \geq 1\}$, $(t = 0, \pm 1, \ldots)$. For an integer $m \geq 0$, $u_t = \Delta^{-m}\zeta_t^{\#}$, $(t = 0, \pm 1, ...)$ is called a Type I $I(m+d)$ process. A zero-mean short-memory covariance stationary process ε_t can be represented as a one-sided moving average: $\varepsilon_t = A(L)v_t = \sum_{j=0}^{\infty} A_j v_{t-j}$ $(t = 0, \pm 1, \ldots)$, where $\psi_0 = 1$, $A(1)^2 > 0$, $\sum_{l=0}^{\infty} l|A_l| < \infty$ and v_t $(t = 0, \pm 1, \ldots)$ are i.i.d. random variables with mean zero and variance $\sigma_v^2 < \infty$. We state the following lemmas, whose proofs follow Iacone et al. (2013), with appropriate modifications for the results to hold under the conditions of Theorem 1. All limit statements are taken as $T \to \infty$.

Lemma A.1. *Suppose that d and σ^2 are known: (i) For $-0.5 < d < 0.5$: $t_{\hat{\beta}_2}(d, \sigma^2) \xrightarrow{d} RC^{-1}L/[RC^{-1}R']^{1/2} := A_1$, where*

$$C = \begin{bmatrix} \frac{1}{[\Gamma(1-d)]^2(1-2d)} & \frac{1}{\Gamma(1-d)\Gamma(2-d)(2-2d)} \\ \frac{1}{\Gamma(1-d)\Gamma(2-d)(2-2d)} & \frac{1}{[\Gamma(2-d)]^2(3-2d)} \end{bmatrix}, \quad L = \begin{bmatrix} \frac{1}{\Gamma(1-d)} \int_0^1 r^{-d} dW(r) \\ \frac{1}{\Gamma(2-d)} \int_0^1 r^{1-d} dW(r) \end{bmatrix},$$

and

$$A_1 = \sqrt{\frac{3-2d}{1-d}} \left[2(1-d) \int_0^1 r^{1-d} dW(r) - (1-2d) \int_0^1 r^{-d} dW(r) \right].$$

(ii) For $0.5 < d < 1.5$: $t_{\hat{\beta}_2}(d, \sigma^2) \xrightarrow{d} R\widetilde{C}^{-1}\widetilde{L}/[R\widetilde{C}^{-1}R']^{1/2} = C_{22}^{-1/2} L_2 := A_2$, where C_{22} and L_2 are the relevant sub-matrix and sub-vector of C and L, and $A_2 = \sqrt{3-2d} \int_0^1 r^{1-d} dW(r)$.

Proof of Lemma A.1 From Robinson (2005, Lemma 1), as $t \to \infty$, for $d \in (-0.5, 1)$, $\Delta^d \mathbf{1}\{t \geq 0\} = \Gamma(1-d)^{-1} t^{-d} + O(t^{-1})$ and $\Delta^d t \mathbf{1}\{t \geq 0\} = \Gamma(2-d)^{-1} t^{1-d} + O(1)$, while for $d \in (1, 1.5)$, $\Delta^d t \mathbf{1}\{t \geq 0\} = \Gamma(2-d)^{-1} t^{1-d} + O(t^{-1})$. From Robinson and Iacone (2005, eq. A.34), for any $r \in (0, 1]$, with $[rT]$ as the integer part of rT, we have (a) for $d \in (-0.5, 1)$: $T^d \Delta^d \mathbf{1}\{[rT] \geq 0\} \to \Gamma(1-d)^{-1} r^{-d}$; (b) for $d \in (-0.5, 1.5)$, $T^{d-1} \Delta^d [rT] \mathbf{1}\{[rT] \geq 0\} \to \Gamma(2-d)^{-1} r^{1-d}$. For part (i), using $K_T := diag\{T^{1/2-d}, T^{3/2-d}\}$, we have:

$$t_{\hat{\beta}_2}(d,\sigma^2) = \frac{R(\hat{\beta}-\beta^0)}{[\sigma^2 R(X^{d'}X^d)^{-1}R']^{1/2}} = \frac{R(K_T^{-1}X^{d'}X^dK_T^{-1})^{-1}(K_T^{-1}X^{d'}\varepsilon)}{[\sigma^2 R(K_T^{-1}X^{d'}X^dK_T^{-1})^{-1}R']^{1/2}} \xrightarrow{d} \frac{RC^{-1}L}{[RC^{-1}R']^{1/2}},$$

where $K_T^{-1}X^{d\prime}X^dK_T^{-1} \to C$ and $K_T^{-1}X^{d\prime}\varepsilon \xrightarrow{d} \sigma L$ are proved in Robinson and Iacone (2005, eq. A.36). For part (ii), using $\widetilde{K}_T := diag\{1, T^{3/2-d}\}$, the t-statistic for β_2 follows:

$$t_{\hat{\beta}_2}(d,\sigma^2) = \frac{R(\widetilde{K}_T^{-1}X^{d'}X^d\widetilde{K}_T^{-1})^{-1}(\widetilde{K}_T^{-1}X^{d'}\varepsilon)}{[\sigma^2 R(\widetilde{K}_T^{-1}X^{d'}X^d\widetilde{K}_T^{-1})^{-1}R']^{1/2}}.$$

The result is obtained given that for $0.5 < d < 1.5$,

$$(\widetilde{K}_T^{-1}X^{d'}X^d\widetilde{K}_T^{-1})^{-1} \to \begin{bmatrix} O(1) & 0 \\ 0 & C_{22}^{-1} \end{bmatrix}, \quad \widetilde{K}_T^{-1}X^{d'}\varepsilon \xrightarrow{d} \sigma\widetilde{L} := \begin{bmatrix} O_p(1) \\ \sigma L_2 \end{bmatrix}.$$

Lemma A.2. *Suppose that d and σ^2 are known: For $d \in (-0.5, 0.5)$: $t_{\hat{\beta}_1}(d,\sigma^2) \xrightarrow{d} R_1C^{-1}L/[R_1C^{-1}R_1']^{1/2} := B_1$, where $B_1 = [(1-2d)/(1-d)]^{1/2}[2(1-d)\int_0^1 r^{-d}dW(r) - (3-2d)\int_0^1 r^{1-d}dW(r)]$.*

Proof of Lemma A.2. The proof can be completed by using $R_1 = [1 \quad 0]$ instead of $R = [0 \quad 1]$ in Lemma A.1, and hence can be omitted.

Lemma A.3. *A_1, A_2 and B_1, defined in Lemmas A.1-A.2 have a $N(0,1)$ distribution.*

Proof of Lemma A.3. When $d \in (-0.5, 0.5)$, it is easy to show that the normally distributed bivariate random vector L has variance-covariance matrix C. Therefore, A_1 is also normally distributed with variance $(RC^{-1}CC^{-1}R')/(RC^{-1}R') = 1$. When $d \in (0.5, 1.5)$, L_2 is a normally distributed random variable with variance C_{22}. Therefore $A_2 = C_{22}^{-1/2}L_2$ is also normally distributed with variance one. It is straightforward to show that B_1 follows a standard normal distribution based on the arguments for A_1.

Lemma A.4. *Under Assumption 1, for $d \in (-0.5, 0.5) \cup (0.5, 1.5)$:*

$$T^{d-3/2}\sum_{t=1}^{T}\mu_{1,t-s}\varepsilon_{t-s} = O_p(1), \tag{A.1}$$

and

$$T^{d-1/2}\sum_{t=1}^{T}\mu_{0,t-s}\varepsilon_{t-s} = O_p(1), \tag{A.2}$$

uniformly in $s \in \mathbb{N}$.

Proof of Lemma A.4. Define $\Xi_s := T^{d-3/2}\sum_{t=1}^{T}\mu_{1,t-s}\varepsilon_{t-s}$, $s = 1, 2, \ldots$. Then, we have

$$\begin{aligned}\sup_{s\in\mathbb{N}} \mathbf{E}[\Xi_s^2] &= T^{2d-3}\sup_{s\in\mathbb{N}} \mathbf{E}[(\textstyle\sum_{t=1}^{T}\mu_{1,t-s}\varepsilon_{t-s})^2] \le T^{2d-3}\mu_{1,T}^2 \sup_{s\in\mathbb{N}} \mathbf{E}[(\sum_{t=1}^{T}\varepsilon_{t-s})^2] \\ &= T^{2d-2}\mu_{1,T}^2 \sup_{s\in\mathbb{N}} \mathbf{E}[(T^{-1/2}\textstyle\sum_{t=1}^{T}\varepsilon_{t-s})^2] = O(1),\end{aligned}$$

where the last equality holds due to (b) in the proof of Lemma A.1 and the functional central limit theorem $T^{-1/2}\sum_{t=1}^{T}\varepsilon_{t-s} \Rightarrow \sigma_v A(1)W(1)$ for all $s \in \mathbb{N}$ under Assumption 1, thereby $\sup_{s\in\mathbb{N}}\mathbf{E}[(T^{-1/2}\sum_{t=1}^{T}\varepsilon_{t-s})^2] \leq \sigma_v^2 A(1)^2$. The symbol $\Rightarrow$ denotes weak convergence in Skorohod topology. Using Proposition 6.2.3 of Brockwell and Davis (1991), Ξ_s is bounded by $O_p(1)$ uniformly in $s \in \mathbb{N}$. Similarly, it is straightforward to show that

$$\begin{aligned} T^{2d-1}\mathbf{E}[(\textstyle\sum_{t=1}^{T}\mu_{0,t}\varepsilon_{t-s})^2] &\leq T^{2d-1}\mu_{0,T}^2\mathbf{E}[(\textstyle\sum_{t=1}^{T}\varepsilon_{t-s})^2] \\ &= T^{2d}\mu_{0,T}^2\mathbf{E}[(T^{-1/2}\textstyle\sum_{t=1}^{T}\varepsilon_{t-s})^2] = O(1), \end{aligned}$$

using (a) in the proof of Lemma A.1, which establishes (A.2).

Lemma A.5. $\sum_{k=1}^{\infty}|\pi_{k,d-\hat{d}}| \xrightarrow{p} 0$.

Proof of Lemma A.5. As in Granger (1980), for $d-\hat{d} \in R$,

$$\pi_{k,d-\hat{d}} = \frac{\Gamma(k+d-\hat{d})}{\Gamma(k+1)\Gamma(d-\hat{d})} \approx k^{d-\hat{d}-1}\frac{1}{\Gamma(d-\hat{d})},$$

using Stirling's approximation $\Gamma(k+a)/\Gamma(k+b) \approx k^{a-b}$ ($\approx$ denoting approximate equality for large k). Note that $\sum_{k=1}^{T}k^{-d} \leq CT^{1-d}$ if $d<1$, and $\sum_{k=1}^{T}k^{-d} \leq C$ if $d>1$. Then,

$$\textstyle\sum_{k=1}^{T}k^{d-\hat{d}-1} \leq \begin{cases} \mathcal{C}T^{d-\hat{d}} & \text{if } d-\hat{d}>0, \\ \mathcal{C} & \text{if } d-\hat{d}<0. \end{cases}$$

Using the continuous mapping theorem, $T^{d-\hat{d}} = (e^{\ln T})^{d-\hat{d}} = e^{(\ln T)(d-\hat{d})} \xrightarrow{p} \exp(0) = 1$ if $d-\hat{d}>0$. Moreover, the reciprocal gamma function $1/\Gamma(z)$ has the infinite product form (see, e.g., Krantz, 1999)

$$\frac{1}{\Gamma(z)} = ze^{\gamma z}\prod_{r=1}^{\infty}\left(1+\frac{z}{r}\right)e^{-z/r} = z\prod_{r=1}^{\infty}\frac{1+\frac{z}{r}}{(1+\frac{1}{r})^z},\ z\in\mathbb{C}$$

that converges uniformly in compact sets. Given that $d-\hat{d} \xrightarrow{p} 0$, there exists some constant C such that

$$P\left(\left|\frac{1}{\Gamma(d-\hat{d})}\right| > \epsilon\right) = P\left(|d-\hat{d}|\left|\prod_{r=1}^{\infty}\frac{1+\frac{d-\hat{d}}{r}}{(1+\frac{1}{r})^{d-\hat{d}}}\right| > \epsilon\right) \leq P\left(|d-\hat{d}| > \mathcal{C}\epsilon\right) \to 0$$

for all $\epsilon > 0$ as $T \to \infty$. Because $\sum_{k=1}^{\infty}|\pi_{k,d-\hat{d}}| = |\Gamma(d-\hat{d})^{-1}|\sum_{k=1}^{\infty}k^{d-\hat{d}-1}$, it follows that $\sum_{k=1}^{\infty}|\pi_{k,d-\hat{d}}| = o_p(1)O_p(1) = o_p(1)$, which establishes Lemma A.5, as desired.

Lemma A.6. *For given $\tau \in [\tau_L,,\tau_U] := \Lambda \subset [0,1]$,*

$$T^{d-3/2}\textstyle\sum_{t=1}^{T}\mu_t(\tau)\varepsilon_{t-s} = O_p(1), \tag{A.3}$$

uniformly in $s \in \mathbb{N}$ where $\mu_t(\tau) = \Delta^d(t-[\tau T])\mathbf{1}\{t>[\tau T]\}$.

Proof of Lemma A.6. It is straightforward to show that

$$\sup_{s\in\mathbb{N}} \mathbf{E}[T^{2d-3}(\textstyle\sum_{t=1}^{T}\mu_t(\tau)\varepsilon_{t-s})^2] \le T^{2d-3}\mu_T(\tau)^2 \sup_{s\in\mathbb{N}} \mathbf{E}[(\textstyle\sum_{t=1}^{T}\varepsilon_{t-s})^2]$$
$$= T^{2d-2}\mu_T(\tau)^2 \sup_{s\in\mathbb{N}} \mathbf{E}[(T^{-1/2}\textstyle\sum_{t=1}^{T}\varepsilon_{t-s})^2] = O(1),$$

using the arguments in the proof of Lemma A.4, which establishes (A.3).

Proof of Theorem 1. We establish the limiting distributions of $t_{\hat{\beta}_l}(\hat{d},\hat{\sigma}^2), l = \{1,2\}$ with consistent estimates $(\hat{d},\hat{\sigma}^2)$. What remains to be shown is that $t_{\hat{\beta}_l}(d,\sigma^2)$ and $t_{\hat{\beta}_l}(\hat{d},\hat{\sigma}^2)$ $(l = \{1,2\})$ share the same limiting distribution. It is trivial to show that the results are the same when using a consistent estimate of σ^2, hence we concentrate on using an estimate of d. We need to show that if $\hat{d} - d = O_p(T^{-\kappa})$ for any $\kappa > 0$, then (a) for $-0.5 < d < 0.5$,

$$K_T^{-1}X^{\hat{d}\prime}X^{\hat{d}}K_T^{-1} - K_T^{-1}X^{d\prime}X^{d}K_T^{-1} \xrightarrow{d} 0, \tag{A.4}$$

and

$$K_T^{-1}X^{\hat{d}\prime}u^{\hat{d}} - K_T^{-1}X^{d\prime}\varepsilon \xrightarrow{d} 0, \tag{A.5}$$

(b) for $0.5 < d < 1.5$,

$$\widetilde{K}_T^{-1}X^{\hat{d}\prime}X^{\hat{d}}\widetilde{K}_T^{-1} - \widetilde{K}_T^{-1}X^{d\prime}X^{d}\widetilde{K}_T^{-1} \xrightarrow{d} 0, \tag{A.6}$$

and

$$\widetilde{K}_T^{-1}X^{\hat{d}\prime}u^{\hat{d}} - \widetilde{K}_T^{-1}X^{d\prime}\varepsilon \xrightarrow{d} 0, \tag{A.7}$$

where $u_t^{\hat{d}} := \Delta^{\hat{d}}u_t\mathbf{1}\{t \ge 1\}$ and $\hat{\mu}_{i,t} := \Delta^{\hat{d}}t^i\mathbf{1}\{t \ge 1\}$ for $i = \{0,1\}$. Consider first (A.4) and (A.6). For (A.4), that is, $-0.5 < d < 0.5$, we show that for $i,j = \{0,1\}$,

$$T^{2d-1-i-j}(\textstyle\sum_{t=1}^{T}\hat{\mu}_{i,t}\hat{\mu}_{j,t} - \sum_{t=1}^{T}\mu_{i,t}\mu_{j,t}) \xrightarrow{d} 0,$$

or equivalently,

$$T^{2d-1-i-j}[\textstyle\sum_{t=1}^{T}(\hat{\mu}_{i,t}-\mu_{i,t})\mu_{j,t} + \sum_{t=1}^{T}\mu_{i,t}(\hat{\mu}_{j,t}-\mu_{j,t}) + \sum_{t=1}^{T}(\hat{\mu}_{i,t}-\mu_{i,t})(\hat{\mu}_{j,t}-\mu_{j,t})] \xrightarrow{d} 0. \tag{A.8}$$

As shown in Iacone et al. (2013, eq. A.22), for $i \in \{0,1\}$ and $d \le 1$, and for $i = 1$ and $d > 1$,

$$\hat{\mu}_{i,t} - \mu_{i,t} = o_p(t^{i-d}). \tag{A.9}$$

By the Cauchy-Schwarz inequality,

$$T^{2d-1-i-j}\textstyle\sum_{t=1}^{T}(\hat{\mu}_{i,t}-\mu_{i,t})\mu_{j,t} \le T^{2d-1-i-j}[\sum_{t=1}^{T}(\hat{\mu}_{i,t}-\mu_{i,t})^2\sum_{t=1}^{T}\mu_{j,t}^2]^{1/2}. \tag{A.10}$$

In the view of (A.9) and $|\mu_{j,t}| \le Ct^{j-d}$, it is straightforward to show that the expression on the RHS of (A.10) is $o_p(T^{2d-1-i-j+i-d+1/2+j-d+1/2}) = o_p(1)$. Following previous arguments, the other terms in (A.8) are $o_p(1)$, which establishes (A.4). For (A.6), that is, $0.5 < d < 1.5$, we show that $T^{(i+j)(d-3/2)}(\sum_{t=1}^T \hat{\mu}_{i,t}\hat{\mu}_{j,t} - \sum_{t=1}^T \mu_{i,t}\mu_{j,t}) \xrightarrow{d} 0$,or equivalently,

$$T^{(i+j)(d-3/2)}[\sum_{t=1}^{T}(\hat{\mu}_{i,t}-\mu_{i,t})\mu_{j,t}+\sum_{t=1}^{T}\mu_{i,t}(\hat{\mu}_{j,t}-\mu_{j,t})+\sum_{t=1}^{T}(\hat{\mu}_{i,t}-\mu_{i,t})(\hat{\mu}_{j,t}-\mu_{j,t})] \xrightarrow{d} 0.$$

For $i = j = 1$, the proof is similar to that for $-0.5 < d < 0.5$. For $i = 0$, $j = 1$,

$$T^{d-3/2}\sum_{t=1}^T(\hat{\mu}_{1,t}-\mu_{1,t})\mu_{0,t} \le T^{d-3/2}[\sum_{t=1}^T(\hat{\mu}_{1,t}-\mu_{1,t})^2\sum_{t=1}^T\mu_{0,t}^2]^{1/2} = o_p\left(T^{d-3/2+3/2-d}\right) = o_p(1).$$

Similar arguments apply when $i = j = 0$. The proof for the other terms is similar, which completes the proof of (A.6). For (A.5) and (A.7), we first show that for $d \in (-0.5, 0.5) \cup (0.5, 1.5)$,

$$T^{d-3/2}\sum_{t=1}^T\hat{\mu}_{1,t}u_t^{\hat{d}} - T^{d-3/2}\sum_{t=1}^T\mu_{1,t}\varepsilon_t \xrightarrow{d} 0,$$

or equivalently,

$$T^{d-3/2}[\sum_{t=1}^T(\hat{\mu}_{1,t}-\mu_{1,t})\varepsilon_t+\sum_{t=1}^T\mu_{1,t}(u_t^{\hat{d}}-\varepsilon_t)+\sum_{t=1}^T(\hat{\mu}_{1,t}-\mu_{1,t})(u_t^{\hat{d}}-\varepsilon_t)] \xrightarrow{d} 0.$$

Now, we show that $T^{d-3/2}\sum_{t=1}^T \mu_{1,t}(u_t^{\hat{d}} - \varepsilon_t) \xrightarrow{p} 0$. As explained in Section 3, $\hat{d}_{ELW}$ is consistent for all values of $d \in [\Delta_1, \Delta_2]$, where $-1/2 < \Delta_1 < \Delta_2 \le 7/4$. For ease of notation, we suppress the subscript ELW and write $\hat{d}$. Since $u_t = \Delta^{-d}\varepsilon_t\mathbf{1}\{t \ge 1\}$, $\hat{d}$-differences of u_t is given by $u_t^{\hat{d}} := \Delta^{\hat{d}}u_t = \Delta^{\hat{d}-d}\varepsilon_t\mathbf{1}\{t \ge 1\}$. Note that $u_t^d := \Delta^d u_t = \varepsilon_t\mathbf{1}\{t \ge 1\}$. Therefore,

$$\begin{aligned}\Delta^{\hat{d}}u_t - \Delta^d u_t &= (\Delta^{\hat{d}-d}\varepsilon_t - \varepsilon_t)\mathbf{1}\{t \ge 1\} = (\textstyle\sum_{k=0}^\infty \pi_{k,d-\hat{d}}\varepsilon_{t-k} - \varepsilon_t)\mathbf{1}\{t \ge 1\} \\ &= (\textstyle\sum_{k=1}^\infty \pi_{k,d-\hat{d}}\varepsilon_{t-k})\mathbf{1}\{t \ge 1\}, \end{aligned} \quad \text{(A.11)}$$

where the last equality holds because $\pi_{0,\hat{d}-d} = 1$. Using (A.11), we have

$$\begin{aligned} T^{d-3/2}\textstyle\sum_{t=1}^T\mu_{1,t}(u_t^{\hat{d}}-\varepsilon_t) &= T^{d-3/2}\textstyle\sum_{t=1}^T\mu_{1,t}(\Delta^{\hat{d}-d}\varepsilon_t\mathbf{1}\{t\ge 1\}-\varepsilon_t) \\ &= T^{d-3/2}\textstyle\sum_{t=1}^T\mu_{1,t}(\sum_{k=1}^{t-1}\pi_{k,d-\hat{d}}\varepsilon_{t-k}) \\ &= T^{d-3/2}\textstyle\sum_{k=1}^{T-1}\pi_{k,d-\hat{d}}(\sum_{t=k+1}^T\mu_{1,t}\varepsilon_{t-k}) \\ &\le (\textstyle\sum_{k=1}^\infty|\pi_{k,d-\hat{d}}|)(\sup_{s\in\mathbb{N}}|T^{d-3/2}\sum_{t=1}^T\mu_{1,t}\varepsilon_{t-s}|) = o_p(1), \end{aligned} \quad \text{(A.12)}$$

where the last equality holds due to Lemmas A.4 and A.5. We need to consider two other terms. In particular, $\hat{\mu}_{1,t}-\mu_{1,t} = \Delta^{\hat{d}}t\mathbf{1}\{t \ge 1\} - \Delta^d t\mathbf{1}\{t \ge 1\} = (\Delta^{\hat{d}-d}-1)\Delta^d t\mathbf{1}\{t \ge 1\} = (\Delta^{\hat{d}-d}-1)\,\mu_{1,t}$ and $\Delta^{\hat{d}-d} = (1-L)^{\hat{d}-d} = \sum_{k=0}^\infty \pi_{k,d-\hat{d}}L^k$, thereby

$$\Delta^{\hat{d}-d} - 1 = \textstyle\sum_{k=1}^\infty\pi_{k,d-\hat{d}}L^k. \quad \text{(A.13)}$$

Using (A.13), we have

$$\begin{aligned} T^{d-3/2}\sum_{t=1}^{T}(\hat{\mu}_{1,t}-\mu_{1,t})\varepsilon_t &= T^{d-3/2}\sum_{t=1}^{T}(\Delta^{\hat{d}-d}-1)\mu_{1,t}\varepsilon_t = T^{d-3/2}\sum_{t=1}^{T}(\sum_{k=1}^{\infty}\pi_{k,d-\hat{d}}\mu_{1,t-k}\,\varepsilon_{t-k}) \\ &= T^{d-3/2}[\sum_{t=1}^{T}\pi_{1,d-\hat{d}}\mu_{1,t-1}\,\varepsilon_{t-1}+\sum_{t=1}^{T}\pi_{2,d-\hat{d}}\mu_{1,t-2}\,\varepsilon_{t-2}+\cdots] \\ &\leq (\sum_{k=1}^{\infty}|\pi_{k,d-\hat{d}}|)(\sup_{s\in\mathbb{N}}|T^{d-3/2}\sum_{t=1}^{T}\mu_{1,t-s}\varepsilon_{t-s}|) = o_p(1), \end{aligned}$$

where the last equality holds due to Lemmas A.4 and A.5. Finally, the proof of $T^{d-3/2}\sum_{t=1}^{T}(\hat{\mu}_{1,t}-\mu_{1,t})(u_t^{\hat{d}}-\varepsilon_t)=o_p(1)$ follows from the previous arguments. We still need to consider the parts of (A.5) and (A.7) that pertain to $\mu_{0,t}$ and $\hat{\mu}_{0,t}$. First, for $-0.5<d<0.5$, we need to show that

$$T^{d-1/2}\sum_{t=1}^{T}\hat{\mu}_{0,t}u_t^{\hat{d}} - T^{d-1/2}\sum_{t=1}^{T}\mu_{0,t}\varepsilon_t \xrightarrow{d} 0,$$

or equivalently,

$$T^{d-1/2}[\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})\varepsilon_t+\sum_{t=1}^{T}\mu_{0,t}(u_t^{\hat{d}}-\varepsilon_t)+\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})(u_t^{\hat{d}}-\varepsilon_t)] \xrightarrow{d} 0.$$

Using Lemmas A.4-A.5, we have

$$\begin{aligned} T^{d-1/2}\sum_{t=1}^{T}\mu_{0,t}(u_t^{\hat{d}}-\varepsilon_t) &= T^{d-1/2}\sum_{t=1}^{T}\mu_{0,t}(\sum_{k=1}^{\infty}\pi_{k,d-\hat{d}}\varepsilon_{t-k}) \\ &\leq (\sum_{k=1}^{\infty}|\pi_{k,d-\hat{d}}|)(\sup_{s\in\mathbb{N}}|T^{d-1/2}\sum_{t=1}^{T}\mu_{0,t}\varepsilon_{t-s}|) = o_p(1), \end{aligned}$$

and

$$\begin{aligned} T^{d-1/2}\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})\,\varepsilon_t &= T^{d-1/2}\sum_{t=2}^{T}(\sum_{k=1}^{t-1}\pi_{k,\hat{d}-d}\mu_{0,t-k}\varepsilon_{t-k}) \\ &\leq (\sum_{k=1}^{\infty}|\pi_{k,d-\hat{d}}|)(\sup_{s\in\mathbb{N}}|T^{d-1/2}\sum_{t=1}^{T}\mu_{0,t-s}\varepsilon_{t-s}|) = o_p(1). \end{aligned}$$

It is easy to show that $T^{d-1/2}\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})(u_t^{\hat{d}}-\varepsilon_t)=o_p(1)$ by combining the previous arguments. Second, for $0.5<d<1.5$, we show that $\sum_{t=1}^{T}\hat{\mu}_{0,t}\,u_t^{\hat{d}}-\sum_{t=1}^{T}\mu_{0,t}\,\varepsilon_t \xrightarrow{d} 0$, or equivalently,

$$\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})\,\varepsilon_t+\sum_{t=1}^{T}\mu_{0,t}(u_t^{\hat{d}}-\varepsilon_t)+\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})(u_t^{\hat{d}}-\varepsilon_t) \xrightarrow{d} 0.$$

Similar to the previous arguments for $d\in(-0.5,0.5)$, using Lemma A.5 and the fact that $\sum_{t=1}^{T}\mu_{0,t}\varepsilon_t=O_p(1)$ (Iacone et al., 2013, p. 41), we have

$$\begin{aligned} \sum_{t=1}^{T}\mu_{0,t}(u_t^{\hat{d}}-\varepsilon_t) &= \sum_{t=1}^{T}\mu_{0,t}(\sum_{k=1}^{\infty}\pi_{k,d-\hat{d}}\,\varepsilon_{t-k}) \\ &\leq (\sum_{k=1}^{\infty}|\pi_{k,d-\hat{d}}|)(\sup_{s\in\mathbb{N}}|\sum_{t=1}^{T}\mu_{0,t}\varepsilon_{t-s}|) = o_p(1), \end{aligned}$$

and

$$\begin{aligned} \sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})\,\varepsilon_t &= \sum_{t=1}^{T}(\sum_{k=1}^{\infty}\pi_{k,d-\hat{d}}\mu_{0,t-k}\varepsilon_{t-k}) \\ &\leq (\sum_{k=1}^{\infty}|\pi_{k,d-\hat{d}}|)(\sup_{s\in\mathbb{N}}|\sum_{t=1}^{T}\mu_{0,t-s}\varepsilon_{t-s}|) = o_p(1). \end{aligned}$$

It is straightforward to show that $\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})(u_t^{\hat{d}}-\varepsilon_t)=o_p(1)$ by combining the previous arguments. From Theorem 4 of Shimotsu (2010), under Assumptions 2-3: $\hat{d}-d=O_p(m^{-1/2})$ with $d\in(\Delta_1,\Delta_2)$ and $-1/2<\Delta_1<\Delta_2\leq 7/4$. Moreover, the long-run variance can be estimated consistently, i.e., $\hat{\sigma}^2-\sigma^2=o_p(1)$ (Andrews, 1991). Therefore, $t_{\hat{\beta}_i}(\hat{d},\hat{\sigma}^2)-t_{\beta_i}(d,\sigma^2)\xrightarrow{d}0$, $i=\{1,2\}$, is satisfied with a bandwidth $m=[C\,T^{2\kappa}]$ for any $\kappa>0$.

Proof of Theorem 2. Under the alternative hypothesis, the DGP is specified as

$$y_t=\beta_1^1+\beta_2^1 t+u_t,\quad t=1,\ldots,T.$$

We can use the arguments in the proof of Theorem 1. For part (i), we have

$$\begin{aligned}t_{\hat{\beta}_2}(d,\sigma^2)&=\frac{R(\hat{\beta}-\beta^0)}{[\sigma^2R(X^{d\prime}X^d)^{-1}R']^{1/2}}\\&=\frac{R(\hat{\beta}-\beta^1+\beta^1-\beta^0)}{[\sigma^2R(X^{d\prime}X^d)^{-1}R']^{1/2}}\\&=\frac{R(\hat{\beta}-\beta^1)}{[\sigma^2R(X^{d\prime}X^d)^{-1}R']^{1/2}}+\frac{R(\beta^1-\beta^0)}{[\sigma^2R(X^{d\prime}X^d)^{-1}R']^{1/2}}.\end{aligned}$$

For $d\in(-0.5,0.5)$, using $K_T:=diag\{T^{1/2-d},T^{3/2-d}\}$, the t-statistic for β_2 follows:

$$\begin{aligned}t_{\hat{\beta}_2}(d,\sigma^2)&=\frac{R(K_T^{-1}X^{d\prime}X^dK_T^{-1})^{-1}(K_T^{-1}X^{d\prime}\varepsilon)}{[\sigma^2R(K_T^{-1}X^{d\prime}X^dK_T^{-1})^{-1}R']^{1/2}}+\frac{RK_T(\beta^1-\beta^0)}{[\sigma^2R(K_T^{-1}X^{d\prime}X^dK_T^{-1})^{-1}R']^{1/2}}\\&=O_p(1)+O_p(T^{3/2-d})=O_p(T^{3/2-d}),\end{aligned}\tag{A.14}$$

where the first term in (A.14) converges in distribution to $(RC^{-1}L)/[RC^{-1}R']^{1/2}$. For $d\in(0.5,1.5)$, using $\widetilde{K}_T:=diag\{1,T^{3/2-d}\}$, the t-statistic for β_2 follows:

$$\begin{aligned}t_{\hat{\beta}_2}(d,\sigma^2)&=\frac{R(\widetilde{K}_T^{-1}X^{d\prime}X^d\widetilde{K}_T^{-1})^{-1}(\widetilde{K}_T^{-1}X^{d\prime}\varepsilon)}{[\sigma^2R(\widetilde{K}_T^{-1}X^{d\prime}X^d\widetilde{K}_T^{-1})^{-1}R']^{1/2}}+\frac{R\widetilde{K}_T(\beta^1-\beta^0)}{[\sigma^2R(\widetilde{K}_T^{-1}X^{d\prime}X^d\widetilde{K}_T^{-1})^{-1}R']^{1/2}}\\&=O_p(T^{3/2-d}).\end{aligned}$$

Now consider part (ii). For $d\in(-0.5,0.5)$,

$$\begin{aligned}t_{\hat{\beta}_1}(d,\sigma^2)&=\frac{R_1(K_T^{-1}X^{d\prime}X^dK_T^{-1})^{-1}(K_T^{-1}X^{d\prime}\varepsilon)}{[\sigma^2R_1(K_T^{-1}X^{d\prime}X^dK_T^{-1})^{-1}R_1']^{1/2}}+\frac{R_1K_T(\beta^1-\beta^0)}{[\sigma^2R_1(K_T^{-1}X^{d\prime}X^dK_T^{-1})^{-1}R_1']^{1/2}}\\&=O_p(T^{1/2-d}),\end{aligned}\tag{A.15}$$

where the first term in (A.15) converges in distribution to $(R_1C^{-1}L)/[R_1C^{-1}R_1']^{1/2}$. For $d\in(0.5,1.5)$, it is straightforward to show that

$$\begin{aligned}t_{\hat{\beta}_1}(d,\sigma^2)&=\frac{R_1(\widetilde{K}_T^{-1}X^{d\prime}X^d\widetilde{K}_T^{-1})^{-1}(\widetilde{K}_T^{-1}X^{d\prime}\varepsilon)}{[\sigma^2R_1(\widetilde{K}_T^{-1}X^{d\prime}X^d\widetilde{K}_T^{-1})^{-1}R_1']^{1/2}}+\frac{R_1\widetilde{K}_T(\beta^1-\beta^0)}{[\sigma^2R_1(\widetilde{K}_T^{-1}X^{d\prime}X^d\widetilde{K}_T^{-1})^{-1}R_1']^{1/2}}\\&=O_p(1).\end{aligned}$$

Proof of Theorem 3. We want to show the weak convergence

$$T^{d-3/2}\textstyle\sum_{t=1}^{T}\hat{\mu}_t(\tau)u_t^{\hat{d}} - T^{d-3/2}\textstyle\sum_{t=1}^{T}\mu_t(\tau)\varepsilon_t \Rightarrow 0 \tag{A.16}$$

uniformly in τ, where $\mu_t(\tau) = \Delta^d(t-[\tau T])\mathbf{1}\{t > [\tau T]\}$, and "$\Rightarrow$" denotes weak convergence in Skorohod topology. This is equivalent to showing that the following three terms weakly converge to 0 uniformly in τ:

$$T^{d-3/2}\textstyle\sum_{t=1}^{T}(\hat{\mu}_t(\tau) - \mu_t(\tau))\varepsilon_t \tag{A.17}$$

$$+\, T^{d-3/2}\textstyle\sum_{t=1}^{T}\mu_t(\tau)(u_t^{\hat{d}} - \varepsilon_t) \tag{A.18}$$

$$+\, T^{d-3/2}\textstyle\sum_{t=1}^{T}(\hat{\mu}_t(\tau) - \mu_t(\tau))(u_t^{\hat{d}} - \varepsilon_t). \tag{A.19}$$

Consider the term (A.18) first. Using (A.11), we have, with ε_t is a short memory process

$$\begin{aligned}
&\sup_{\tau\in(0,1)} |T^{d-3/2}\textstyle\sum_{t=1}^{T}\mu_t(\tau)(u_t^{\hat{d}} - \varepsilon_t)| \\
&= \sup_{\tau\in(0,1)} |T^{d-3/2}\textstyle\sum_{t=1}^{T}\mu_t(\tau)(\Delta^{\hat{d}-d}\varepsilon_t\mathbf{1}\{t \geq 1\} - \varepsilon_t)| \\
&= \sup_{\tau\in(0,1)} |T^{d-3/2}\textstyle\sum_{t=1}^{T}\mu_t(\tau)(\sum_{k=1}^{t-1}\pi_{k,d-\hat{d}}\varepsilon_{t-k})| \\
&= \sup_{\tau\in(0,1)} |T^{d-3/2}\textstyle\sum_{k=1}^{T-1}\pi_{k,d-\hat{d}}(\sum_{t=k+1}^{T}\mu_t(\tau)\varepsilon_{t-k})| \\
&\leq (\textstyle\sum_{k=1}^{\infty}|\pi_{k,d-\hat{d}}|)\sup_{\tau\in(0,1)}\{\sup_{s\in\mathbb{N}}|T^{d-3/2}\textstyle\sum_{t=1}^{T}\mu_t(\tau)\,\varepsilon_{t-s}|\} \Rightarrow 0,
\end{aligned}$$

where the last equality holds due to Lemmas A.5 and A.6. We next consider the term (A.17). Because

$$\begin{aligned}
\hat{\mu}_t(\tau) - \mu_t(\tau) &= \Delta^{\hat{d}}(t-[\tau T])\mathbf{1}\{t > [\tau T]\} - \Delta^d(t-[\tau T])\mathbf{1}\{t > [\tau T]\} \\
&= (\Delta^{\hat{d}-d} - 1)\Delta^d(t-[\tau T])\mathbf{1}\{t > [\tau T]\} = (\Delta^{\hat{d}-d} - 1)\mu_t(\tau),
\end{aligned}$$

we have

$$\begin{aligned}
&\sup_{\tau\in(0,1)} |T^{d-3/2}\textstyle\sum_{t=1}^{T}(\hat{\mu}_t(\tau) - \mu_t(\tau))\varepsilon_t| = \sup_{\tau\in(0,1)} |T^{d-3/2}\textstyle\sum_{t=1}^{T}(\Delta^{\hat{d}-d} - 1)\mu_t(\tau)\varepsilon_t| \\
&= \sup_{\tau\in(0,1)} |T^{d-3/2}\textstyle\sum_{t=1}^{T}(\sum_{k=1}^{\infty}\pi_{k,d-\hat{d}}\mu_{t-k}(\tau)\varepsilon_{t-k})| \\
&\leq \sup_{\tau\in(0,1)}\{(\textstyle\sum_{k=1}^{\infty}|\pi_{k,d-\hat{d}}|)\sup_{s\in\mathbb{N}}|T^{d-3/2}\textstyle\sum_{t=1}^{T}\mu_{t-s}(\tau)\varepsilon_{t-s}|\} \Rightarrow 0,
\end{aligned}$$

where the last equality holds due to Lemmas A.5 and A.6. Finally, (A.19) also weakly converges to 0 uniformly in τ using similar arguments.

Table 1: Finite Sample Size; Pure Fractional Processes

T		d					
		-0.4	0.2	0.4	0.8	1	1.4
500	$t_{\hat{\beta}_2}$	0.054	0.056	0.057	0.047	0.062	0.051
	$t_{\beta}^{FS}(MU)$	0.000	0.098	0.157	0.155	0.053	0.462
	$t_{\beta}^{FS}(UB)$	0.000	0.093	0.139	0.100	0.052	0.462
1000	$t_{\hat{\beta}_2}$	0.047	0.051	0.059	0.034	0.043	0.049
	$t_{\beta}^{FS}(MU)$	0.000	0.138	0.173	0.134	0.049	0.495
	$t_{\beta}^{FS}(UB)$	0.000	0.138	0.163	0.099	0.049	0.495
2000	$t_{\hat{\beta}_2}$	0.054	0.047	0.044	0.051	0.043	0.039
	$t_{\beta}^{FS}(MU)$	0.000	0.178	0.277	0.108	0.046	0.559
	$t_{\beta}^{FS}(UB)$	0.000	0.178	0.277	0.097	0.046	0.559

Note: $t_{\hat{\beta}_2}$ is the slope test proposed in this paper; $t_{\beta}^{FS}(MU)$ and $t_{\beta}^{FS}(UB)$ are the Perron and Yabu (2009) tests using a Median Unbiased (MU) and Upper Biased (UB) correction, respectively.

Table 2: Finite Sample Size; AR(1) Processes with $d = 0$

T		AR					
		0	0.3	0.5	0.7	0.9	0.95
500	$t_{\hat{\beta}_2}$	0.074	0.068	0.049	0.018	0.017	0.005
	$t_{\beta}^{FS}(MU)$	0.051	0.059	0.030	0.037	0.049	0.045
	$t_{\beta}^{FS}(UB)$	0.051	0.059	0.030	0.037	0.047	0.034
1000	$t_{\hat{\beta}_2}$	0.085	0.068	0.057	0.008	0.017	0.005
	$t_{\beta}^{FS}(MU)$	0.057	0.031	0.046	0.046	0.042	0.053
	$t_{\beta}^{FS}(UB)$	0.057	0.031	0.046	0.046	0.042	0.051
2000	$t_{\hat{\beta}_2}$	0.064	0.067	0.069	0.017	0.029	0.005
	$t_{\beta}^{FS}(MU)$	0.066	0.045	0.058	0.041	0.044	0.049
	$t_{\beta}^{FS}(UB)$	0.066	0.045	0.058	0.041	0.044	0.049

Note: $t_{\hat{\beta}_2}$ is the slope test proposed in this paper; $t_{\beta}^{FS}(MU)$ and $t_{\beta}^{FS}(UB)$ are the Perron and Yabu (2009) tests using a Median Unbiased (MU) and Upper Biased (UB) correction, respectively.

Table 3: Finite Sample Size of $t_{\hat{\beta}_2}$; DGPs 1-6 with $d = 0.4$

T	DGP-1 AR=0.4	DGP-1 AR=-0.4	DGP-2 MA=0.4	DGP-2 MA=-0.4	DGP-3 AR_1=0.3 AR_2=0.5	DGP-4 Measurement error	DGP-5 GARCH	DGP-6 Heavy-tailed
500	0.061	0.059	0.084	0.107	0.033	0.052	0.084	0.056
1000	0.057	0.058	0.076	0.094	0.056	0.031	0.074	0.044
2000	0.048	0.065	0.080	0.069	0.045	0.052	0.069	0.046

Table 4: Results of Tests Applied to Equity Indices

	$t_{\hat{\beta}_2}$	$t_{\beta}^{FS}(MU)$	$t_{\beta}^{FS}(UB)$	$\hat{d}_{ELW}$	ADF
S&P 500 index	3.314***	3.396***	3.396***	0.946	-3.430^{**} (0)
NASDAQ composite index	3.949***	3.809***	3.809***	0.941	-3.028 (0)
Dow Jones Industrial Average	2.912***	2.196**	2.196**	0.934	-4.064^{***} (0)
Wilshire 5000 Total Market Index	3.296***	3.575***	3.575***	0.960	-3.263^{*} (0)

Note: $t_{\hat{\beta}_1}$ and $t_{\hat{\beta}_2}$ are the intercept and slope test proposed in this paper; $t_{\beta}^{FS}(MU)$ and $t_{\beta}^{FS}(UB)$ are the Perron and Yabu (2009) tests using a Median Unbiased (MU) and Upper Biased (UB) correction, respectively. Test rejected at *10% level, **5% level, ***1% level. The numbers in parentheses are the values of the autoregressive order selected by the Bayesian information criterion when constructing the ADF test.

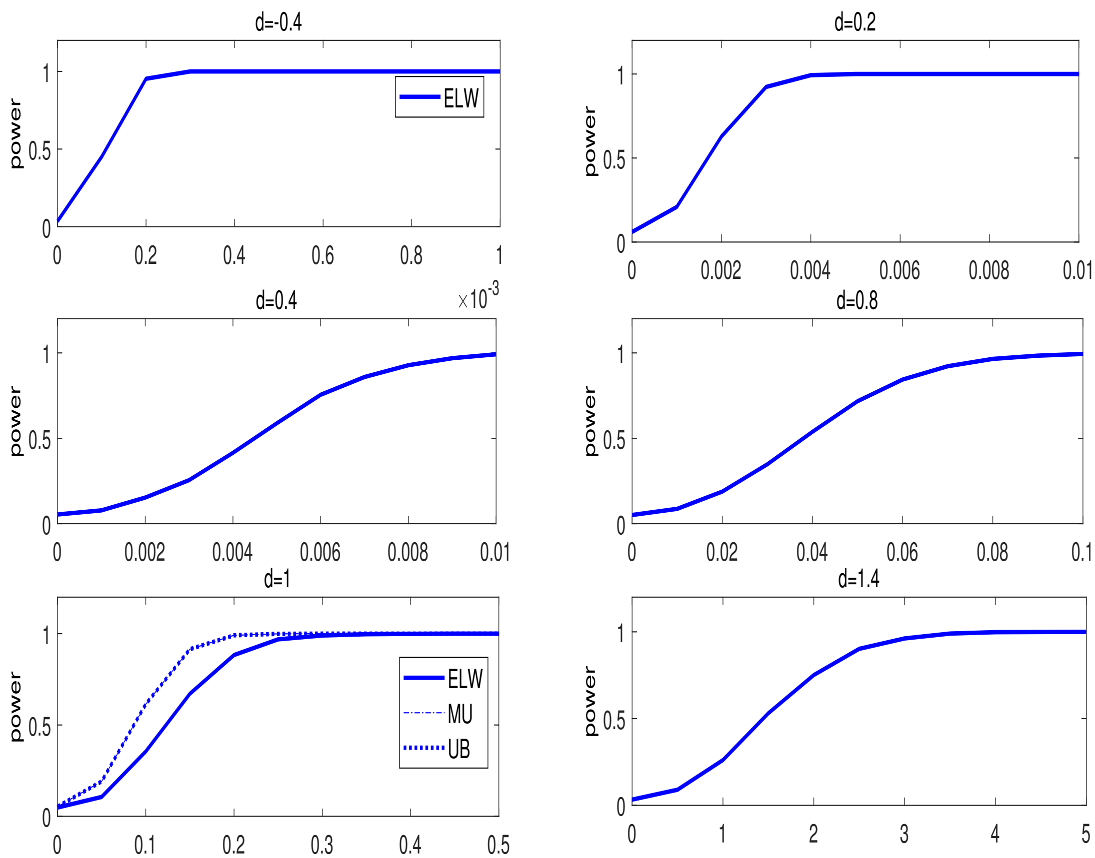


Figure 1: Unadjusted power for pure fractional processes

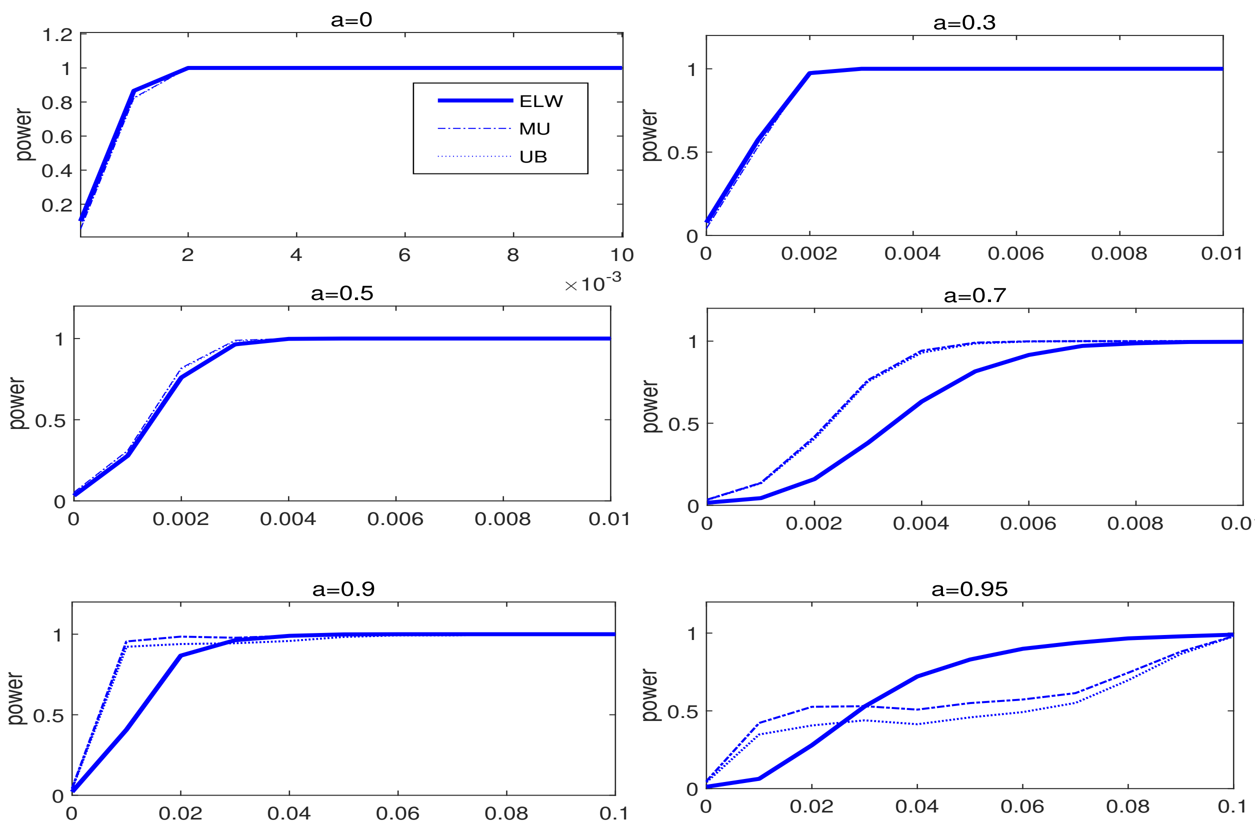


Figure 2: Unadjusted power for AR(1) processes

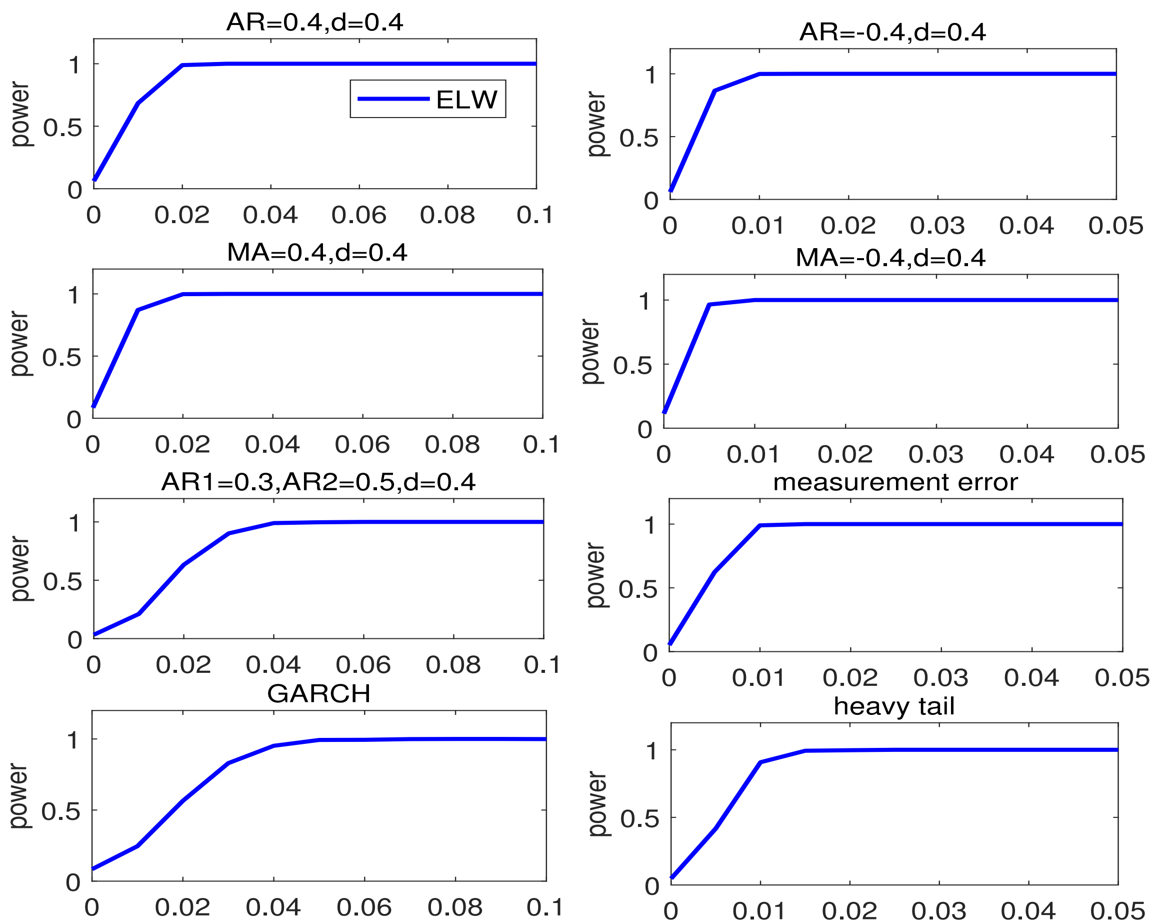
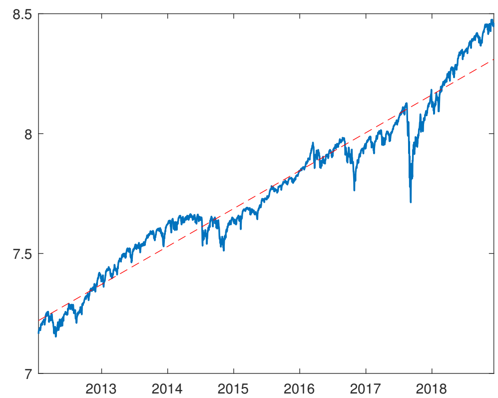
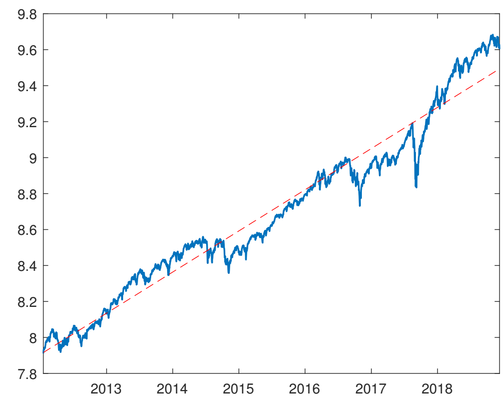


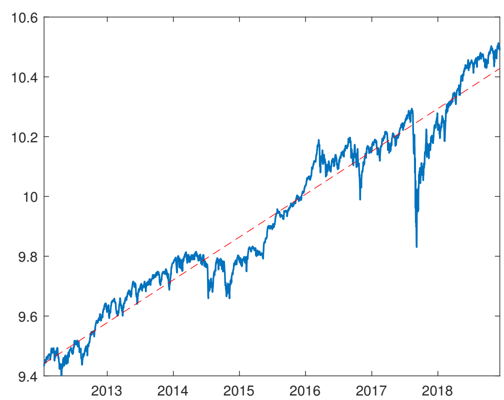
Figure 3: Unadjusted power for DGPs 1-6



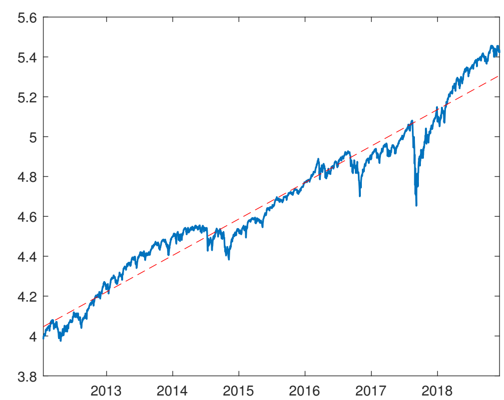
(a) S&P 500



(b) Nasdaq

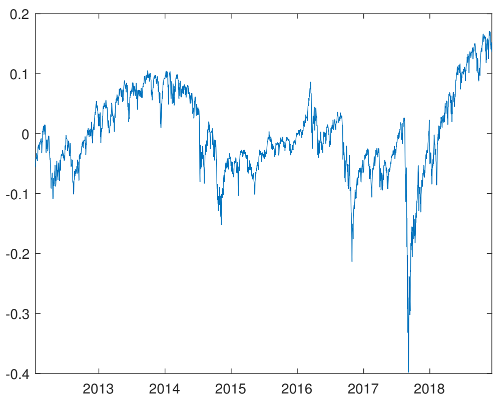


(c) Dow Jones

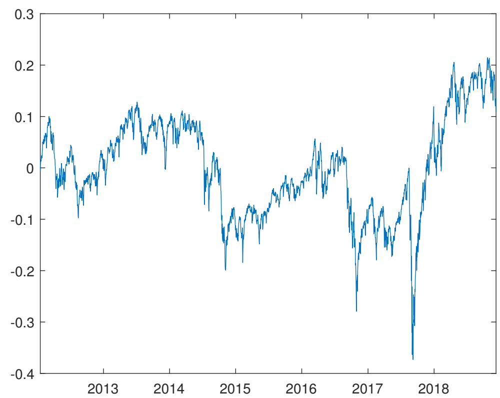


(d) Wilshire 5000

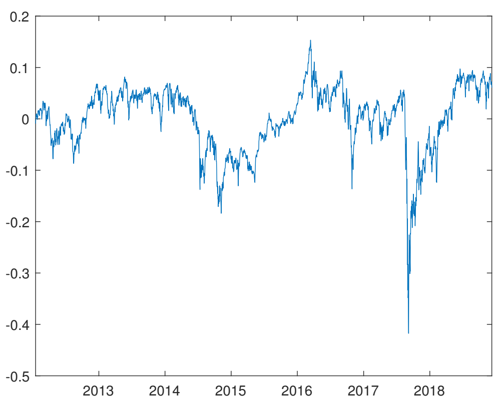
Figure 4: Trend fitted to equity indices



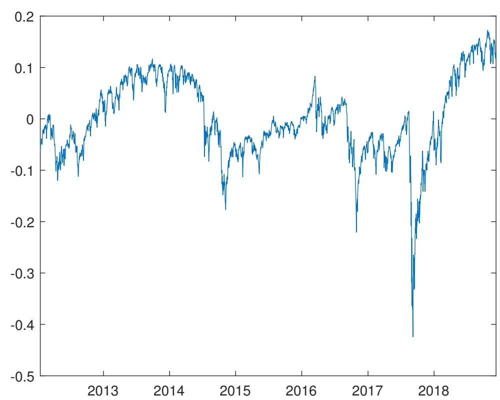
(a) S&P 500



(b) Nasdaq



(c) Dow Jones



(d) Wilshire 5000

Figure 5: Detrended equity indices