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

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

Xiamen University **Boston University** **SUFE**

January 13, 2018

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 $d \in (-0.5, 1.5)$ by applying a quasi-GLS procedure using d-differences of the data. Doing so, the asymptotic distribution of the OLS 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 of Shimotsu (2010), valid for processes with a linear trend. The finite sample size and power of the tests are investigated via simulations. We also provide a comparison with the tests of Perron and Yabu (2009) valid for a noise component that is $I(0)$ or $I(1)$. The results are encouraging in that our test is valid under more general conditions, yet has similar power as those that apply to the dichotomous cases with d either 0 or 1. We also use our method of proof to show that the main result of Iacone, Leybourne and Taylor (2013), who considered testing for a break in the slope of a trend function with a fractionally integrated noise, is valid for the full range $d \in (-.5, 1.5)$.

JEL Classification: C22

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

*We are grateful to two referees, the editor and Zhongjun Qu for useful comments. One referee, in particular, provided detailed comments to make the proofs more rigorous.

[†]Wang Yanan Institute for Studies in Economics & Department of Statistics, Xiamen University, 422 Ximing South Road, Xiamen, China 361005, and MOE Key Laboratory of Econometrics, and Fujian Key Laboratory of Statistical Science (sychang@xmu.edu.cn).

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

[§]School of Economics, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai, China 200433, and Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai 200433, China (xu.jiawen@mail.shufe.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:

$$y_t = \beta_1 + \beta_2 t + u_t, \tag{1}$$

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 (e.g., Perron, 1988). Third, tests for hypotheses about the values of β_1 and β_2 allow constructing confidence intervals via inversions. There is a large literature on issues pertaining to inference about the slope of a linear trend function, most related to the case where the noise component is stationary, i.e., integrated of order zero, $I(0)$. A classic result due to Grenander and Rosenblatt (1957) states that the estimate of β_2 obtained from a simple least-squares regression of the form (1) is asymptotically as efficient as that obtained from a Generalized Least Squares (GLS) regression 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 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 test of the unit root hypothesis 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 the cases where 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 superefficient estimate of the sum of the autoregressive parameters α when $\alpha = 1$. The estimate of α is the OLS estimate 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 long-memory process extends the classical dichotomy of $I(0)$ and $I(1)$ processes. Our aim is to provide tests to perform inference on the coefficients 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 OLS procedures can be applied with the limit normal 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 $(-0.5, 1.5)$ and yields tests with good finite sample properties. Of related interest, Abadir et al. (2011) considered an $I(d)$ model with trend 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 the full range $d \in (-.5, 1.5)$.

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) 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 similar power. Section 5 considers generalyzing the main result of Iacone et al. (2013). Section 6 provides brief conclusions and techical derivations are collected in an appendix.

2 The Model and Test Statistics

The data-generating process is assumed to be:

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

for $t = 1, ..., T$, with u_t a fractionally integrated process satisfying the following assumptions.

Assumption 1 *The process u_t is generated by $\Delta^d u_t = (1-L)^d u_t = \varepsilon_t \P\{t \geq 1\}$ with $d \in (-0.5, 0.5) \cup (0.5, 1.5)$, where Δ^d is the fractional difference operator and $\P\{A\}$ 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 $f_\varepsilon(\lambda) \sim G$ for $\lambda \sim 0$.

Assumption 2 *$\{\gamma_j\}_{j=-\infty}^{\infty}$ is the sequence of autocovariances of ε_t; then (i) $f_\varepsilon(\lambda) \sim G_0 \in (0, \infty)$ and, for some $\beta \in (0, 2]$, $f_\varepsilon(\lambda) = G_0(1 + O(\lambda^\beta))$ 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 mostly 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 a functional central limit theorem for the partial sums of u_t. Applying a d-differencing transformation, the DGP is:

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

Note that $\Delta^d u_t = \varepsilon_t$ and $\Delta^d y_1 = y_1$. We also define $X_t = [1, t]'$ and $X_t^d \equiv \Delta^d X_t = [\Delta^d \P\{t \geq 1\}, \Delta^d t \P\{t \geq 1\}]'$ with $\Delta^d X_1 = [1, 1]'$. Hence, the GLS transformed regression is:

$$y_t^d = X_t^{d'} \beta + \varepsilon_t, \qquad (t = 1, ..., T).$$

To obtain a feasible regression, we need to replace d by some consistent estimate $\hat{d}$ to be discussed in the next section. The tests will then be based on the regression

$$y_t^{\hat{d}} = X_t^{\hat{d}'} \beta + u_t^{\hat{d}}, \qquad (t = 1, ..., T) \tag{3}$$

[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 and Iacone et al, 2013.)

where $u^{\hat{d}} = \Delta^{\hat{d}} u_t \P\{t \geq 1\}$. Let $\hat{\beta} = (X^{\hat{d}\prime} X^{\hat{d}})^{-1} X^{\hat{d}\prime} y^{\hat{d}}$ denote the OLS estimator of $\beta = [\beta_1, \beta_2]'$, 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:

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

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 $H_0 : \beta_1 = \beta_1^0$ can also be constructed as usual with:

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

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. Let "$\xrightarrow{d}$" denote weak convergence in distribution under the Skorohod topology. Suppose that we have estimated $\hat{d}$ and $\hat{\sigma}^2$ such that $\hat{d} - d = O_p(T^{-\kappa})$ for any $\kappa > 0$ and $\hat{\sigma}^2 - \sigma^2 = o_p(1)$. Then, (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 $d \in (-0.5, 0.5)$.*

Remark 1 *Iacone et al. (2013) considered the Fully Extended Local Whittle (FELW) estimator of Abadir, Distaso, and Giraitis (2007) to construct the test statistic 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 user chosen constant $c_1 > 0$ satisfy a condition that $0 < n < 0.8$ (see Corollary 2.1 of Abadir et al., 2007), where $[x]$ denotes the integer part of x. As addressed in Remark 3 of Iacone et al. (2013), this requirement is somewhat restrictive because it imposes $d < 1.33$ with $n = 0.65$ and $d < 1.40$ with $n = 0.79$. Unlike Iacone et al. (2013), we only require that $\hat{d}$ be consistent at any polynomial rate for all values of d. 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.*

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 $h(\cdot)$ 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 estimation procedure for d was studied by Shimotsu and Phillips (2005) and extended by Shimotsu (2010) for the case with an unknown trend function, needed 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

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

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 by an OLS regression of y_t on $(1, t)$ with the residuals denoted $\hat{y}_t$. The modified objective function is:

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

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:

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

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 $T \to \infty$, $m^{-1} + m^{1+2\beta}(\log m)^2 T^{-2\beta} + m^{-\gamma} \log T \to 0$ 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

In this section, we consider the size and power of the test $t_{\hat{\beta}_2}$ for the slope of the trend via simulations, using 1,000 replications throughout (the results for the test $t_{\hat{\beta}_1}$ for the mean are qualitatively similar for the range $d \in (-0.5, 0.5)$). The data are generated by (2) with u_t an autoregressive fractionally integrated moving average process (ARFIMA(p, d, q)) of the form $(1 - L)^d u_t = \varepsilon_t \P\{t \geq 1\}$ 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, respectively, 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 exact sizes of the test $t_{\hat{\beta}_2}$ are 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 $d = 0, 1$. When d is negative the tests are very conservative, while for $0 < d < 1$, the tests are liberal. The liberal size distortions 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 different from 0 and 1, we include them only for the case $d = 1$ (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. This is due to the fact that the PY tests apply a truncation to 1 when the autoregressive parameter is in a neighborhood of 1 leading to a smaller bias when $d = 1$. However, the differences are not large and decrease as T increases (from unreported simulations). As expected, the power of $t_{\hat{\beta}_2}$ is highest when d is small with power decreasing as d increases (note the different scaling on the horizontal axis).

Table 2 presents results about 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 exact size remains close to the nominal 5% level, unless a is close to 1, in which case the exact size of $t_{\hat{\beta}_2}$ is below nominal size. It is well known that in the presence of a short-run component that has 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 have slightly higher power. When $a = 0.95$, $t_{\hat{\beta}_2}$ has much higher power, despite being conservative, unless the alternative is close to the null value.

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. These are given by:

DGP 1. ARFIMA$(1, d, 0)$: $(1 - a_1 L)(1 - L)^{0.4}\varepsilon_t = e_t$, where $a_1 = 0.4$ and -0.4.

DGP 2. ARFIMA$(0, d, 1)$: $(1 - L)^{0.4}\varepsilon_t = (1 + b_1 L)e_t$, where $b_1 = 0.4$ and -0.4.

DGP 3. ARFIMA$(2, d, 0)$: $(1 - a_1 L)(1 - a_2 L)(1 - L)^{0.4}\varepsilon_t = e_t$, 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, $e_t \sim i.i.d.\ N(0, 1)$. DGPs 1-3 are different cases of ARFIMA processes, DGP 4 is a fractionally integrated process with measurement errors and DGP 5 is a generalized autoregressive conditional heteroskedasticity (GARCH) process. Note that DGPs 4 and 5 do not satisfy the conditions of Assumptions 1-2. We nevertheless include them to assess the robustness of the results given that conditional heteroskedasticity and measurement errors 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 presents the exact sizes of the tests. In all cases, the exact size of $t_{\hat{\beta}_2}$ 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, with except perhaps with GARCH errors. Comparing across DGPs, power decreases when additional short-run dynamics is present. The effect of measurement errors on the power is minor.

5 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 $DT_t(\tau_0) := (t - [\tau_0 T])\P\{t > [\tau_0 T]\}$, τ_0 unknown and u_t satifying 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 $d \in [0, 1/2) \cup (1/2, 3/2)$. 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, show the results to hold for any $n > 0$.

Theorem 2 (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 user chosen 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,*

$$\mathcal{W}(\hat{d}, \tau, \hat{\sigma}^2) - \mathcal{W}(d, \tau, \sigma^2) = o_p(1), \; \mathcal{SW}(\hat{d}, \hat{\sigma}^2) - \mathcal{SW}(d, \sigma^2) = o_p(1).$$

6 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 the sense that our test is valid under much more general conditions, yet has power similar to the Perron and Yabu (2009) tests that apply only to the dichotomous cases with d either 0 or 1. When d is different from 0 or 1, its exact 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 on 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 $d \in (-.5, 1.5)$ excluding $d = 0.5$.

References

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

Abadir, K. M., Distaso, W. and 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.

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

Giraitis, L., Koul, H. L., and Surgailis, D. (2012) *Large Sample Inference for Long Memory Processes*, London: Imperial College Press.

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

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

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

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

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

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

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

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

Sun, H. and 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. and 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 $X^d = [X_1^d, ..., X_T^d]'$, $X_t^d = [\mu_{0,t}\ \mu_{1,t}]'$, $\mu_{i,t} = \Delta^d t^i \P\{t \geq 1\}$ 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 start with the following lemmas, whose proofs follow Iacone et al. (2013, p.40) with appropriate modifications for the results to hold under the conditions of Theorem 1. We start with the following Lemma from Robinson (2005, Lemma 1, p.1824)

Lemma 1 *As $t \to \infty$ and for an m such that $d - 1 < m \leq d$ (i) For $d \in (-0.5, 1)$, $\Delta^d \P\{t \geq 1\} = \Gamma(1-d)^{-1} t^{-d} + O(t^{-d-1} + t^{-m-1}\P\{d > 0\})$; (ii) For $d \in (-0.5, 2)$, $\Delta^d t \P\{t \geq 1\} = \Gamma(2-d)^{-1} t^{1-d} + O(t^{-d} + t^{-m}\P\{d > 0\})$; (iii) For $d > -0.5$, $\Delta^d \P\{t \geq 1\} = O(t^{-d+\delta})$, for any $\delta > 0$.*

Note that Lemma 1 implies that for $d \in (-0.5, 1)$: $|\mu_{1,t+1} - \mu_{1,t}| = \Delta^d \P\{t \geq 1\} \leq \mathcal{C} t^{-d}$ and $\sum_{t=1}^{T-1} |\mu_{1,t+1} - \mu_{1,t}| = O(xx)$, from part (i), while for $d \in [1, 1.5)$: $|\mu_{1,t+1} - \mu_{1,t}| = \Delta^d \P\{t \geq 1\} \leq (xx)$ and $\sum_{t=1}^{T-1} |\mu_{1,t+1} - \mu_{1,t}| = O(T^{1-d+\delta})$ for any $\delta > 0$ These results will be used throughout.

Lemma 2 *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

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

(ii) For $0.5 \leq d < 1.5$: $t_{\hat{\beta}_2}(d, \sigma^2) \xrightarrow{d} R\widetilde{C}^{-1}\widetilde{L}/[R\widetilde{C}^{-1}R']^{1/2} = C_{22}^{-1/2} L_2 := A_2$, 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.2: From Robinson and Iacone (2005, A.34), for any $\tau \in (0, 1]$, with $[\tau T]$ the integer part of τT, we have (a) for $d \in (-0.5, 1)$: $T^d \Delta^d \P\{[\tau T] \geq 0\} \to \Gamma(1-d)^{-1}\tau^{-d}$; (b) for $d \in (-0.5, 1.5)$, $T^{d-1}\Delta^d[\tau T]\P\{[\tau T] \geq 0\} \to \Gamma(2-d)^{-1}\tau^{1-d}$. For part (i), using $K_T := diag\{T^{1/2-d}, T^{3/2-d}\}$, the t-statistic for β_2 follows:

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

where $K_T^{-1}X^{d\prime}X^d K_T^{-1} \to C$ and $K_T^{-1}X^{d\prime}\varepsilon \xrightarrow{d} \sigma L$ are proved in Robinson and Iacone (2005, 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 follows given that for $0.5 < d < 1.5$,

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

and for $d = 0.5$, the limit distribution of $\widetilde{K}_T^{-1} X^{d'} \varepsilon$ is the same, while

$$\left(\widetilde{K}_T^{-1} X^{d'} X^d \widetilde{K}_T^{-1} \right)^{-1} \to \begin{bmatrix} 0 & 0 \\ 0 & C_{22}^{-1} \end{bmatrix}.$$

Lemma 3 *Suppose that d and σ^2 are known: For $d \in (-0.5, 0.5)$: $t_{\hat{\beta}_1}(d, \sigma^2) \xrightarrow{d} R_1 C^{-1} L / [R_1 C^{-1} R_1']^{1/2} := B_1 t$, where $B_1 = \sqrt{1-2d}[2(1-d)\int_0^1 r^{-d} dW(r) - (3-2d)\int_0^1 r^{1-d} dW(r)]$.*

Proof of Lemma A.3: The proof follows using $R_1 = [1\ 0]$ instead of R in Lemma A.2.

Lemma 4 *A_1, A_2 and B_1, defined in Lemmas A.1-A.2 have a $N(0,1)$ distribution.*

Proof of Lemma A.4: 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.

Proof of Theorem 1: Here, we establish the limiting distributions of $t_{\hat{\beta}_l}(\hat{d}, \hat{\sigma}^2)$, $l = \{1, 2\}$ with consistent estimates $(\hat{d}, \hat{\sigma}^2)$. More specifically, we show 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 remain the same using a consistent estimate of σ^2, hence we concentrate on using an estimate of d. We need to show that if $\hat{d} - d = O_p(T^{-\varphi})$ for any $\varphi > 0$, then (a) for $-0.5 < d < 0.5$,

$$K_T^{-1} X^{\hat{d}'} X^{\hat{d}} K_T^{-1} - K_T^{-1} X^{d'} X^d K_T^{-1} \xrightarrow{d} 0, \tag{1}$$

and

$$K_T^{-1} X^{\hat{d}'} u^{\hat{d}} - K_T^{-1} X^{d'} \varepsilon \xrightarrow{d} 0, \tag{2}$$

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

$$\widetilde{K}_T^{-1} X^{\hat{d}'} X^{\hat{d}} \widetilde{K}_T^{-1} - \widetilde{K}_T^{-1} X^{d'} X^d \widetilde{K}_T^{-1} \xrightarrow{d} 0, \tag{3}$$

and

$$\widetilde{K}_T^{-1} X^{\hat{d}'} u^{\hat{d}} - \widetilde{K}_T^{-1} X^{d'} \varepsilon \xrightarrow{d} 0, \tag{4}$$

where $u_t^{\hat{d}} := \Delta^{\hat{d}} u_t \P\{t \geq 1\}$ and $\hat{\mu}_{i,t} := \Delta^{\hat{d}} t^i \P\{t \geq 1\}$ for $i = \{0, 1\}$. Consider first (1) and (3). For (1) with $-0.5 < d < 0.5$, we show that for $i, j = \{0, 1\}$, $T^{2d-1-i-j}(\sum_{t=1}^T \hat{\mu}_{i,t}\hat{\mu}_{j,t} - \sum_{t=1}^T \mu_{i,t}\mu_{j,t}) \xrightarrow{d} 0$,or equivalently,

$$T^{2d-1-i-j}[\sum_{t=1}^T(\hat{\mu}_{i,t} - \mu_{i,t})\mu_{j,t} + \sum_{t=1}^T \mu_{i,t}(\hat{\mu}_{j,t} - \mu_{j,t}) + \sum_{t=1}^T(\hat{\mu}_{i,t} - \mu_{i,t})(\hat{\mu}_{j,t} - \mu_{j,t})] \xrightarrow{d} 0. \quad (5)$$

Note that from Iacone et al. (2013, A.22), for $i \in \{0, 1\}$ and $d \leq 1$, and for $i = 1$ and $d > 1$,

$$\hat{\mu}_{i,t} - \mu_{i,t} = o_p(t^{i-d}). \quad (6)$$

By the Cauchy-Schwarz inequality,

$$T^{2d-1-i-j}\sum_{t=1}^T(\hat{\mu}_{i,t} - \mu_{i,t})\mu_{j,t} \leq T^{2d-1-i-j}[\sum_{t=1}^T(\hat{\mu}_{i,t} - \mu_{i,t})^2 \sum_{t=1}^T \mu_{j,t}^2]^{1/2}. \quad (7)$$

In the view of (6) and $|\mu_{j,t}| \leq \mathcal{C}t^{j-d}$, it is straightforward to show that the expression on the RHS of (7) is $o_p(T^{2d-1-i-j+i-d+1/2+j-d+1/2}) = o_p(1)$. Following the same arguments, we can show that the other terms in (5) are $o_p(1)$, which establishes (1). For (3) with $0.5 < d < 1.5$, 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 $d \in (-0.5, 0.5)$. Note that $\hat{\mu}_{0,t} - \mu_{0,t} = o_p(t^{-d})$ for $d \in (0.5, 1)$, which implies $\sum_{t=1}^T(\hat{\mu}_{0,t} - \mu_{0,t})^2 = o_p(T^{1-2d})$. For $d \in [1, 1.5)$, $\sum_{t=1}^T(\hat{\mu}_{0,t} - \mu_{0,t})^2 = o_p(1)$ using Robinson (2005, Lemma 4) and Iacone et al. (2013, Lemma A.2). Then, for $i = 0$, $j = 1$,

$$\begin{aligned}
T^{d-3/2}\sum_{t=1}^T(\hat{\mu}_{0,t} - \mu_{0,t})\mu_{1,t} &\leq T^{d-3/2}[\sum_{t=1}^T(\hat{\mu}_{0,t} - \mu_{0,t})^2 \sum_{t=1}^T \mu_{1,t}^2]^{1/2} \\
&= \begin{cases} o_p(T^{d-3/2+(1-2d+3-2d)/2}) = o_p(T^{-d+1/2}) = o_p(1) & \text{for } d \in (0.5, 1), \\ o_p(T^{d-3/2+(3-2d)/2}) = o_p(1) & \text{for } d \in [1, 1.5). \end{cases}
\end{aligned}$$

For $i = 1$, $j = 0$, using Lemma 1(iii),

$$\begin{aligned}
T^{d-3/2}\sum_{t=1}^T(\hat{\mu}_{1,t} - \mu_{1,t})\mu_{0,t} &\leq T^{d-3/2}[\sum_{t=1}^T(\hat{\mu}_{1,t} - \mu_{1,t})^2 \sum_{t=1}^T \mu_{0,t}^2]^{1/2} \\
&= \begin{cases} o_p\left(T^{d-3/2+(3-2d+1-2d)/2}\right) = o_p(T^{-d+1/2}) = o_p(1) & \text{for } d \in (0.5, 1), \\ o_p\left(T^{d-3/2+(3-2d+1-2d+2\delta)/2}\right) = o_p(T^{-d+1/2+\delta}) & \text{for } d \in [1, 1.5), \end{cases}
\end{aligned}$$

where $o_p(T^{-d+1/2+\delta}) = o_p(1)$ when $d > 1/2+\delta$, that is, $\delta \in (0, 0.5)$. Similar arguments apply when $i = j = 0$. The proofs for the other terms are similar, which completes the proof of (3). For (2) and (4), we first show that for $d \in (-0.5, 1.5)$,

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

or equivalently,

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

The following lemmas are useful for that purpose.

Lemma 5 *Under Assumption 1, for* $d \in (-0.5, 1.5)$, $T^{d-3/2}\sum_{t=1}^{T}\mu_{1,t}(u_t^{\hat{d}}-\varepsilon_t) \xrightarrow{p} 0$.

Proof of Lemma A.5: As Robinson and Hualde (2013, p.1758) and Iacone et al. (2013, p.44), let

$$a_s^{(r)}(\nu) := \frac{\partial^r \Delta_s^{(\nu)}}{\partial \nu^r}, \quad g^{(r)}(\varepsilon_t, \nu) := \sum_{s=1}^{t-1} a_s^{(r)}(\nu)\varepsilon_{t-s}$$

for any $\nu \geq 0$ where the process u_t is such that $u_t = \sum_{s=-\infty}^{t}\Delta_{t-s}^{(d)}\varepsilon_s\P\{s \geq 1\}$. Using Robinson and Hualde (2003, C.19 and C.20), we can rewrite

$$T^{d-3/2}\sum_{t=1}^{T}\mu_{1,t}(u_t^{\hat{d}}-\varepsilon_t) = T^{d-3/2}\sum_{t=1}^{T}\mu_{1,t}\left(\sum_{r=1}^{B-1}\frac{1}{r!}(d-\hat{d})^r g^{(r)}(\varepsilon_t, 0)\right) \tag{8}$$

$$+ T^{d-3/2}\sum_{t=1}^{T}\mu_{1,t}\frac{1}{B!}(d-\hat{d})^B g^{(B)}(\varepsilon_t, d-\hat{d}). \tag{9}$$

Iacone et al. (2013, p.44) showed that

$$(\ln T)^{-(r+1)}T^{-1/2}\sum_{t=1}^{[\tau T]} g^{(r)}(\varepsilon_t, 0) \xrightarrow{p} 0 \tag{10}$$

uniformly in τ. Using summation by parts, we have

$$\begin{aligned}
&T^{d-3/2}T^{-r\varphi}|\sum_{t=1}^{T}\mu_{1,t}g^{(r)}(\varepsilon_t, 0)| \\
&\leq T^{d-3/2}T^{-r\varphi}\left[\sum_{t=1}^{T-1}|\mu_{1,t+1}-\mu_{1,t}|\left|\sum_{s=t+1}^{T}g^{(r)}(\varepsilon_s, 0)\right| + \left|\mu_{1,1}\sum_{t=1}^{T}g^{(r)}(\varepsilon_t, 0)\right|\right] \\
&\leq T^{d-3/2}T^{-r\varphi}\left[\sup_t\left(\left|\sum_{s=t+1}^{T}g^{(r)}(\varepsilon_s, 0)\right|\right)\sum_{t=1}^{T-1}|\mu_{1,t+1}-\mu_{1,t}| + \left|\mu_{1,1}\sum_{t=1}^{T}g^{(r)}(\varepsilon_t, 0)\right|\right]. \qquad (11)
\end{aligned}$$

Note that $\mu_{1,1} = \Delta^d$ and $|\mu_{1,t+1} - \mu_{1,t}| = \Delta^d\P\{t \geq 1\}$. From Lemma 1 (i), $|\mu_{1,t+1} - \mu_{1,t}| = \Delta^d\P\{t \geq 1\} \leq \mathcal{C}t^{-d}$ and $\sum_{t=1}^{T-1} |\mu_{1,t+1} - \mu_{1,t}| = O(T^{1-d})$ for $d \in (-0.5, 1)$ (??check) From Lemma 1(iii), $\sum_{t=1}^{T-1} |\mu_{1,t+1} - \mu_{1,t}| = O(T^{1-d+\delta})$ for any $\delta > 0$. Using (10), the expression (11) is $o_p(T^{d-3/2-r\varphi}T^{1/2}(\ln T)^{r+1}T^{1-d+\delta})$ for any $\delta > 0$. This is the largest for $r = 1$, and it is $o_p(T^{-\varphi+\delta}(\ln T)^2)$, which is $o_p(1)$ when $-\varphi + \delta < 0$, that is, $0 < \delta < \varphi$. Since the aforementioned results hold for any $\delta > 0$, the condition $0 < \delta < \varphi$ always holds for any $\varphi > 0$ with δ depending on φ, say $\delta = \varphi/2$. Moreover, each summand in r in (8) is $o_p(1)$ since $(d - \hat{d})^r T^{r\varphi} = O_p(1)$. Lastly, consider the expression (9). Using the Cauchy-Schwarz inequality, we have

$$\sum_{t=1}^{T} \mu_{1,t} g^{(B)}(\varepsilon_t, d - \hat{d}) \leq \sqrt{\left(\sum_{t=1}^{T} \mu_{1,t}^2\right)\left(\sum_{t=1}^{T} [g^{(B)}(\varepsilon_t, d - \hat{d})]^2\right)} = O_p(\sqrt{T^{2(1-d)+1}T^2}),$$

where the equality holds due to Lemma 1(ii) and Robinson and Hualde (2003, C.13). Hence, the expression (9) is $O_p(T^{d-3/2}T^{5/2-d}T^{-B\varphi}) = O_p(T^{1-B\varphi}) = o_p(1)$ upon extending the Taylor series expansion up to $B > 1/\varphi$.

Lemma 6 *Under Assumption 1, for $d \in (-0.5, 1.5)$, $T^{d-3/2}\sum_{t=1}^{T}(\hat{\mu}_{1,t} - \mu_{1,t})\varepsilon_t \xrightarrow{p} 0$.*

Proof of Lemma A.6: Iterating the mean value theorem application in Robinson (2005, Lemma 4), for $|\widetilde{d} - d| \leq |\hat{d} - d|$,

$$\hat{\mu}_{1,t} = \mu_{1,t} + \sum_{r=1}^{B-1} \frac{1}{r!}(\hat{d} - d)^r \mu_{1,t}^{(r)} + \frac{1}{B!}(\hat{d} - d)^B \widetilde{\mu}_{1,t}^{(B)}$$

for a user chosen B, where $\widetilde{\mu}_{1,t}^{(B)} = (\ln T)^B \Delta^{\widetilde{d}-\bar{m}}\Delta^{\bar{m}} t\P\{t \geq 1\}$ with $\mu_{1,t}^{(r)}$ defined similarly (??check), and $\bar{m} = 0$ if $d \in (-0.5, 0.5)$ and $\bar{m} = 1$ if $d \in (0.5, 1.5)$. By choosing $B > (1 + d)/\varphi$, we have $(1/B!)(\hat{d} - d)^B \widetilde{\mu}_{1,t}^{(B)} = o_p(t^{1/2-d})$ (see Iacone et al., 2013, p.43). Then,

$$T^{d-3/2}\sum_{t=1}^{T}(\hat{\mu}_{1,t} - \mu_{1,t})\varepsilon_t = T^{d-3/2}\sum_{t=1}^{T}\left(\sum_{r=1}^{B-1}\frac{1}{r!}(\hat{d} - d)^r \mu_{1,t}^{(r)}\right)\varepsilon_t \tag{12}$$

$$+ T^{d-3/2}\sum_{t=1}^{T}\left(\frac{1}{B!}(\hat{d} - d)^B \widetilde{\mu}_{1,t}^{(B)}\right)\varepsilon_t. \tag{13}$$

For the expression (12), using summation by parts, we have

$$\begin{aligned}
T^{d-3/2}T^{-r\varphi}\sum_{t=1}^{T}\mu_{1,t}^{(r)}\varepsilon_t &\leq T^{d-3/2}T^{-r\varphi}\left|\sum_{t=1}^{T}\mu_{1,t}^{(r)}\varepsilon_t\right| \\
&\leq T^{d-3/2}T^{-r\varphi}\left[\sum_{t=1}^{T-1}|\mu_{1,t+1}^{(r)} - \mu_{1,t}^{(r)}|\left|\sum_{s=t+1}^{T}\varepsilon_s\right| + \left|\mu_{1,1}^{(r)}\sum_{t=1}^{T}\varepsilon_t\right|\right] \\
&\leq T^{d-3/2}T^{-r\varphi}\left[\sup_t\left|\sum_{s=t+1}^{T}\varepsilon_s\right|\sum_{t=1}^{T-1}|\mu_{1,t+1}^{(r)} - \mu_{1,t}^{(r)}| + \left|\mu_{1,1}^{(r)}\sum_{t=1}^{T}\varepsilon_t\right|\right],
\end{aligned}$$

which is $O_p(T^{-r\varphi+\delta}(\ln\Delta)^r)$ because $|\mu_{1,t+1}^{(r)} - \mu_{1,t}^{(r)}| = (\ln\Delta)^r\Delta^d\P\{t \geq 1\} = O(t^{-d+\delta}(\ln\Delta)^r)$ from Lemma 1(iii), $\sum_{t=1}^T |\mu_{1,t+1}^{(r)} - \mu_{1,t}^{(r)}| = O(t^{1-d+\delta}(\ln\Delta)^r)$ for any $\delta > 0$. This is the largest for $r = 1$, and it is $O_p(T^{-\varphi+\delta}(\ln\Delta))$, which is $o_p(1)$ when $0 < \delta < \varphi$. Furthermore, each summand in r in (12) is $o_p(1)$ because $(d-\hat{d})^r T^{r\varphi} = O_p(1)$. For the expression (13), using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}T^{d-3/2}\sum_{t=1}^T \left(\frac{1}{B!}(\hat{d}-d)^B\widetilde{\mu}_{1,t}^{(B)}\right)\varepsilon_t &\leq \left[T^{d-3/2}\sum_{t=1}^T \left(\frac{1}{B!}(\hat{d}-d)^B\widetilde{\mu}_{1,t}^{(B)}\right)^2 T^{d-3/2}\sum_{t=1}^T \varepsilon_t^2\right]^{1/2} \\ &= \sqrt{o_p(T^{d-3/2}T^{2-2d})O_p(T^{d-3/2}T)} = o_p(1),\end{aligned}$$

upon extending the Taylor series expansion up to $B > (1+d)/\varphi$.

Finally the proof that $T^{d-3/2}\sum_{t=1}^T(\hat{\mu}_{1,t}-\mu_{1,t})(u_t^{\hat{d}}-\varepsilon_t) \xrightarrow{p} 0$ is a straightforward combination of the aforementioned Lemmas A.5-A.6, hence omitted. We still need to consider the parts of (A.2) and (A.4) which pertain to $\mu_{0,t}$ and $\hat{\mu}_{0,t}$. First, for $-0.5 < d < 0.5$, we need to show that

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

or equivalently,

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

Lemma 7 *Under Assumption 1, for $d \in (-0.5, 0.5)$, $T^{d-1/2}\sum_{t=1}^T \mu_{0,t}(u_t^{\hat{d}}-\varepsilon_t) \xrightarrow{p} 0$.*

Proof of Lemma A.7: Note that $|\mu_{0,t+1}-\mu_{0,t}| = |\Delta^d(t+1)^0\P\{t \geq 1\} - \Delta^d t^0\P\{t \geq 1\}| = 0$ for $t = 1, \ldots, T-1$. Similar to the arguments in Lemma 5, using summation by parts,

$$\begin{aligned}&T^{d-1/2}T^{-r\varphi}|\sum_{t=1}^T \mu_{0,t}g^{(r)}(\varepsilon_t, 0)| \\ &\leq T^{d-3/2}T^{-r\varphi}\left[\sum_{t=1}^{T-1}|\mu_{1,t+1}-\mu_{1,t}||\sum_{s=t+1}^T g^{(r)}(\varepsilon_s,0)| + |\mu_{1,1}\sum_{t=1}^T g^{(r)}(\varepsilon_t,0)|\right] \\ &\leq T^{d-1/2}T^{-r\varphi}\left[\sup_t\left(|\sum_{s=t+1}^T g^{(r)}(\varepsilon_s,0)|\right)\sum_{t=1}^{T-1}|\mu_{0,t+1}-\mu_{0,t}| + |\mu_{0,1}\sum_{t=1}^T g^{(r)}(\varepsilon_t,0)|\right],\end{aligned}$$

which is $O_p(T^{d-1/2}T^{-r\varphi}(\ln T)^{r+1}T^{1/2}T^{-d})$ using Lemma 1(i). This is the largest for $r = 1$, and it is $O_p(T^{-\varphi}(\ln T)^2)$, which is $o_p(1)$ when $\varphi > 0$. Also, from Cauchy-Schwarz's inequality,

$$\sum_{t=1}^T \mu_{0,t}g^{(B)}(\varepsilon_t, d-\hat{d}) \leq \sqrt{\left(\sum_{t=1}^T \mu_{0,t}^2\right)\left(\sum_{t=1}^T[g^{(B)}(\varepsilon_t, d-\hat{d})]^2\right)} = O_p(\sqrt{T^{2(-d)+1}T^2}),$$

where the equality follows from Lemma 1(i) and Robinson and Hualde (2003, C.13). Hence, the expression $T^{d-1/2}\sum_{t=1}^{T}\mu_{0,t}(1/B!)(d-\hat{d})^B g^{(B)}(\varepsilon_t, d-\hat{d})$ is $O_p(T^{d-1/2}T^{3/2-d}T^{-B\varphi}) = O_p(T^{1-B\varphi}) = o_p(1)$ upon extending the Taylor series expansion up to $B > 1/\varphi$.

Lemma 8 *Under Assumption 1, for $d \in (-0.5, 0.5)$, $T^{d-1/2}\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})\varepsilon_t \xrightarrow{p} 0$.*

Proof of Lemma A.8: Iterating the mean value theorem application in Robinson (2005, Lemma 4), for $|\widetilde{d}-d| \leq |\hat{d}-d|$,

$$\hat{\mu}_{0,t} = \mu_{0,t} + \sum_{r=1}^{B-1}\frac{1}{r!}(\hat{d}-d)^r\mu_{0,t}^{(r)} + \frac{1}{B!}(\hat{d}-d)^B\widetilde{\mu}_{0,t}^{(B)}$$

for a user chosen B, where $\widetilde{\mu}_{0,t}^{(B)} = (\ln T)^B\Delta^{\widetilde{d}-\bar{m}}\Delta^{\bar{m}}1\P\{t \geq 1\}$. By choosing $B > (1+d)/\varphi$, we have $(1/B!)(\hat{d}-d)^B\widetilde{\mu}_{0,t}^{(B)} = o_p(t^{-1/2-d})$. Then,

$$T^{d-1/2}\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})\varepsilon_t = T^{d-1/2}\sum_{t=1}^{T}\left(\sum_{r=1}^{B-1}\frac{1}{r!}(\hat{d}-d)^r\mu_{0,t}^{(r)}\right)\varepsilon_t + T^{d-1/2}\sum_{t=1}^{T}\left(\frac{1}{B!}(\hat{d}-d)^B\widetilde{\mu}_{0,t}^{(B)}\right)\varepsilon_t$$

Using summation by parts, we have

$$T^{d-1/2}T^{-r\varphi}\sum_{t=1}^{T}\mu_{0,t}^{(r)}\varepsilon_t \leq T^{d-1/2}T^{-r\varphi}[\sup_t|\sum_{s=t+1}^{T}\varepsilon_s|\sum_{t=1}^{T-1}|\mu_{0,t+1}^{(r)}-\mu_{0,t}^{(r)}| + |\mu_{0,1}^{(r)}\sum_{t=1}^{T}\varepsilon_t|],$$

which is $O_p(T^{d-1/2-r\varphi}T^{-d+1/2})$ because $|\mu_{0,t+1}^{(r)}-\mu_{0,t}^{(r)}| = 0$ for $t = 1,\ldots,T-1$. This is the largest for $r = 1$, and it is $O_p(T^{-\varphi})$, which is $o_p(1)$ when $\varphi > 0$. Furthermore, each summand in r is $o_p(1)$ because $(d-\hat{d})^r T^{r\varphi} = O_p(1)$. Using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} T^{d-1/2}\sum_{t=1}^{T}\left(\frac{1}{B!}(\hat{d}-d)^B\widetilde{\mu}_{0,t}^{(B)}\right)\varepsilon_t &\leq [T^{d-1/2}\sum_{t=1}^{T}\left(\frac{1}{B!}(\hat{d}-d)^B\widetilde{\mu}_{0,t}^{(B)}\right)^2 T^{d-1/2}\sum_{t=1}^{T}\varepsilon_t^2]^{1/2} \\ &= \sqrt{o_p(T^{d-1/2}T^{-2d})O_p(T^{d-1/2}T)} = o_p(1), \end{aligned}$$

upon extending the Taylor series expansion up to $B > (1+d)/\varphi$.

The proof that $T^{d-1/2}\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})(u_t^{\hat{d}}-\varepsilon_t) \xrightarrow{p} 0$ is a straightforward combination of the aforementioned arguments and omitted. Second, for $0.5 < d < 1.5$, we need to show that $\sum_{t=1}^{T}\hat{\mu}_{0,t}\,u_t^{\hat{d}} - \sum_{t=1}^{T}\mu_{0,t}\,\varepsilon_t \xrightarrow{p} 0$, or equivalently,

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

Lemma 9 *Under Assumption 1, for $d \in (0.5, 1.5)$, $\sum_{t=1}^{T}\mu_{0,t}(u_t^{\hat{d}}-\varepsilon_t) \xrightarrow{p} 0$.*

Proof of Lemma A.9: Using summation by parts, we have

$$T^{-r\varphi}|\sum_{t=1}^{T}\mu_{0,t}g^{(r)}(\varepsilon_t,0)| \leq T^{d-3/2}T^{-r\varphi}\left[\sum_{t=1}^{T-1}|\mu_{1,t+1}-\mu_{1,t}||\sum_{s=t+1}^{T}g^{(r)}(\varepsilon_s,0)|+|\mu_{1,1}\sum_{t=1}^{T}g^{(r)}(\varepsilon_t,0)|\right]$$
$$\leq T^{-r\varphi}\left[\sup_t\left(|\sum_{s=t+1}^{T}g^{(r)}(\varepsilon_s,0)|\right)\sum_{t=1}^{T-1}|\mu_{0,t+1}-\mu_{0,t}|+|\mu_{0,1}\sum_{t=1}^{T}g^{(r)}(\varepsilon_t,0)|\right],$$

which is $O_p(T^{-r\varphi}(\ln T)^{r+1}T^{1/2}T^{-d+\delta})$ for any $\delta>0$ using Lemma 1(iii). This is the largest for $r=1$, and it is $O_p(T^{-\varphi+1/2-d+\delta}(\ln T)^2)$, which is $o_p(1)$ when $0<\delta\leq\varphi$. Each summand in r is also $o_p(1)$. Moreover, using the Cauchy-Schwarz inequality, we have

$$\sum_{t=1}^{T}\mu_{0,t}g^{(B)}(\varepsilon_t,d-\hat{d}) \leq \sqrt{\left(\sum_{t=1}^{T}\mu_{0,t}^2\right)\left(\sum_{t=1}^{T}[g^{(B)}(\varepsilon_t,d-\hat{d})]^2\right)} = O_p(\sqrt{T^{2(-d+\delta)+1}T^2}),$$

where the equality follows from Lemma 1(iii) and Robinson and Hualde (2003, C.13). Hence, the expression $\sum_{t=1}^{T}(1/B!)(\hat{d}-d)^B\widetilde{\mu}_{0,t}^{(B)}\varepsilon_t = O_p(T^{3/2-d+\delta-B\varphi}) = o_p(1)$ upon extending the Taylor series expansion up to $B>(3/2-d+\delta)/\varphi$.

Lemma 10 *Under Assumption 1, for $d\in(0.5,1.5)$, $\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})\varepsilon_t \xrightarrow{p} 0$.*

Proof of Lemma A.10: We have:

$$\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})\varepsilon_t = \sum_{t=1}^{T}\left(\sum_{r=1}^{B-1}\frac{1}{r!}(\hat{d}-d)^r\mu_{0,t}^{(r)}\right)\varepsilon_t + \sum_{t=1}^{T}\left(\frac{1}{B!}(\hat{d}-d)^B\widetilde{\mu}_{0,t}^{(B)}\right)\varepsilon_t$$

and using summation by parts,

$$T^{-r\varphi}\sum_{t=1}^{T}\mu_{0,t}^{(r)}\varepsilon_t \leq T^{-r\varphi}[\sup_t|\sum_{s=t+1}^{T}\varepsilon_s|\sum_{t=1}^{T-1}|\mu_{0,t+1}^{(r)}-\mu_{0,t}^{(r)}|+|\mu_{0,1}^{(r)}\sum_{t=1}^{T}\varepsilon_t|],$$

which is $O_p(T^{-r\varphi-d+\delta+1/2})$ because $|\mu_{0,t+1}^{(r)}-\mu_{0,t}^{(r)}|=0$ for $t=1,\ldots,T-1$. This is the largest for $r=1$, and it is $O_p(T^{-\varphi-d+\delta+1/2})$, which is $o_p(1)$ when $0<\delta\leq\varphi$. Also, each summand in r is $o_p(1)$ since $(d-\hat{d})^rT^{r\varphi}=O_p(1)$. Using Cauchy-Schwarz's inequality,

$$\sum_{t=1}^{T}\left(\frac{1}{B!}(\hat{d}-d)^B\widetilde{\mu}_{0,t}^{(B)}\right)\varepsilon_t \leq [\sum_{t=1}^{T}\left(\frac{1}{B!}(\hat{d}-d)^B\widetilde{\mu}_{0,t}^{(B)}\right)^2\sum_{t=1}^{T}\varepsilon_t^2]^{1/2}$$
$$= \sqrt{o_p(T^{-2d})O_p(T)} = O_p(T^{1/2-d}) = o_p(1).$$

Finally the proof that $\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})(u_t^{\hat{d}}-\varepsilon_t) \xrightarrow{p} 0$ is a straightforward combination of the aforementioned arguments, hence omitted.

From Shimotsu (2010, Theorem 4), under Assumptions 1-2, $\hat{d} - d = O_p(m^{-1/2})$ 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, that is, $\hat{\sigma}^2 - \sigma^2 = o_p(1)$ (see, e.g., 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 the bandwidth $m = [\mathcal{C}\, T^{2\varphi}]$ for any $\varphi > 0$.

Proof of Theorem 2: Here, we suggest a simple modification for the proof of Iacone et al. (2013, Lemma A.2), which plays a crucial role in establishing the result in Theorem 2. In the proof of Lemma A.2, Iacone et al. (2013, p.44) considered, using summation by parts,

$$
\begin{aligned}
&T^{d-3/2}T^{-r\varphi} \sup_{\tau} |\sum_{t=1+[\tau T]}^{T} \mu_t(\tau) g^{(r)}(\varepsilon_t, 0)| \\
&\leq T^{d-3/2}T^{-r\varphi} \left(\sum_{t=1}^{T} |\mu_{1,t+1} - \mu_{1,t}| \sup_{t} |\sum_{s=t}^{T} g^{(r)}(\varepsilon_s, 0)| + \sup_{\tau} |\sum_{t=1+[\tau T]}^{T} g^{(r)}(\varepsilon_t, 0)| \right), \qquad (14)
\end{aligned}
$$

which is $o_p(1)$ when $d \leq 1$, using the uniform convergence of $\sum_{t=1}^{[\tau T]} g^{(r)}(\varepsilon_t, 0)$ and $|\mu_{1,t+1} - \mu_{1,t}| \leq \mathcal{C}t^{-d}$. This result is based on Robinson (2005, Lemma 1), more specifically, Lemma 1(i). For $d > 1$, Iacone et al. (2013) used Lemma 1(i) to show that the same expression is $o_p(T^{d-3/2-r\varphi+1/2}(\ln T)^{r+1}) = o_p(T^{d-1-r\varphi}(\ln T)^{r+1})$. For $r = 1$, this is the largest as $o_p(T^{d-1-\varphi}(\ln T)^2)$, which is $o_p(1)$ when $\varphi > d - 1$. This follows since from Lemma 1(i) $|\mu_{1,t+1} - \mu_{1,t}| = \Delta^d \P\{t \geq 1\} \leq \mathcal{C}t^{-d}$ and $\sum_{t=1}^{T} |\mu_{1,t+1} - \mu_{1,t}| = O(T^{1-d}) = o(1)$ for $d > 1$. However, the bound in Lemma 1(i) is not the strongest possible using Lema 1(i). Instead, using Lemma 1(iii), $|\mu_{1,t+1} - \mu_{1,t}| = \Delta^d \P\{t \geq 1\} = O_p(T^{-d+\delta})$ for any $\delta > 0$. So the expression (14) is $o_p(T^{d-3/2-r\varphi+1/2}(\ln T)^{r+1}T^{1-d+\delta}) = o_p(T^{\delta-r\varphi}(\ln T)^{r+1})$. When $r = 1$, it is $o_p(T^{\delta-\varphi}(\ln T)^2)$, which is $o_p(1)$ when $0 < \delta < \varphi$. Since the aforementioned results hold for any $\delta > 0$, the condition $0 < \delta < \varphi$ always holds for any $\varphi > 0$ with δ depending on φ, say $\delta = \varphi/2$. With this modification, the arguments in Iacone et al. (2013, Lemma A.2) hold for any $\varphi > 0$, which establish the results in Theorem 2 for any $\varphi > 0$ as desired.

Table 1: Finite Sample Size; Pure Fractional Processes.

T	d	-0.4	0.2	0.4	0.8	1	1.4
500	ELW	0.054	0.056	0.057	0.047	0.062	0.051
	MU	0.000	0.098	0.157	0.155	0.053	0.462
	UB	0.000	0.093	0.139	0.100	0.052	0.462
1000	ELW	0.047	0.051	0.059	0.034	0.043	0.049
	MU	0.000	0.138	0.173	0.134	0.049	0.495
	UB	0.000	0.138	0.163	0.099	0.049	0.495
2000	ELW	0.054	0.047	0.044	0.051	0.043	0.039
	MU	0.000	0.178	0.277	0.108	0.046	0.559
	UB	0.000	0.178	0.277	0.097	0.046	0.559

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	ELW	0.074	0.068	0.049	0.018	0.017	0.005
	MU	0.051	0.059	0.030	0.037	0.049	0.045
	UB	0.051	0.059	0.030	0.037	0.047	0.034
1000	ELW	0.085	0.068	0.057	0.008	0.017	0.005
	MU	0.057	0.031	0.046	0.046	0.042	0.053
	UB	0.057	0.031	0.046	0.046	0.042	0.051
2000	ELW	0.064	0.067	0.069	0.017	0.029	0.005
	MU	0.066	0.045	0.058	0.041	0.044	0.049
	UB	0.066	0.045	0.058	0.041	0.044	0.049

Table 3 Finite Sample Sizes; DGP 1-5 with d=0.4

	DGP-1	DGP-1	DGP-2	DGP-2	DGP-3	DGP-4	DGP-5
T	AR=0.4	AR=-0.4	MA=0.4	MA=-0.4	AR1=0.3, AR2=0.5	Measurement error	GARCH
500	0.061	0.059	0.084	0.107	0.033	0.052	0.084
1000	0.057	0.058	0.076	0.094	0.056	0.031	0.074
2000	0.048	0.065	0.08	0.069	0.045	0.052	0.069

Figure 1: Unadjusted power for pure fractional processes

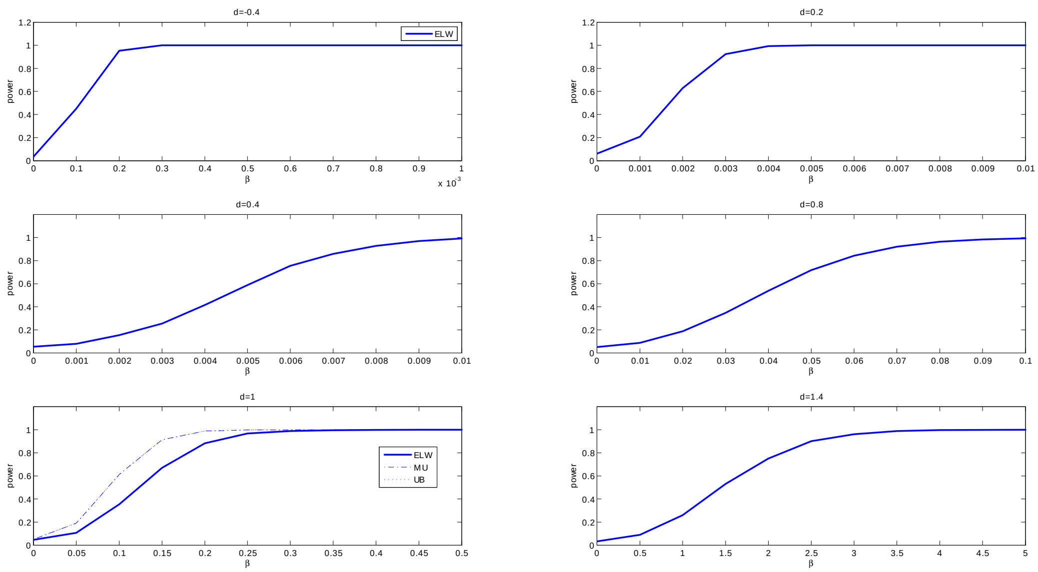

Figure 2: Unadjusted power for AR(1) processes

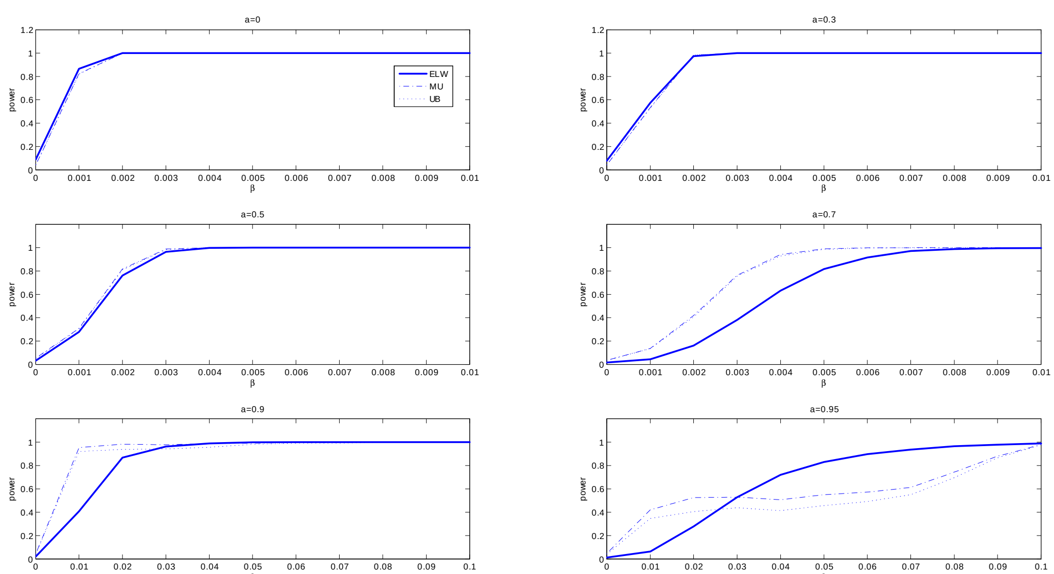

Figure 3: Unadjusted power for DGP 1-5