Partially one-sided semiparametric inference for trending persistent and antipersistent processes

Karim M. Abadir[1] Walter Distaso [1]

Liudas Giraitis[2] *

[1] *Business School, Imperial College London, London SW7 2AZ, UK*

[2] *School of Economics and Finance, Queen Mary University of London*

January 3, 2022

Abstract

Hypothesis testing in models allowing for trending processes that are possibly non-stationary and non-Gaussian is considered. Using semiparametric estimators, joint hypothesis testing for these processes is developed, taking into account the one-sided nature of typical hypotheses on the persistence parameter in order to gain power. The results are applicable for a wide class of processes and are easy to implement. They are illustrated with an application to the dynamics of GDP.

Key words: fractional integration and trends, partially one-sided joint hypotheses, fully-extended local Whittle estimation.

JEL Classification: C22.

*Corresponding Author.

E-mail addresses: k.m.abadir@imperial.ac.uk (K.M. Abadir), w.distaso@imperial.ac.uk (W. Distaso), l.giraitis@qmul.ac.uk (L. Giraitis).

1 Introduction

Testing for persistence is an important subject in time series. It is particularly of interest to macroeconomists to determine how Gross Domestic Product (GDP) and other variables evolve. Due to its potential impact on the choice of economic stabilization policies, this question has generated a huge literature that was largely started by Nelson and Plosser (1982). The current paper considers testing joint hypotheses about the extent of persistence and the possibility of a trend in time series. We use the results of Abadir and Distaso (2007), with the implication here that power can be gained by modifying tests of joint hypotheses to take into account the fact that inference on the persistence parameter is typically one-sided, whereas inference on the remaining components are two-sided. The trend parameters are estimated in the time domain, but we estimate the persistence in the frequency domain using the Fully-Extended Local Whittle (FELW) estimator of Abadir, Distaso, and Giraitis (2007, 2011) which extends to nonstationarity the classical local Whittle estimator proposed by Künsch (1987) and Robinson (1995). By virtue of the specification being semiparametric, it generates robust inference: it allows for seasonality and other effects to be present at nonzero spectral frequencies and it is valid for a wide class of generating processes that include non-Gaussian ones. It is also easily usable in applied work.

There are precursors to using frequency-domain estimators (including ones obtained via autocorrelation functions) in testing hypotheses about trending persistent series. First, Robinson (1994) introduces such tests that are applied in Gil-Alaña and Robinson (1997). However, the partially one-sided nature of the joint hypotheses is not taken into account in their setup (see the alternative hypothesis in their (28)) and there is power to be gained from doing so. Second, Dolado, Gonzalo, and Mayoral (2008, 2009) allow for trends in the efficient formulation which Lobato and Velasco (2006) introduce as a modification of the original one in Dolado, Gonzalo, and Mayoral (2002). They generalize the tests of Dickey and Fuller (1979) to allow for fractional persistence. They detrend the series but do not consider joint hypotheses on the trend as well as persistence, which is done by Gil-Alaña and Robinson (1997) and by Dickey and Fuller (1981). Our procedure also differs from Dolado *et al.* (2002) in the robustness indicated in the previous paragraph when estimating the degree of persistence.

In this paper, $\overset{p}{\to}$ and $\overset{d}{\to}$ denote respectively convergence in probability and in distribution. We write 1_A for the indicator of a set A, $\lfloor \nu \rfloor$ for the integer part of ν, C for a generic constant but $c_\bullet$ for specific constants. The lag operator is denoted by L, such that $\mathrm{L}u_t = u_{t-1}$, and the backward difference operator by $\nabla := 1 - \mathrm{L}$. We write i for the imaginary unit (principal value of $\sqrt{-1}$), in roman typeface to distinguish it from the index i. Consider the process

$$X_t = \alpha + \beta t + u_t, \quad t = 1, 2, ..., n, \tag{1.1}$$

where the sequence $\{u_t\} \sim \mathrm{I}(d)$ satisfies the following definition.

DEFINITION 1.1. *For $d = k + d_\xi$, where $k \in \mathbb{Z}$ is an integer and $d_\xi \in (-1/2, 1/2)$, we say that $\{u_t\}$ is an $\mathrm{I}(d)$ process (also denoted by $u_t \sim \mathrm{I}(d)$) if*

$$\nabla^k u_t = \xi_t, \quad t = 1 - k, 2 - k, \ldots,$$

where $\{\xi_t\}$ *is a second order stationary sequence with spectral density*

$$f_\xi(\lambda) = b_0\,|\lambda|^{-2d_\xi} + o(|\lambda|^{-2d_\xi}), \text{ as } \lambda \to 0 \tag{1.2}$$

where $b_0 > 0$.

Note that we use the term "stationarity" in a weaker sense than usual, only requiring the leading term of the spectrum to be as in (1.2). Few papers have so far considered such settings with an extended range for d to include regions of nonstationarity and to estimate a time trend, and to conduct joint hypothesis testing, as discussed earlier.

We will assume that the process $\{\xi_t\}$ is a linear sequence as follows.

Assumption A.1. $\{\xi_t\}$ is a linear sequence

$$\xi_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \tag{1.3}$$

where $\{a_j,\, j \geq 0\}$ are real nonrandom weights, $\sum_{j=0}^{\infty} a_j^2 < \infty$, and $\{\varepsilon_j\}$ are i.i.d. variates with zero mean, unit variance, and finite fourth moment $\mathrm{E}\varepsilon_0^4 < \infty$. Moreover, the spectral density $f_\xi(\lambda)$ of $\{\xi_t\}$ has the property

$$f_\xi(\lambda) = |\lambda|^{-2d_\xi}\big(b_0 + b_1\lambda^2 + o(\lambda^2)\big), \text{ as } \lambda \to 0, \tag{1.4}$$

for some $d_\xi \in (-1/2, 1/2)$, $b_0 > 0$, and finite b_0, b_1. Defining $A(\lambda) := \sum_{j=0}^{\infty} \mathrm{e}^{\mathrm{i}j\lambda} a_j$, it is also required that

$$\frac{\mathrm{d}A(\lambda)}{\mathrm{d}\lambda} = O(|A(\lambda)|/\lambda), \quad \text{as} \quad \lambda \to 0^+.$$

For convenience, we need the following assumption on the true d.

Assumption A.2. $\{u_t\} \sim \mathrm{I}(d)$, with $d \in (-1/2, 3/2)$, $d \neq 1/2$.

It is technically straightforward to extend our results to all values of $d > 3/2$ that give rise to nonstationarity, as well as to higher-order polynomials. We do not report such extensions, in order to keep the exposition as clear as possible and because the applications' literature that we just cited requires at most linear trends.

In Section 2, we present the estimators and their basic properties for later use. Section 3 contains the construction of the new tests and their limiting distributions under the null and alternatives. Section 4 illustrates the gains of our approach by means of a simulation study. Section 5 demonstrates the ease of our approach by applying it to the dynamics of GDP. Proofs of the main results are given in Section 6.

2 The estimators and their properties

This brief section is not new, but it collects results we need from Abadir *et al.* (2007, 2011) mainly and lays the ground for the derivations in the following sections. In order to estimate the slope parameter β and the location parameter α of (1.1), we use the standard least squares (LS) estimators

$$\widehat{\beta} = \frac{\sum_{t=1}^{n}(X_t - \bar{X})(t - \bar{t})}{\sum_{t=1}^{n}(t - \bar{t})^2}, \qquad \widehat{\alpha} = \bar{X} - \widehat{\beta}\bar{t},$$

where $\bar{X} = n^{-1}\sum_{t=1}^{n} X_t$ and $\bar{t} = n^{-1}\sum_{t=1}^{n} t = (n+1)/2$ are the sample means of the variables. To estimate d, we start by calculating the detrended data

$$\widehat{u}_t = X_t - \widehat{\alpha} - \widehat{\beta}t = u_t + \alpha - \widehat{\alpha} + (\beta - \widehat{\beta})t, \quad t = 0, 1, ..., n.$$

Let

$$I_{n,u}(\lambda_j) := |w_u(\lambda_j)|^2, \qquad w_u(\lambda_j) := (2\pi n)^{-1/2}\sum_{t=1}^{n} \mathrm{e}^{\mathrm{i}t\lambda_j} u_t$$

the periodogram and discrete Fourier transform of $\{u_t\}$, where $\lambda_j = 2\pi j/n$, $j = 1, \ldots, n$ denote the Fourier frequencies. The FELW estimator $\widehat{d}$ of d based on the residuals $\{\widehat{u}_t\}$ is defined as

$$\widehat{d} := \mathrm{argmin}_{\delta\in[-1/2,3/2]}\, U_n(\delta), \tag{2.1}$$

where

$$U_n(\delta) := \begin{cases} \log\left(\frac{1}{m}\sum_{j=1}^{m} j^{2\delta}\, I_{n,\hat{u}}(\lambda_j)\right) - \frac{2\delta}{m}\sum_{j=1}^{m}\log j, & \text{if } \delta \in [-1/2, 1/2], \\ \log\left(\frac{1}{m}\sum_{j=1}^{m} j^{2\delta}|1 - \mathrm{e}^{\mathrm{i}\lambda_j}|^{-2}\, I_{n,\nabla\hat{u}}(\lambda_j)\right) - \frac{2\delta}{m}\sum_{j=1}^{m}\log j, & \text{if } \delta \in (1/2, 3/2], \end{cases} \tag{2.2}$$

where the bandwidth parameter m is such that $m \to \infty$ and $m = o(n)$. In applying the estimation method, the sample data must be enumerated as $\widehat{u}_0, ..., \widehat{u}_n$ when $\widehat{d} \in [1/2, 3/2]$. Note that the minimization in (2.1) that yields $\widehat{d}$ is carried out over $[-1/2, 3/2]$, so that $\widehat{d}$ is restricted to this range. For the limiting distributions of the next section, we will also need to introduce the following two estimators.

The estimator of the scale parameter b_0 of (1.2) is defined by

$$\widehat{b}_{m,\widehat{u}}(\widehat{d}) := \begin{cases} \frac{1}{m}\sum_{j=1}^{m} \lambda_j^{2\widehat{d}}\, I_{n,\widehat{u}}(\lambda_j), & \text{if } \widehat{d} \in [-1/2, 1/2], \\ \frac{1}{m}\sum_{j=1}^{m} \lambda_j^{2\widehat{d}}|1 - \mathrm{e}^{\mathrm{i}\lambda_j}|^{-2}\, I_{n,\Delta\widehat{u}}(\lambda_j), & \text{if } \widehat{d} \in (1/2, 3/2]. \end{cases} \tag{2.3}$$

Abadir *et al.* (2011) show the consistencies

$$\widehat{d} \xrightarrow{p} d \quad \text{and} \quad \widehat{b}_{m,\widehat{u}}(\widehat{d}) \xrightarrow{p} b_0 \tag{2.4}$$

as $n \to \infty$, $m \to \infty$ and $m = O(n^{4/5})$. Moreover, if $m = o(n^{4/5})$, the estimator $\widehat{d}$ satisfies

$$(\sum_{j=1}^{m}\nu_j^2)^{1/2}(\widehat{d} - d) \xrightarrow{d} \mathrm{N}(0, 1/4), \qquad \nu_j := \log j - m^{-1}\sum_{k=1}^{m}\log k. \tag{2.5}$$

The factor $\sum_{j=1}^{m}\nu_j^2$ leads to a better finite-sample approximation than the asymptotic $\sqrt{m}(\widehat{d} - d) \xrightarrow{d} \mathrm{N}(0, 1/4)$ that was reported in Section 2.3 and Section 3 in Abadir *et al.*

(2007, 2011), respectively. The validity of (2.5) follows from the proof of Theorem 2.4 in Abadir *et al.* (2007); see also the proof of Theorem 2 in Robinson (1995) as well as Theorem 8.5.2 and Remark 8.5.2 in Giraitis, Koul, and Surgailis (2012). Since the convergence $(\sum_{j=1}^{m}\nu_j^2)/m \to 1$ as $m \to \infty$ is rather slow, we recommend (2.5) for use in applied work. Also, Cheung and Hassler (2018) and Cheung (2020) noted that the finite-sample normal approximation (2.5) can be distorted when d is close to discontinuity points $-0.5, 0.5, 1.5$ of the objective function $U_n(\delta)$ and they suggested a solution to this problem.

The testing procedures considered below require us to estimate the long run variance s_ξ^2 of $\{\xi_t\}$. We recall that Property (1.2) of the spectral density f_ξ implies that

$$s_\xi^2 := \lim_{n\to\infty} \mathrm{E}\left(n^{-1/2-d_\xi}\sum_{t=1}^{n}\xi_t\right)^2 = \lim_{n\to\infty} n^{-1-2d_\xi}\int_{-\pi}^{\pi}\left(\frac{\sin(n\lambda/2)}{\sin(\lambda/2)}\right)^2 f_\xi(\lambda)\mathrm{d}\lambda = p(d_\xi)b_0, \tag{2.6}$$

where

$$p(d) := \int_{-\infty}^{\infty}\left(\frac{\sin(\lambda/2)}{\lambda/2}\right)^2|\lambda|^{-2d}\mathrm{d}\lambda = \begin{cases} 2\frac{\Gamma(1-2d)\sin(\pi d)}{d(1+2d)}, & \text{if } d\neq 0,\\ 2\pi, & \text{if } d=0.\end{cases}$$

Under Assumption A.1 and the consistencies in (2.4), the probability limit of

$$\widehat{s}_m^2(\widehat{d}) := \begin{cases} p(\widehat{d})\widehat{b}_{m,\widehat{u}}(\widehat{d}), & \text{if } \widehat{d}\in[-1/2,1/2),\\ p(\widehat{d}-1)\widehat{b}_{m,\widehat{u}}(\widehat{d}), & \text{if } \widehat{d}\in[1/2,3/2]\end{cases} \tag{2.7}$$

equals the long-run variance s_ξ^2, and we use the same bandwidth parameter m used to calculate the FELW estimator (2.1). In determining $p\,(d-1)$ as $d \to 3/2$ from below, one should note that $\lim_{c\to 0^-}\Gamma(-c) = +\infty$. For the long-run variance when $d \in (-1/2, 1/2)$, see Robinson (2005).

To describe the asymptotic distribution of $\widehat{\beta}$, define

$$\sigma_\beta^2(d) := \begin{cases} (12)^2\left(\frac{1}{2d+3}-\frac{1}{4}\right), & \text{if } d\in[-1/2,1/2),\\ (12)^2\frac{2d-1}{8d(2d+1)(2d+3)}, & \text{if } d\in[1/2,3/2].\end{cases}$$

Then,

$$\frac{n^{3/2-d}}{s_\xi\sigma_\beta(d)}(\widehat{\beta}-\beta) \xrightarrow{d} \mathrm{N}(0,1). \tag{2.8}$$

This result follows from Theorem 2.1 in Abadir *et al.* (2011). Recall that $\{\xi_t\}$ is a linear process which satisfies Assumption A.1 and appears in Definition 1.1 of u_t. The conditions of that theorem require that finite-dimensional distributions of the partial sums process $Y_n(r) := n^{-1/2-d_\xi}\sum_{t=1}^{\lfloor nr\rfloor+1}\xi_t$, $0 \le r \le 1$, converge to those of the Gaussian process $Y_\infty(r) = s_\xi B_{1/2+d_\xi}(r)$,

$$Y_n(.) \xrightarrow{d} Y_\infty(.) \tag{2.9}$$

where $B_{1/2+d_\xi}(.)$ is Gaussian process (fractional Brownian motion) with zero mean and covariance function

$$\mathrm{cov}\big(B_{1/2+d_\xi}(r), B_{1/2+d_\xi}(s)\big) = (1/2)\big(r^{1+2d\xi} + s^{1+2d\xi} - |r-s|^{1+2d\xi}\big) =: R(r,s).$$

Convergence (2.9) is shown in Proposition 3.1 of Abadir *et al.* (2013). Moreover, property (1.2) together with definition of $Y_n(r)$ implies

$$\mathrm{E}\big(Y_n(r)Y_n(s)\big) \to s_\xi^2 R(r,s) = \mathrm{E}\big(Y_\infty(r)Y_\infty(s)\big). \tag{2.10}$$

3 Testing joint hypotheses

In this section, we discuss testing joint hypotheses in (1.1). Recall that in the case $d \in (-1/2, 1/2)$, we have that $u_t = \xi_t$ is stationary with memory parameter d. If $d \in (1/2, 3/2)$, then u_t can be written as

$$u_t = u_{t-1} + \xi_t, \quad t = 1, 2, ..., n,$$

where $\{\xi_t\}$ is a stationary $\mathrm{I}(d_\xi)$ process with the memory parameter $d_\xi = d-1 \in (-1/2, 1/2)$. We assume that $\{\xi_t\}$ is a linear process which satisfies Assumption A.1. Note that fractional ARIMA(p, d_ξ, q) sequences $\{\xi_t\}$ with memory parameter $d_\xi \in (-1/2, 1/2)$ satisfy Assumption A.1.

Definition 3.1. *Let $d_0 \in (-1/2, 3/2)$, $d_0 \neq 1/2$, and $\beta_0 \in \mathbb{R}$. We say that $\{X_t\}$ given by (1.1) satisfies:*

(a) *the null hypothesis* $\mathrm{H}_0(d_0, \beta_0)$ *if $d = d_0$ and $\beta = \beta_0$;*

(b) *the alternative hypothesis* $\mathrm{H}_1(< d_0, \beta_0)$ *if $d < d_0$ or if*

$$d \leq d_0 \text{ and } \beta \neq \beta_0;$$

(c) *the alternative hypothesis* $\mathrm{H}_1(> d_0, \beta_0)$ *if $d > d_0$ or if*

$$d \geq d_0 \text{ and } \beta \neq \beta_0.$$

Remark 3.1. Notice in the displayed cases of H_1 that d is restricted to be at one side of d_0, even when the violation of H_0 occurs because of $\beta \neq \beta_0$. Compared to the usual approaches in the literature where $\mathrm{H}_1 : d \neq d_0$ or $\beta \neq \beta_0$, we exclude the case of two-sided alternatives on d in order to gain power, for any given size, and focus on only one side of potential violation of $d = d_0$ at a time (case b then case c). The following examples apply:

(a) Under $\mathrm{H}_0(0, 0)$, $X_t = \alpha + \xi_t$ is a short memory I(0) sequence. Under $\mathrm{H}_0(1, 0)$, $X_t = \alpha + u_t$ where $u_t = u_{t-1} + \xi_t$ is a unit root I(1) process with short memory $\{\xi_t\}$.

(b) Alternative $\mathrm{H}_1(< 1, 0)$ covers processes $X_t = \alpha + u_t$ with $u_t \sim \mathrm{I}(d)$, $d < 1$, and processes with a linear trend and $d \leq 1$. Alternative $\mathrm{H}_1(< 0, 0)$ covers antipersistent stationary processes $X_t = \alpha + \xi_t$ with $d < 0$, due in practice to overdifferencing, and processes with a linear trend and no long memory.

(c) Alternative $\mathrm{H}_1(> 0, 0)$ covers processes with a linear trend and no antipersistence. It also covers processes with long memory or nonstationary $u_t \sim \mathrm{I}(d)$, subject to $d < 3/2$.

Note that Dickey and Fuller (1981) have alternatives of the form $\mathrm{H}_1(< 1, 0) \cup \mathrm{H}_1(> 1, 0)$, allowing explosive roots and $d > 1$ (but no fractional d). Given the nature of the alternative hypothesis, which is one-sided for the memory parameter d, it is possible to modify the conventional test statistics to construct more powerful ones. We use the general principle introduced in Abadir and Distaso (2007) to devise our tests, and then we work out their asymptotic distributions.

Let us start with the setup for testing $\mathrm{H}_0(d_0,\beta_0)$ versus $\mathrm{H}_1(>d_0,\beta_0)$. We use

$$\tau_r := \tau_{d_0}^2 1_{\widehat{d}>d_0} + \tau_{\beta_0}^2(\widehat{d}), \tag{3.1}$$

as a modification of $\tau := \tau_{d_0}^2 + \tau_{\beta_0}^2(\widehat{d})$, where

$$\tau_{d_0} := 2(\sum_{j=1}^{m} \nu_j^2)^{1/2}(\widehat{d}-d_0), \qquad \tau_{\beta_0}(\widehat{d}) := \frac{n^{3/2-\widehat{d}}}{\widehat{s}_m(\widehat{d})\sigma_\beta(\widehat{d})}(\widehat{\beta}-\beta_0)$$

with

$$m \to \infty, \quad m = o(n^{4/5}), \quad m^{-1}\log^4 n = o(1). \tag{3.2}$$

The modification of the test statistics comes from the indicator function associated with the first component of (3.1). Unlike in Abadir and Distaso (2007), there is no need to orthogonalize the two components of the test statistic since, as will be shown in Theorem 3.1, $\mathrm{cov}(\tau_{d_0},\tau_{\beta_0}(\widehat{d})) \to 0$: consequently, the inner boundaries of the critical regions of the unmodified tests are circular rather than elliptical.

The following theorem establishes some asymptotic properties of the normalized estimators of d and β. These are required for the application of the joint testing procedure.

Theorem 3.1. *Suppose that $X_1,...,X_n$ are given by (1.1), $u_t \sim \mathrm{I}(d_0)$ with $d_0 \in (-1/2,3/2)$, $d_0 \neq 1/2$, and $\{\xi_t\}$ satisfies Assumption A.1. Then, under hypothesis $\mathrm{H}_0(d_0,\beta_0)$, as $n\to\infty$,*

$$\left(\tau_{d_0},\tau_{\beta_0}(\widehat{d})\right) \xrightarrow{d} (Z_1,Z_2), \qquad \tau \xrightarrow{d} Z_1^2+Z_2^2 \sim \chi^2(2), \tag{3.3}$$

where $Z_1 \sim \mathrm{N}(0,1)$ and $Z_2 \sim \mathrm{N}(0,1)$ are independent.

To derive the asymptotic distribution of τ_r, we know from (2.5) and Theorem 2.1 of Abadir *et al.* (2011) that, as $n\to\infty$,

$$\tau_{d_0} \xrightarrow{d} \mathrm{N}(\zeta,1) \text{ and } \tau_{\beta_0}^2(\widehat{d}) \xrightarrow{d} \chi^2(1,\delta)$$

independently, where $\zeta = 2\sqrt{m}(d-d_0)$ and the χ^2 has one degree of freedom and noncentrality parameter

$$\delta = \frac{n^{3-2d}}{s_\xi^2\sigma_\beta^2(d)}(\beta-\beta_0)^2.$$

Therefore, under the null hypothesis both noncentrality parameters are zero, $\zeta=\delta=0$; whereas under the alternative $\mathrm{H}_1(>d_0,\beta_0)$ at least one of them is greater than zero and diverges as $n\to\infty$. The asymptotic distribution function of τ_r, denoted in the limit by $G_{\zeta,\delta}(c) := \Pr(\tau_r \le c)$ where $c \ge 0$, has been derived by Abadir and Distaso (2007) and is now specialized to the case of interest:

$$G_{\zeta,\delta}(c) = \frac{\sqrt{c}\exp(-\delta/2)}{2\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{c^{j/2} D_{j-1}^{-}(\zeta)}{j!}$$

$$\times \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{j}{2}+1\right)\left(\frac{\delta c}{4}\right)^{k}}{\Gamma\left(k+\frac{j+3}{2}\right) k!} {}_1F_1\left(k+\frac{1}{2}; k+\frac{j+3}{2}; -\frac{c}{2}\right), \tag{3.4}$$

where $D_{j-1}^{-}(\zeta) := \exp\left(-\zeta^2/4\right) D_{j-1}(\zeta)$ is the modified parabolic cylinder function whose series expansion is derived in Abadir (1993). The gamma function $\Gamma(\cdot)$, Kummer's hypergeometric function ${}_1F_1$, and the (unmodified) parabolic cylinder function $D_{j-1}(\zeta)$ can be found in Erdélyi's (1953) chs.1, 6, and 8, respectively. Their main properties and computational aspects are summarized in Abadir (1999). Each of these "infinite" series are as fast to compute as the exponential series, with terms decaying at an exponential rate.

The distribution under the null hypothesis is readily obtained by calculating

$$\begin{aligned} G_{0,0}(c) &= \sqrt{\frac{\pi c}{2}} \sum_{j=0}^{\infty} \frac{\left(\frac{c}{2}\right)^{j/2} {}_1F_1\left(\frac{1}{2}; \frac{j+3}{2}; -\frac{c}{2}\right)}{\Gamma\left(1-\frac{j}{2}\right)\left[\Gamma\left(\frac{j+1}{2}\right)\right]^2 (j+1)} \\ &= \Phi\left(\sqrt{c}\right) - \frac{1}{2} + \frac{c\sqrt{\pi}}{4} \sum_{j=0}^{\infty} \frac{\left(\frac{c}{2}\right)^{j} {}_1F_1\left(\frac{1}{2}; j+2; -\frac{c}{2}\right)}{\Gamma\left(\frac{1}{2}-j\right) j!\,(j+1)!}, \end{aligned} \tag{3.5}$$

where we have split the first sum into even j and odd j, and used the definition of the standard normal c.d.f. $\Phi(\cdot)$ in terms of ${}_1F_1$ as

$$\Phi(\zeta) = \frac{1}{2} + \frac{\zeta}{\sqrt{2\pi}} {}_1F_1\left(\frac{1}{2}; \frac{3}{2}; -\frac{\zeta^2}{2}\right).$$

The test of size γ is constructed as follows:

reject $\mathrm{H}_0(d_0, \beta_0)$ in favour of $\mathrm{H}_1(> d_0, \beta_0)$ if $\tau_r > c_\gamma$,

where c_γ is the quantile of τ_r defined by $G_{0,0}(c_\gamma) = 1-\gamma$, and we compute the exact (to 2 decimal places) quantiles $c_{1\%} = 8.27$, $c_{5\%} = 5.14$, and $c_{10\%} = 3.81$. The asymptotic power function of the test is obtained by calculating $1 - G_{\zeta,\delta}(c_\gamma)$, which tends to 1 because $|\zeta| \to \infty$ or $\delta \to \infty$ as $n \to \infty$.

Similarly, the test of the hypothesis $\mathrm{H}_0(d_0, \beta_0)$ versus the alternative $\mathrm{H}_1(< d_0, \beta_0)$ is based on the statistic

$$\tau_l := \tau_{d_0}^2 1_{\widehat{d}<d_0} + \tau_{\beta_0}^2(\widehat{d}). \tag{3.6}$$

We reject $\mathrm{H}_0(d_0, \beta_0)$ in favour of $\mathrm{H}_1(< d_0, \beta_0)$ if $\tau_l > c_\gamma$, where the quantiles c_γ are the same as above.

As $n \to \infty$ under $\mathrm{H}_0(d_0, \beta_0)$, we have

$$\Pr(\tau_r > c) \to 1 - G_{0,0}(c), \qquad \Pr(\tau_l > c) \to 1 - G_{0,0}(c). \tag{3.7}$$

We conclude this section with the following consistency result, which follows from applying the result of Theorem 2.1 of Abadir *et al.* (2011) to Theorem 3 in Abadir and Distaso (2007).

THEOREM 3.2. *As $n \to \infty$ under the alternatives $\mathrm{H}_1(> d_0, \beta_0)$ and $\mathrm{H}_1(< d_0, \beta_0)$, we have*

$$\Pr(\tau_r > c_\gamma) \to 1 \text{ and } \Pr(\tau_l > c_\gamma) \to 1,$$

respectively, for any quantile $c_\gamma > 0$ of asymptotic size $1 - G_{0,0}(c_\gamma) = \gamma \in (0,1)$.

Therefore, the tests τ_r and τ_l are consistent.

4 Simulation results

We simulate the following Data Generating Process (DGP)

$$y_t = \alpha + \beta t + u_t, \quad t = 1, \ldots, n,$$

where $u_t \sim \mathrm{I}(d)$, for $\alpha = 0$, $\beta = 0, 0.05, 0.1$, $d = 0.94, 0.96, 0.98, 1$ and $n = 250, 500, 1000$. We consider three different bandwidths for the estimation of d, namely $m = \lfloor n^{0.65} \rfloor, \lfloor n^{0.7} \rfloor, \lfloor n^{0.75} \rfloor$. The DGP is simulated 10,000 times.

The theory covers both persistent $\mathrm{I}(d)$ processes with $d \in (0, 1.5)$ and antipersistent ones with $d \in (-0.5, 0)$. To save space, we include simulation results only for testing the null hypothesis H_0: $d = 1$ and $\beta = 0$ against nearby alternatives, since these are the ones of most interest in economic applications. Other simulation results can be obtained from the authors upon request. The results are reported in Tables 1 to 3, for the three values of m. For a nominal size of 10%, the tables contain size and power for the null hypothesis $\mathrm{H}_0(1, 0)$ for inference on the existence of a unit root and no deterministic trend. The alternative of interest is $\mathrm{H}_1(< 1, 0)$, which covers processes with a finite impulse-response function and possibly a linear trend. The tests used are based on the traditional joint statistic given by

$$\tau = \tau^2_{c,d_0=1} + \tau^2_{\beta_0=0}(\widehat{d}_c), \qquad \tau^2_{c,d_0=1} = 4(\sum_{j=1}^{m} \nu_j^2)(\widehat{d}_c - 1)^2,$$

and the corresponding one-sided modified statistic

$$\tau_l = \tau^2_{c,d_0=1} 1_{\widehat{d}_c < 1} + \tau^2_{\beta_0=0}(\widehat{d}_c)$$

where

$$\widehat{d}_c := \begin{cases} \widehat{d}, & \text{if } \widehat{d} \in [-1/2, 1/2], \\ \widehat{d} + \frac{1}{108}\lambda_m^2, & \text{if } \widehat{d} \in (1/2, 3/2] \end{cases} \tag{4.1}$$

is the version of the FELW estimator $\widehat{d}$ corrected for the bias generated by the estimation procedure in the interval $\widehat{d} \in (1/2, 3/2]$, see Theorem 2.2 in Abadir et al. (2007). Such correction leads to minor improvement of size properties of the tests τ and τ_l in finite samples. Since for $m = o(n^{0.8})$ it holds $\widehat{d}_c = \widehat{d} + o_p(m^{-1/2})$, the tests τ and τ_l based on $\widehat{d}_c$ and $\widehat{d}$ have the same asymptotics properties.

The actual size is read off the top-right corner entry in each block ($n = 250, 500, 1000$) of each table, and is highlighted in italics. Both tests are oversized, due to some negative bias arising from the estimation of d. The size distortion seems to be increasing in m, albeit not uniformly. Size distortions are similar in magnitude for τ and τ_l, except for the largest $m = \lfloor n^{0.75} \rfloor$ when both distortions increase. We therefore do not recommend such a bandwidth unless the sample size is very large.

Powers also are seen to be better for larger values of m. The power gains from our one-sided modified τ_l can be substantial. For example, when $d = 0.94$ (the smallest d that we report) and $\beta = 0$, the power gains of τ_l over τ are of the order of 5–10%. Even in the case of the nearest to the unit root, $d = 0.98$, the power gains remain high at around 4%. These examples show that, even when the alternative to a unit root does not contain a visibly distinguishing trend (i.e., $\beta = 0$), our modified test is able to detect small departures from $d = 1$ better than the traditional $\chi^2(2)$ test.

We now compare the size-adjusted powers of τ and τ_l. It is possible to get a representative idea of the power surfaces of the two tests by looking at Figure 1. The test based on τ_l seems to be uniformly more powerful than its unmodified counterpart τ, especially when violations of the null occur on the long memory parameter. This demonstrates the usefulness of the modified test.

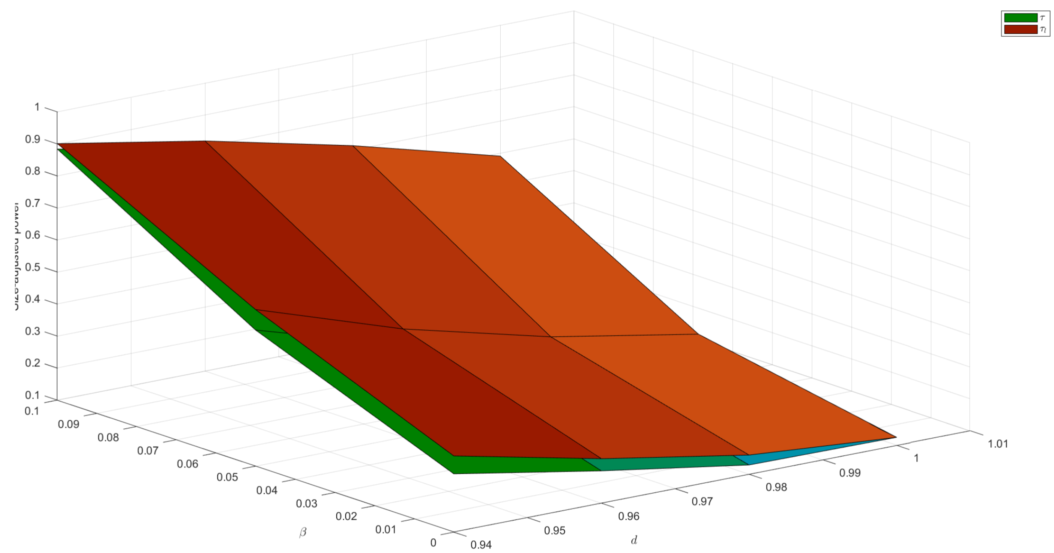

Figure 1: Size-adjusted power surfaces of the tests based on τ (green) and τ_l (red), for $n = 500$ and $m = \lfloor n^{0.75} \rfloor$.

Table 1: Rejection frequencies of $H_0(1,0)$ at the 10% significance level, for τ and τ_l, with $m = \lfloor n^{0.65} \rfloor$.

		$d = 0.94$	$d = 0.96$	$d = 0.98$	$d = 1$
		$n = 250$			
$\beta = 0$	τ	0.178	0.139	0.127	*0.108*
	τ_l	0.226	0.181	0.155	*0.123*
$\beta = 0.05$	τ	0.292	0.236	0.186	0.175
	τ_l	0.352	0.281	0.230	0.201
$\beta = 0.1$	τ	0.559	0.474	0.397	0.343
	τ_l	0.611	0.534	0.458	0.368
		$n = 500$			
$\beta = 0$	τ	0.198	0.145	0.117	*0.116*
	τ_l	0.256	0.188	0.148	*0.121*
$\beta = 0.05$	τ	0.435	0.359	0.269	0.220
	τ_l	0.498	0.418	0.328	0.247
$\beta = 0.1$	τ	0.854	0.755	0.661	0.561
	τ_l	0.887	0.805	0.715	0.602
		$n = 1000$			
$\beta = 0$	τ	0.251	0.164	0.129	*0.111*
	τ_l	0.308	0.221	0.164	*0.117*
$\beta = 0.05$	τ	0.680	0.546	0.422	0.352
	τ_l	0.735	0.612	0.484	0.388
$\beta = 0.1$	τ	0.990	0.967	0.913	0.835
	τ_l	0.995	0.977	0.935	0.856

Table 2: Rejection frequencies of $H_0(1,0)$ at the 10% significance level, for τ and τ_l, with $m = \lfloor n^{0.7} \rfloor$.

		$d = 0.94$	$d = 0.96$	$d = 0.98$	$d = 1$
			$n = 250$		
$\beta = 0$	τ	0.181	0.154	0.122	*0.104*
	τ_l	0.227	0.189	0.149	*0.118*
$\beta = 0.05$	τ	0.293	0.241	0.179	0.171
	τ_l	0.351	0.287	0.222	0.192
$\beta = 0.1$	τ	0.570	0.489	0.406	0.346
	τ_l	0.636	0.547	0.450	0.391
			$n = 500$		
$\beta = 0$	τ	0.222	0.164	0.128	*0.117*
	τ_l	0.292	0.222	0.154	*0.127*
$\beta = 0.05$	τ	0.462	0.349	0.284	0.221
	τ_l	0.528	0.420	0.333	0.244
$\beta = 0.1$	τ	0.873	0.767	0.671	0.538
	τ_l	0.905	0.816	0.709	0.582
			$n = 1000$		
$\beta = 0$	τ	0.296	0.204	0.136	*0.121*
	τ_l	0.374	0.257	0.173	*0.124*
$\beta = 0.05$	τ	0.725	0.579	0.428	0.339
	τ_l	0.785	0.643	0.493	0.380
$\beta = 0.1$	τ	0.993	0.971	0.905	0.828
	τ_l	0.996	0.982	0.931	0.855

Table 3: Rejection frequencies of $H_0(1,0)$ at the 10% significance level, for τ and τ_l, with $m = \lfloor n^{0.75} \rfloor$.

		$d = 0.94$	$d = 0.96$	$d = 0.98$	$d = 1$
		$n = 250$			
$\beta = 0$	τ	0.204	0.150	0.116	*0.115*
	τ_l	0.271	0.191	0.151	*0.121*
$\beta = 0.05$	τ	0.314	0.239	0.195	0.160
	τ_l	0.382	0.306	0.242	0.191
$\beta = 0.1$	τ	0.583	0.490	0.425	0.338
	τ_l	0.661	0.556	0.477	0.372
		$n = 500$			
$\beta = 0$	τ	0.268	0.187	0.115	*0.111*
	τ_l	0.338	0.239	0.160	*0.125*
$\beta = 0.05$	τ	0.511	0.374	0.271	0.211
	τ_l	0.589	0.438	0.323	0.241
$\beta = 0.1$	τ	0.870	0.777	0.658	0.553
	τ_l	0.901	0.819	0.714	0.591
		$n = 1000$			
$\beta = 0$	τ	0.410	0.237	0.124	*0.106*
	τ_l	0.491	0.300	0.170	*0.107*
$\beta = 0.05$	τ	0.754	0.604	0.439	0.338
	τ_l	0.808	0.671	0.499	0.368
$\beta = 0.1$	τ	0.996	0.969	0.908	0.827
	τ_l	0.998	0.979	0.932	0.855

5 Empirical illustration: the case of US quarterly GDP

The methods outlined earlier will now be illustrated with the quarterly series of US GDP. This should be viewed as testing for persistence and trend in a context that is more general than Dickey-Fuller tests. It is *not* an attempt to model GDP, as we will explain that this requires further analysis. Data have been obtained form the Bureau of Economic Analysis and refer to seasonally-adjusted quarterly GDP values, expressed in billions of dollars at 2000 prices. The series is available from the first quarter of 1947 to the third quarter of 2019, avoiding the covid pandemic period which would require further modelling of this exceptional event. We can see from Figure 2 that a linear trend is sufficient to capture any deterministic growth that may be present in the evolution of log(GDP), but we also see a long cycle that our paper is not estimating and could inflate the estimate of our memory parameter d which would mop up the excess volatility. Estimating a cycle is not included in the Dickey-Fuller models either, but was found by Abadir, Caggiano, and Talmain (2013) to be prevalent in macro series. We leave this aspect to further research, to develop the theory of estimating a memory parameter d at a nonzero cyclical spectral frequency when there are deterministic trends.

For $y_t := \log(\text{GDP})$, we fit

$$y_t = \alpha + \beta t + u_t,$$

where $u_t \sim \mathrm{I}(d)$ with $d \in (-1/2, 3/2)$. The parameters of the process have been estimated using LS estimators for α and β and the extended Whittle estimator for d. In particular, d has been estimated using the two recommended bandwidths for this sample size, $m = \lfloor n^{.65} \rfloor, \lfloor n^{.7} \rfloor$.

The resulting estimator for $m = \lfloor n^{.65} \rfloor$ is $\widehat{d} = 1.002$, and the corresponding 95% confidence interval is $(0.8215, 1.1809)$. The fitted process is given by

$$\widehat{y}_t = 7.716 + 0.0079t + \ u_t,$$

where the 95% confidence interval for β is $(0.0062, 0.0095)$. The confidence intervals for d and β are valid for a wide range of processes that lead to Gaussian limits for the estimators.

We caution against the omission of long cycles from the model we fitted here, with the effect that our estimate of d could be inflated. We leave this for future research. The cycles are visible in our Figure 1 (where we also see the trend), and these cycles have been found in Abadir, Caggiano, and Talmain (2013) through the estimation and testing of a parametric autocorrelation function that represents very accurately almost all macroeconomic and aggregate financial series.

It is also of interest to test for the popular hypothesis of driftless unit-root nonstationarity. Using the methodology described in Section 3, the null and alternative hypotheses of interest would be

$$\mathrm{H}_0(1,0) \text{ versus } \mathrm{H}_1(<1,0).$$

The test statistic defined in (3.6) yields

$$\tau_l = 4\Big(\sum_{j=1}^{m} \nu_j^2\Big)(\widehat{d}_c - 1)^2 1_{\widehat{d}_c<1} + \left(\frac{n^{3/2-\widehat{d}}\widehat{\beta}}{\widehat{s}_m(\widehat{d}_c)\sigma_\beta(\widehat{d}_c)}\right)^2 = 83.93$$

and the null hypothesis is massively rejected even at the 1% level.

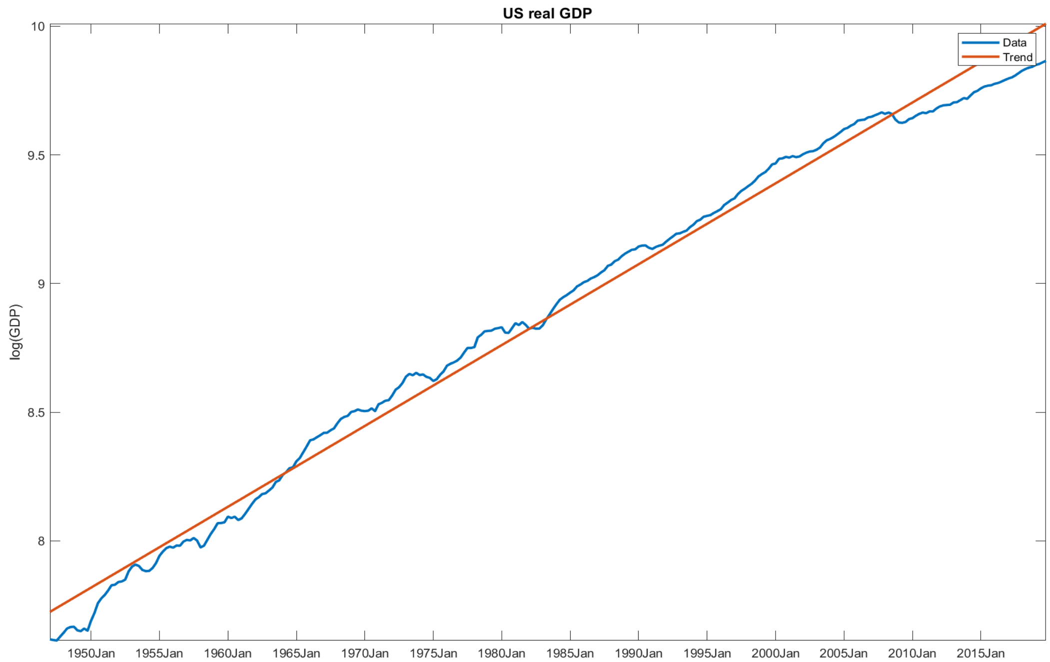

Figure 2: Time series of log(GDP).

6 Proofs

PROOF OF THEOREM 3.1. We assume that $\mathrm{H}_0(d_0, \beta_0)$ is true. Notice that the first claim in (3.3) implies the second claim $\tau \xrightarrow{d} Z_1^2 + Z_2^2$.

By the Cramér-Wold device, to prove the weak convergence of the finite dimensional distributions (3.3), $(\tau_{d_0}, \tau_{\beta_0}(\widehat{d})) \xrightarrow{d} (Z_1, Z_2)$, it suffices to show that for any real numbers a_1, a_2,

$$a_1 \tau_{d_0} + a_2 \tau_{\beta_0}(\widehat{d}) \xrightarrow{d} a_1 Z_1 + a_2 Z_2. \tag{6.1}$$

First we outline some properties of τ_{d_0} and $\tau_{\beta_0}(\widehat{d})$, used in the proof.

By Theorem 2.1 in Abadir *et al.* (2011), we have that $\mathrm{E}(\widehat{\beta} - \beta_0)^2 \leq Cn^{-3+2d_0}$. Now, the estimated residuals can be written as

$$\widehat{u}_t = X_t - \widehat{\alpha} - \widehat{\beta}t = u_t + \alpha - \widehat{\alpha} + g_{n,t}$$

where $g_{n,t} := (\beta - \widehat{\beta})t$. We will estimate d_0 by the FELW estimator $\widehat{d}$ computed using the residuals $\widehat{u}_1, ..., \widehat{u}_n$. Recall that $\sum_{j=1}^m \nu_j^2/m \to 1$. Then,

$$\tau_{d_0} = 2(\sum_{j=1}^m \nu_j^2)^{1/2}(\widehat{d} - d_0) = 2\sqrt{m}(\widehat{d} - d_0) + o_p(1).$$

Since $\widehat{d}$ is data shift-invariant, in estimation of d_0, $\widehat{u}_t$ can be replaced by $u_t + g_{n,t}$. Moreover, since $\sum_{t=1}^n g_{n,t}^2 = O_p(n^{2d_\xi})$, then $g_{n,t}$ satisfies (4.41) of Theorem 2.5 (iii) in Abadir *et al.* (2007), which implies

$$\tau_{d_0} = 2\sqrt{m}(\widehat{d}_u - d_0)(1 + o_p(1)) + o_p(1),$$

where $\widehat{d}_u$ denotes the FELW estimator computed as if $u_0, u_1, ..., u_n$ were observed. Together with (2.5), under Assumption A.1 and $m = o(n^{4/5})$, this implies that the extended local Whittle estimator $\widehat{d}$ has the property

$$\tau_{d_0} = 2\sqrt{m}(\widehat{d}_u - d_0) + o_p(1) \xrightarrow{d} \mathrm{N}(0, 1). \tag{6.2}$$

Together with Theorem 2.1 of Abadir *et al.* (2007) applied to $\widehat{d}_u - d_0$, this implies that

$$\tau_{d_0} = m^{-1/2} \sum_{j=1}^m (\log(j/m) + 1)\eta_j + o_p(1),$$

where $\eta_j = b_0^{-1} \lambda_j^{2d_\xi}\ I_{n,\xi}(\lambda_j)$ and $I_{n,\xi}(\lambda_j) = (2\pi n)^{-1} \left|\sum_{t=1}^n \mathrm{e}^{\mathrm{i}t\lambda_j}\xi_t\right|^2$. Write

$$\eta_j = \eta_{j,\varepsilon} + r_j,$$

where $\eta_{j,\varepsilon} := 2\pi\ I_{n,\varepsilon}(\lambda_j)$ and $r_j := \eta_j - \eta_{j,\varepsilon}$ and set

$$T_{n,\varepsilon} = m^{-1/2} \sum_{j=1}^m \big(\log(j/m) + 1\big)\eta_{j,\varepsilon}, \quad R_n = m^{-1/2} \sum_{j=1}^m \big(\log(j/m) + 1\big) r_j.$$

Robinson (1995, page 1644) showed that, under (3.2) and Assumption A.1, $R_n \xrightarrow{p} 0$, as $n \to \infty$. The latter can be also verified using (6.2.15) of Lemma 6.2.1 in Giraitis *et al.* (2012). Therefore, by (6.2),

$$\tau_{d_0} = T_{n,\varepsilon} + o_p(1). \tag{6.3}$$

Write

$$\begin{aligned} T_{n,\varepsilon} &= n^{-1} \sum_{t,s=1}^n [m^{-1/2} \sum_{j=1}^m (\log(j/m) + 1) e^{i(t-s)\lambda_j}] \varepsilon_t \varepsilon_s \\ &= \sum_{1 \le s < t \le n} c_n(t-s)\varepsilon_t\varepsilon_s + c_n(0) \sum_{t=1}^n \varepsilon_t^2, \end{aligned}$$

where

$$c_s := 2n^{-1} m^{-1/2} \sum_{j=1}^m (\log(j/m) + 1) \cos(s\lambda_j).$$

Property $m^{-1/2} \sum_{j=1}^m (\log(j/m) + 1) = o(1)$ implies

$$c_n(0) = o(n^{-1}), \quad c_n(0) \sum_{t=1}^n \varepsilon_t^2 = o(1)\big(n^{-1} \sum_{t=1}^n \varepsilon_t^2\big) = o_p(1). \tag{6.4}$$

Denote

$$S_{n,\varepsilon} = \sum_{t=1}^{n} z_t$$

where $z_t := \varepsilon_t \zeta_t$, $\zeta_t = \sum_{s=1}^{t-1} c_{t-s}\varepsilon_s$ for $2 \le t \le n$ and is 0 otherwise. Then, $T_{n,\varepsilon} = S_{n,\varepsilon} + o_p(1)$ and therefore (6.3) reduces to

$$\tau_{d_0} = S_{n,\varepsilon} + o_p(1) \xrightarrow{d} Z_1 \sim \mathrm{N}(0,1). \tag{6.5}$$

Next, we evaluate $\tau_{\beta_0}(\widehat{d})$. Note that by (2.8), consistency of the long-run variance estimator (2.7), and (2.4), we get

$$\tau_{\beta_0}(\widehat{d}) = \tau_{\beta_0}(d_0)(1 + o_p(1)) \xrightarrow{d} Z_2 \sim \mathrm{N}(0,1). \tag{6.6}$$

Here, $\tau_{\beta_0} = (s_\xi \sigma_\beta(d_0))^{-1} n^{3/2-d_0}(\widehat{\beta} - \beta_0)$. As in the proof of Theorem 1 in Giraitis *et al.* (2011), we can write

$$n^{3/2-d_0}(\widehat{\beta} - \beta_0) = V_n/D_n, \quad \text{where} \tag{6.7}$$

$$V_n := n^{-3/2-d_0}\sum_{t=1}^{n}(u_t - \bar{u})t, \quad \bar{u} := n^{-1}\sum_{j=1}^{n} u_j, \quad D_n := n^{-3}\sum_{t=1}^{n}(t - \bar{t})^2.$$

Notice that $D_n \to 1/12$. In the proof of Theorem 1 in Giraitis *et al.* (2011), using convergence (2.9), it is shown that for $d_0 \in (-1/2, 1/2)$,

$$\begin{aligned} V_n &= -\sum_{j=1}^{n} n^{-3/2-d_0}\left(\sum_{k=1}^{j}\xi_k - \frac{j}{n}\sum_{k=1}^{n}\xi_k\right) \\ &= -\int_0^1 (Y_n(r) - rY_n(1))\,\mathrm{d}r + o_p(1) \xrightarrow{d} V_\infty := -\int_0^1 (Y_\infty(r) - rY_\infty(1))\,\mathrm{d}r, \end{aligned}$$

whereas for $d_0 \in (1/2, 3/2)$,

$$\begin{aligned} V_n &= -n^{-3/2-d_0}\sum_{j=1}^{n}\sum_{k=1}^{j}\left(\sum_{t=1}^{k}\xi_t - n^{-1}\sum_{s=1}^{n}\sum_{t=1}^{s}\xi_t\right) \\ &= -\int_0^1\int_0^r\left(Y_n(s) - \int_0^1 Y_n(u)\mathrm{d}u\right)\mathrm{d}s\mathrm{d}r + o_p(1) \\ &\quad \xrightarrow{d} V_\infty := -\int_0^1\int_0^r\left(Y_\infty(s) - \int_0^1 Y_\infty(u)\mathrm{d}u\right)\mathrm{d}s\mathrm{d}r \end{aligned}$$

and $Y_n(,)$, $Y_\infty(.)$ are the same as in (2.9).

In view of (6.5) and (6.6)–(6.7), to prove (6.1) it suffices to show that

$$a_1 S_{n,\varepsilon} + a_2 V_n \xrightarrow{d} a_1 Z_1 + a_2 V_\infty. \tag{6.8}$$

In turn, to prove (6.8), in view of the definition of V_n, it suffices to show that for any $m \ge 2$ and $0 < r_1 < r_2 < ... < r_m \le 1$, and real numbers $a_1, b_1, ..., b_m$, it holds

$$a_1 S_{n,\varepsilon} + b_1 Y_n(r_1) + ... + b_m Y_n(r_m) \xrightarrow{d} a_1 Z_1 + b_1 Y_\infty(r_1) + ... + b_m Y_\infty(r_m), \tag{6.9}$$

where $Z_1 \sim \mathrm{N}(0,1)$ and $Y_\infty(r) \sim \mathrm{N}\big(0, \mathrm{E}(Y_\infty^2(r))\big)$ are Gaussian variables such that

$$\mathrm{E}(Z_1 Y_\infty(r)) = 0, \quad 0 < r \le 1. \tag{6.10}$$

Then, (6.10) implies $\mathrm{E}(Z_1 V_\infty) = 0$ and $\mathrm{E}(Z_1 Z_2) = 12\mathrm{E}(Z_1 V_\infty)(s_\xi \sigma_\beta(d_0))^{-1} = 0$.

For $d_0 \in (-1/2, 1/2)$ we have that

$$\mathrm{E}\left(Z_1 V_\infty\right) = -\int_0^1 \mathrm{E}\left[Z_1\left(Y_\infty(r) - rY_\infty(1)\right)\right] \mathrm{d}r = 0,$$

whereas for $d_0 \in (1/2, 3/2)$

$$\mathrm{E}\left(Z_1 V_\infty\right) = -\int_0^1 \int_0^r \mathrm{E}\left[Z_1\left(Y_\infty(s) - \int_0^1 Y_\infty(u)\mathrm{d}u\right)\right] \mathrm{d}s\mathrm{d}r = 0.$$

Proof of (6.9). To this end, set $a_j = 0$ for $j < 0$ in (1.3), denote $n_i = \lfloor nr_i \rfloor + 1$, $i = 1, ..., m$ and define

$$d_{nk,i} := n^{-1/2-d_\xi} \sum_{j=1}^{n_i} a_{j-k}, \quad k \le n_i.$$

We set $d_{nk,i} = 0$ for $n_i < k \le n$. Then,

$$\begin{aligned} Y_n(r_i) &= n^{-1/2-d_\xi} \sum_{j=1}^{n_i} \xi_j = n^{-1/2-d_\xi} \sum_{j=1}^{n_i} \sum_{k=-\infty}^{j} a_{j-k}\varepsilon_k \\ &= \sum_{k=-\infty}^{n_i} d_{nk,i}\varepsilon_k = \sum_{k=-\infty}^{n} d_{nk,i}\varepsilon_k. \end{aligned} \tag{6.11}$$

So, we can write

$$U_n = b_1 Y_n(r_1) + ... + b_m Y_n(r_m) = \sum_{k=-\infty}^{n} \nu_{nk}\varepsilon_k, \quad \nu_{nk} := \sum_{i=1}^{m} b_i d_{nk,i}.$$

Note that by (2.10),

$$\begin{aligned} \mathrm{E}\left(Y_n(r_i)Y_n(r_j)\right) &= \sum_{k=-\infty}^{\min(n_i,n_j)} d_{nk,i}d_{nk,j} \to s_\xi^2 R(t,s) = \mathrm{E}\left(Y_\infty(r_i)Y_\infty(r_j)\right), \\ \mathrm{E}\left(U_n^2\right) &= \sum_{k=-\infty}^{n} \nu_{nk}^2 \to \mathrm{E}\left(U_\infty^2\right) < \infty. \end{aligned} \tag{6.12}$$

Hence, we can write

$$U_n = \sum_{k=-M}^{n} \nu_{nk}\varepsilon_k + \sum_{k=-\infty}^{-M-1} \nu_{nk}\varepsilon_k =: U_{n,1} + U_{n,2},$$

where $M = M(n)$ is selected such that

$$\mathrm{E}\left(U_{n,2}^2\right) = \sum_{k=-\infty}^{-M-1} \nu_{nk}^2 = o(1). \tag{6.13}$$

Hence

$$a_1 S_{n,\varepsilon} + b_1 Y_n(r_1) + ... + b_m Y_n(r_m) = a_1 S_{n,\varepsilon} + U_{n,1} + o_p(1).$$

To prove (6.9), it remains to show that

$$a_1 S_{n,\varepsilon} + U_{n,1} \xrightarrow{d} a_1 Z_1 + U_\infty, \quad U_\infty := b_1 Y_\infty(r_1) + ... + b_m Y_\infty(r_m). \tag{6.14}$$

Defining $v_t := \nu_{nt}\varepsilon_t$ if $-M \le t \le n$, we can write

$$a_1 S_{n,\varepsilon} + U_{n,1} = \sum_{t=-M}^{n} (a_1 z_t + v_t),$$

where $z_t = 0$ for $t \le 2$. Notice that $a_2 z_t + v_t$ is a zero mean martingale difference array with respect to the sigma algebra $\mathcal{F}_{t-1}$ generated by the variables ε_s with $s \le t-1$. Hence, to verify convergence (6.14), it suffices to prove that with Z_1 and $Y_\infty(r)$ as in (6.9) and (6.10), it holds

$$\sum_{t=-M}^{n} \mathrm{E}[(a_1 z_t + v_t)^2 | \mathcal{F}_{t-1}] \xrightarrow{p} \mathrm{E}(a_1 Z_1 + U_\infty)^2 = a_1^2 + \mathrm{E}\left(U_\infty^2\right) \tag{6.15}$$

and

$$\sum_{t=-M}^{n} \mathrm{E}\left[(a_1 z_t + v_t)^2 1_{|a_1 z_t + v_t| \ge \delta}\right] \to 0 \quad \text{for all } \delta > 0. \tag{6.16}$$

Proof of (6.15). Using $(a_1 z_t + v_t)^2 = a_1^2 z_t^2 + v_t^2 + 2a_1 z_t v_t$, we can write

$$\sum_{t=-M}^{n} \mathrm{E}[(a_1 z_t + v_t)^2 | \mathcal{F}_{t-1}] = a_1^2 i_{n,1} + i_{n,2} + 2a_1 i_{n,3},$$

$$i_{n,1} = \sum_{t=2}^{n} \mathrm{E}[z_t^2 | \mathcal{F}_{t-1}], \quad i_{n,2} = \sum_{t=-M}^{n} \mathrm{E}[v_t^2 | \mathcal{F}_{t-1}], \quad i_{n,2} = \sum_{t=2}^{n} \mathrm{E}[z_t v_t | \mathcal{F}_{t-1}].$$

To prove (6.15), it suffices to show that

$$i_{n,1} \xrightarrow{p} 1, \quad i_{n,2} \xrightarrow{p} \mathrm{E}\left(U_\infty^2\right), \quad i_{n,3} \xrightarrow{p} 0. \tag{6.17}$$

Convergence $i_{n,1} \xrightarrow{p} 1$ was shown in (4.12) of Robinson (1995).

Next we evaluate $i_{n,2}$. Using $\mathrm{E}[v_t^2 | \mathcal{F}_{t-1}] = \nu_{nt}^2$, (6.13) and (6.12), we obtain

$$\begin{aligned} i_{n,2} &= \sum_{t=-M}^{n} \nu_{tk}^2 = \sum_{t=-\infty}^{n} \nu_{nt}^2 + o(1) = \mathrm{E}(U_n^2) + o(1) \qquad (6.18) \\ &= \mathrm{E}(b_1 Y_n(r_1) + ... + b_m Y_n(r_m))^2 + o(1) \end{aligned}$$

$$\to \quad \mathrm{E}(b_1 Y_\infty(r_1) + ... + b_m Y_\infty(r_m))^2 = \mathrm{E}(U_\infty^2),$$

which proves (6.17) for $i_{n,2}$.

To bound $i_{n,3}$, note that $\mathrm{E}[z_t v_t | \mathcal{F}_{t-1}] = \mathrm{E}[\zeta_t \nu_{nt} \varepsilon_t^2 | \mathcal{F}_{t-1}] = \zeta_t \nu_{nt}$. Hence,

$$\begin{aligned} i_{n,3} &= \sum_{t=2}^{n} \zeta_t \nu_{nt} = \sum_{t=2}^{n} (\sum_{s=1}^{t-1} c_{t-s} \varepsilon_s) \nu_{nt} = \sum_{s=1}^{n-1} (\sum_{t=s+1}^{n} c_{t-s} \nu_{nt}) \varepsilon_s, \\ \mathrm{E}(i_{n,3}^2) &\le \sum_{s=1}^{n-1} (\sum_{t=s+1}^{n} c_{t-s} \nu_{nt})^2 \le \sum_{t,k=2}^{n} |\nu_{nt} \nu_{nk}| \sum_{s=1}^{\min(t,k)-1} |c_{t-s} c_{t-k}|. \end{aligned}$$

Bounding $|\nu_{nt}\nu_{nk}| \le \nu_{nt}^2 + \nu_{nk}^2$ and noting that in the sum above $1 \le t-s, t-k \le n$, we obtain

$$\mathrm{E}(i_{n,3}^2) \le 2 \sum_{t,k=2}^{n} \nu_{nt}^2 \sum_{s=1}^{\min(t,k)-1} |c_{t-s} c_{t-k}| \le 2 (\sum_{t=2}^{n} \nu_{nt}^2)(\sum_{s=1}^{n} |c_s|)(\sum_{k=1}^{n} |c_k|).$$

By (6.12),

$$\sum_{t=-M}^{n} \nu_{nt}^2 = O(1).$$

In Robinson (1995, equation (4.21)), it is shown that

$$|c_s| = O(m^{-1/2} s^{-1} \log m), \quad s \ge 1, \tag{6.19}$$

which yields

$$\begin{aligned} \max_{1 \le t \le n} \sum_{s=1:s \ne t}^{n} c_{t-s}^2 &\le Cm^{-1} (\log^2 m) \sum_{s=1}^{n} s^{-2} \le Cm^{-1} \log^2 m = o(1), \qquad (6.20) \\ \sum_{s=1}^{n} |c_s| &\le Cm^{-1/2} (\log m) \sum_{s=1}^{n} s^{-1} \le Cm^{-1/2} \log^2 n = o(1) \end{aligned}$$

by assumption (3.2) on m. Hence,

$$\mathrm{E}(i_{n,3}^2) \le Cm^{-1} \log^4 n = o(1),$$

which proves (6.17) for $i_{n,3}$.

Proof of (6.16). Using the inequality $(a+b)^4 \le (2a^2 + 2b^2)^2 \le 8(a^4 + b^4)$, we can bound

$$\begin{aligned} &\sum_{t=-M}^{n} \mathrm{E}\left[(a_1 z_t + v_t)^2 1_{|a_1 z_t + v_t| \ge \delta}\right] \\ &\le \delta^{-2} \sum_{t=-M}^{n} \mathrm{E}\left[(a_1 z_t + v_t)^4\right] \le \delta^{-2} 8 \sum_{t=-M}^{n} \mathrm{E}[a_1^4 z_t^4 + v_t^4] = \delta^{-2} 8 (a_1^4 j_{n,1} + j_{n,2}), \\ &j_{n,1} = \sum_{t=-M}^{n} \mathrm{E}[z_t^4], \quad j_{n,2} = \sum_{t=-M}^{n} \mathrm{E}[v_t^4]. \end{aligned}$$

It suffices to show that

$$j_{n,1} \to 0, \quad j_{n,2} \to 0. \tag{6.21}$$

We start with the proof of the first claim. Recall that $\{\varepsilon_t\}$ are i.i.d. random variables, and $\mathrm{E}\left(\varepsilon_1^4\right) < \infty$. Therefore, for $z_t = \varepsilon_t \zeta_t$,

$$j_{n,1} = \mathrm{E}\left(\varepsilon_1^4\right) \sum_{t=1}^{n} \mathrm{E}\left(\zeta_t^4\right).$$

Since by (6.26),

$$\mathrm{E}\left(\zeta_t^4\right) = \mathrm{E}[(\sum_{s=1}^{t-1} c_{t-s}\varepsilon_s)^4] \le C(\sum_{s=1}^{t-1} c_{t-s}^2)^2,$$

where C does not depend on t and c_{t-s}, we obtain

$$j_{n,1} \le C \sum_{t=1}^{n} (\sum_{s=1}^{t-1} c_{t-s}^2)^2 \le (\sum_{t=2}^{n} \sum_{s=1}^{t-1} c_{t-s}^2)(\max_{t=1,\ldots,n} \sum_{s=1:s\neq t}^{n} c_{t-s}^2). \tag{6.22}$$

In the proof of (4.12) of Robinson (1995), it was shown that

$$\sum_{t=2}^{n} \mathrm{E}\left(z_t^2\right) = \sum_{t=2}^{n} \sum_{s=1}^{t-1} c_{t-s}^2 \to 1. \tag{6.23}$$

Applying (6.23) and (6.20) in (6.22), we obtain $j_{n,1} \to 0$.

Using definition $v_t = \nu_{nt}\varepsilon_t$, we have

$$j_{n,2} = \sum_{t=-M}^{n} \mathrm{E}\left(v_t^4\right) = \mathrm{E}\left(\varepsilon_1^4\right) \sum_{t=-M}^{n} \nu_{nt}^4 \le C(\sum_{t=-\infty}^{n} \nu_{nt}^2)(\max_{t\le n} \nu_{nt}^2). \tag{6.24}$$

Property (6.12) implies that the standardized sum $Y_n(r_i)$ of linear process ξ_t given by (6.11) has the property

$$\mathrm{E}[Y_n^2(r_i)] = \mathrm{E}[(n^{-1/2-d_\xi} \sum_{j=1}^{n_i} \xi_j)^2] = \sum_{k=-\infty}^{n} d_{nk,i}^2 \to \mathrm{E}[Y_\infty^2(r_i)] < \infty.$$

In Abadir *et al.* (2014, equation (2.4)), it is shown that this fact implies $\max_{k\le n} |d_{nk,i}| = o(1)$ which, in turn, implies

$$\max_{t\le n} \nu_{nt}^2 = \left(\max_{t\le n}(\sum_{i=1}^{m} |b_i d_{nk,i}|\right)^2 \to 0. \tag{6.25}$$

Using (6.18) and (6.25) in (6.24) we obtain $j_{n,2} \to 0$ which proves (6.21) and concludes the proof of the theorem. □

LEMMA 6.1. *(Abadir et al.(2014), Lemma 3.1). Let* $s_n = \sum_{j=1}^{n} b_j \varepsilon_j$ *where* $\{\varepsilon_j\}$ *are i.i.d. random variables with zero mean,* $\mathrm{E}\left(\varepsilon_1^4\right) < \infty$*, and* $\{b_j\}$ *with* $j \geq 1$ *is a sequence of real numbers. Then,*

$$\mathrm{E}\left(s_n^4\right) \leq C(\mathrm{E}\left(s_n^2\right))^2 \leq C(\sum_{j=1}^{n} b_j^2)^2, \tag{6.26}$$

where C *does not depend on* n *and* b_j*.*

Acknowledgement. We would like to the thank the Associate Editor and two Referees for constructive comments on an earlier version of the paper, and Violetta Dalla for expert advice on simulations. We are grateful for the comments of seminar participants at the Bank of Italy, Bilbao, Econometric Society World Congress (London), GREQAM, Liverpool, LSE, Oxford, Queen Mary, Tilburg, Tinbergen Institute, Vilnius Academy of Science, York. This research is supported by the ESRC grants R000239538, RES000230176, and RES062230790.

REFERENCES

Abadir, K.M., 1993. Expansions for some confluent hypergeometric functions. Journal of Physics A 26, 4059–4066 (Corrigendum for printing error, 1993, 7663).

Abadir, K.M., 1999. An introduction to hypergeometric functions for economists. Econometric Reviews 18, 287–330.

Abadir, K.M., Caggiano, G. and Talmain, G., 2013. Nelson-Plosser revisited: the ACF approach. Journal of Econometrics 175, 22–34.

Abadir, K.M., Distaso, W., 2007. Testing joint hypotheses when one of the alternatives is one-sided. Journal of Econometrics 140, 695–718.

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.

Abadir, K.M., Distaso, W., Giraitis, L., Koul, H.L., 2014. Asymptotic normality for weighted sums of linear processes. Econometric Theory 30, 252–284.

Cheung, Y.L., Hassler, U., 2018. Discontinuity of fully extended (local) Whittle estimation. Available at SSRN 3120575.

Cheung, Y.L., 2020. Nonstationarity-extended Whittle estimation with discontinuity: a correction. Economic Letters 187, article 108914.

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

Dickey, D.A., Fuller, W.A., 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49, 1057–1072.

Dolado, J.J., Gonzalo, J., Mayoral, L., 2002. A fractional Dickey-Fuller test for unit roots. Econometrica 70, 1963–2006.

Dolado, J.J., Gonzalo, J., Mayoral, L., 2008. Wald tests of I(1) against I(d) alternatives: Some new properties and an extension to processes with trending components. Studies in Nonlinear Dynamics and Econometrics 12, 1–32.

Dolado, J.J., Gonzalo, J., Mayoral, L., 2009. Simple Wald tests of the fractional integration parameter: an overview of new results. In The Methodology and Practice of Econometrics, edited by J. Castle and N. Shephard, Oxford University Press, Oxford.

Erdélyi, A. (Ed.), 1953. Higher Transcendental Functions, volumes 1–2. McGraw-Hill, New York.

Gil-Alaña, L.A., Robinson, P.M., 1997. Testing of unit roots and other nonstationary hypotheses in macroeconomic time series. Journal of Econometrics 80, 241–268.

Giraitis, L., Koul, H.L., Surgailis, D., 2012. *Large Sample Inference for Long memory Processes.* Imperial College Press, London.

Künsch, H., 1987. Statistical aspects of self-similar processes, in Yu.A. Prohorov, V.V. Sazonov (eds). Proceedings of the 1st World Congress of the Bernoulli Society, Vol.1, Science Press, Utrecht, 67–74.

Lobato, I., Velasco, C., 2006. Optimal fractional Dickey–Fuller tests for unit roots. Econometrics Journal 9, 492–510.

Nelson, C.R., Plosser, C.I., 1982. Trends and random walks in macroeconomic time series: some evidence and implications. Journal of Monetary Economics 10, 139–162.

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

Robinson, P.M., 1995. Gaussian semiparametric estimation of long range dependence. Annals of Statistics 23, 1630–1661.

Robinson, P.M., 2005. Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction. Econometric Theory 21, 171–180.