THE ROLE OF DETERMINISTIC COMPONENTS IN THE FRACTIONAL DICKEY-FULLER TEST FOR UNIT ROOTS

BY JUAN J. DOLADO[a], JESUS GONZALO[b], AND LAURA MAYORAL[c] *

[a]Dept. of Economics, Universidad Carlos III de Madrid.

[b]Dept. of Statistics and Econometrics, Universidad Carlos III de Madrid.

[c]Dept. of Economics, Universidad Autonoma de Barcelona.

July, 2002

Abstract

This paper discusses the role of deterministic components in the DGP and in the auxiliary regression model which underlies the implementation of the Fractional Dickey-Fuller (FDF) test for $I(1)$ against $FI(d)$ processes with $d \in [0, 1)$. It is found that most of the results obtained for the standard DF test in the $I(1)$ *vs.* $I(0)$ framework remain valid in our framework. An exception is that the introduction of a linear trend in the regression, when the DGP is an I(1) process with drift, leads to an asymptotically normal distribution of the test instead of a nonstandard distribution, as is the case with the DF test. A simple testing strategy entailing only aymptotically normally-diistributed tests is proposed and an empirical application is provided where the FDF test with deterministic components is used to test for long-memory in the per capita GDP of several OCDE countries, a property that could explain a puzzle in the empirical literature on convergence and growth.

KEYWORDS: Deterministic components, Dickey-Fuller test, Fractionally Dickey-Fuller test, Fractional processes, Long memory, Trends, Unit roots.

---

*Corresponding E-mail:jgonzalo@est-econ.uc3m.es.

We are grateful to Claudio Michelacci and Bart Verspagen for providing us with their data bases and to Francesc Mármol and participants in a seminar at Ente Luigi Einaudi (Rome) for useful comments.

INTRODUCTION

In a recent paper, Dolado, Gonzalo and Mayoral (2002, DGM henceforth) have proposed an extension of the well-known Dickey-Fuller (DF henceforth) test of $I(1)$ against $I(0)$ processes to the more general framework of testing for a unit root against long-range dependence. Relying upon the DF approach, the underlying idea is to run a potentially unbalanced regression where the dependent variable is the filtered series under the null and the explanatory variable is the lagged value of filtered series under the alternative hypothesis. Specifically, within the class of fractionally-integrated, $FI(d)$, processes, the so-called Fractional Dickey-Fuller test (FDF test henceforth) tests the null hypothesis of $d = 1$ against the alternative of $0 \leq d < 1$ by running the OLS regression $\Delta y_t = \phi\ \Delta^d y_{t-1} + \varepsilon_t$, where ε_t is an i.i.d. disturbance, where $\Delta = (1 - L)$ and L is the lag operator. If the error term under the null hypothesis happens to be autocorrelated, then the above regression should be augmented by a suitable number of lags of Δy_t in the RHS of the equation, giving rise to the Augmented Fractional Dickey-Fuller test (AFDF, henceforth). The regressor $\Delta^d y_{t-1}$ is constructed by applying the truncated binomial expansion of the filter $(1 - L)^d$ to y_{t-1}, so that $\Delta^d y_t = \sum_0^{t-1} \pi_i(d)\ y_{t-i}$ where $\pi_i(d)$ is the i-th coefficient in that expansion. The FDF test is based upon the t-ratio of $\hat{\phi}_{ols}$, t_ϕ, so that non-rejection of H_0: $\phi = 0$ against H_1: $\phi < 0$, implies that the process is $I(1)$, namely, $\Delta y_t = \varepsilon_t$. Conversely, rejection of the null implies that the process is $FI(d)$, $0 \leq d < 1$, that is, $\Delta^d y_t = C(L)\varepsilon_t$, where the lag polynomial $C(L)$ has absolutely summable coefficients and a finite gain, such that $C(0) = 1$ and $C(1) = -\phi^{-1} > 0$. The basic results derived by DGM depend on whether the value of d under the alternative is assumed to be known or needs to be estimated as, for instance, in the case where the alternative hypothesis is a composite one. In the first case, it can be proved that the asymptotic distribution of the t_ϕ is a $N(0,1)$ variate when $0.5 \leq d < 1$, whilst it is nonstandard, i.e., a functional of Fractional Brownian motion, when $0 \leq d < 0.5$. In the second case, whenever d is pre-estimated using any $T^{1/2}-$consistent estimator of $d \in [0, 1)$, the asymptotic distribution of the t_ϕ is always $N(0,1)$ for any value of d within the pre-specified range. The intuition of these results is that whenever the values of d under the null and the alternative hypothesis are close (i.e., when d belongs to the nonstationary range or when d is estimated using a consistent estimator) asymptotic normality follows under the null hypothesis, whereas when they are far away from each other (i.e., when d belongs to the stationary range) the limiting distributions are nonstandard. Fractiles for the nonstandard distributions are reported in DGM (see Tables X, XI and XII) for the cases where the DGP contains no deterministic components and the regression model is either driftless (denoted as FDF distribution), contains a constant term ($FDF^{\mu}_{t_\phi}$), or contains both a drift and a linear trend ($FDF^{\tau}_{t_\phi}$). The advantages of this test, in parallel with the DF approach, rely on its simplicity and on its good performance, in terms of size and power, in finite samples when compared to other available tests in the literature (see DGM).

Following the development of unit root tests in the past, where the initial canonical AR(1) model was subsequently augmented with further deterministic terms (including linear, non-linear and broken trends), our goal in this paper is precisely to investigate how the limiting distribution of the FDF test changes when deterministic components are considered in the DGP and in the regression model. In particular, we will focus on the role of *drifts* and *linear trends*. After all, most (macro) economic time series have a trending behaviour, which should be carefully treated when extracting the stochastic component of the series which is commonly subject to unit root tests. In the $I(1)$ *vs.* $I(0)$ framework, a constant term and a linear trend are typically included in the auxiliary regression model in such cases so that, if a unit root exists, the constant term becomes a trend under the null hypothesis. As DF (1981) showed, including the linear trend in the model permits to capture the trending behavior of the series even if, under the alternative, the root lies in the stationary range. In this context, as West (1988) highlighted a particularly interesting case that arises when both the DGP and the model share the same deterministic terms. For example, if the DGP is $\Delta y_t = \mu + \varepsilon_t$ and the model is $\Delta y_t = \alpha + \phi y_{t-1} + \varepsilon_t$, then t_ϕ is asymptotically normal since y_t can be decomposed into the sum of a linear trend, whose variability is of order $O_p(T^3)$ and a stochastic trend whose variability is of order $O_p(T^2)$, so that the former dominates the latter. Hence, the regressor y_{t-1} in the auxiliary equation of the DF test behaves like a linear trend and from Theil (1971) we know that the asymptotically normality results of least squares hold whenever the variables are linear trends, leading in this case to the result that $T^{3/2}\ \hat{\phi}_{ols}$ has a limiting distribution corresponding to a $N(0, 12/\mu^2)$. Of course, if a linear trend, t, is included in the auxiliary model, t and the linear trend embedded in y_{t-1} will become colinear and the stochastic component of y_{t-1} (a random walk in this case) will dominate its limiting behavior, leading to the traditional nonstandard DF distribution for random $I(1)$ series with nonzero drift.

Along the lines of above discussion, we will proceed in the rest of the paper to analyze whether the results derived for the standard DF test when deterministic terms are present both in the DGP and in the model still remain valid when considering the more general FDF test. Specifically, we will start from the following auxiliary regression model which underlies the implementation of the FDF test:

$$\Delta y_t = \mu\,(t) + \phi \Delta^d y_{t-1} + a_t,$$

where $\mu\,(t)$ represents some deterministic components, typically a constant (α) and a linear time trend (δt). As in the DF case, the asymptotic behaviour of t_ϕ depends on the nature of $\mu(t)$ and whether the true DGP also contains deterministic components or not. Abstracting from the case where $\mu(t) = 0$ both in the model and in the DGP, which is the case studied at length in DGM, three situations can be considered, namely, *Case I*: the DGP is a driftless random walk and the auxiliary regression contains a constant or a constant and a linear trend; *Case II*: the DGP is a random walk with drift and the regression also contains a constant term; and *Case III*: the DGP is also a random walk with drift and the regression model contains a constant term and a time trend.

A preview of our main conclusions can be summarized as follows:

- *Case I:* The results are similar to those found for the FDF test when $\mu(t) = 0$. The same type of nonstandard limiting distributions (functionals of fractional brownian motions –demeaned or detrended in this case– when $0 \leq d < 0.5$, and normal distributions when $0.5 \leq d < 1$) are found when d is assumed to be *a priori* known. When a $\mathrm{T}^{1/2}-$consistent estimate of d is used as a pre-estimate, the asymptotic distribution of the t_ϕ is always $N(0,1)$.

- *Case II*: The results are similar to those found for the DF test, namely, the existence of a drift in the DGP implies that the asymptotic distributions are standard for all $d \in [0,1)$, as in the case where a $\mathrm{T}^{1/2}-$consistent estimate of d is being used. However, in finite samples, if the drift is small relative to the standard deviation of the error term, the $N(0,1)$ limiting distribution may present size distortions so that for small values of the drift it behaves closely to the $FDF_{t_\phi}^{\mu}$ distribution .

- *Case III*: The only difference with respect to the DF test is found in this third case. In the DF framework ($d = 0$), the inclusion of a time trend in the regression generates non-standard distributions as in Case I. By contrast, we find that the limiting distributions are still normal for any $d \in (0,1)$ used in the regression. The intuition behind this result is that while in the DF case, the term t and y_{t-1} are colinear in large samples, and this cancels the effect of the deterministic components in the DGP, the terms t and $\Delta^d y_{t-1}$ are no longer colinear in the FDF case and therefore the linear trend is still the leading term in the regression, generating standard limiting distributions, as in Case II. Similarly, the relative size of the drift to the standard deviation of the error term under the null of $I(1)$, may lead to strong divergences between the finite-sample distribution and the limiting one.

Table 1 gathers the above conclusions in compact way.

TABLE 1

ASYMPTOTIC DISTRIBUTIONS OF $t_{\hat{\phi}_{ols}}$ UNDER THE NULL HYPOTHESIS

DGP: $\Delta y_t = \alpha + \varepsilon_t$			
	$\alpha = 0$		
d | Regression	$\phi \Delta^d y_{t-1} + a_t$	$\alpha + \phi \Delta^d y_{t-1} + a_t$	$\alpha + \delta t + \phi \Delta^d y_{t-1} + a_t$
Fixed $d \in [0, 0.5)$	FDF	$FDF^{\mu}_{t_\phi}$	$FDF^{\tau}_{t_\phi}$
Fixed $d \in [0.5, 1)$	$N(0,1)$	$N(0,1)$	$N(0,1)$
Estimated d	$N(0,1)$	$N(0,1)$	$N(0,1)$
	$\alpha \neq 0$		
Fixed d	–	$N(0,1)$	$DF^{\tau}_{t_\phi}$ if $d = 0$ $N(0,1)$ $d \in (0,1)$
Estimated d	–	$N(0,1)$	$N(0,1)$

Note: Critical values of the FDF, $FDF^{\mu}_{t_\phi}$ and $FDF^{\tau}_{t_\phi}$ distributions can be found in Tables 11.a, b and c of DGM (2002).

It is important to note at this stage that, although the analysis in the sequel is restricted to the case where the DGP is a driftless random walk or a random walk with drift −and therefore the error term is i.i.d− the asymptotic results obtained in this paper remain valid when the disturbance is allowed to be autocorrelated, as it happens in the ADF case. In this respect, DGM (Theorems 6 and 7) have proved that, in order to remove the correlation, it is sufficient to augment the set of regressors in the auxiliary regression described above with k lags of the dependent variable such that $k \uparrow \infty$ as $T \uparrow \infty$,and $k^3/T \uparrow 0$, as in Said and Dickey (1984). As mentioned above, this procedure gives rise to the AFDF test which will be used in the empirical section below and whose properties, being the same as those of the FDF test, are omitted to save space.

It should be remarked at the outset that our results provide a statistical foundation for a testing strategy of $H_0 : d = 1$ vs. $H_0 : 0 < d < 1$ when d is taken to be unknown and therefore needs to be *a priori* estimated to implement the AFDF test in the presence of deterministic terms. In contrast to the usual methods to deal with deterministic terms in applied work on long-memory processes, we do not either need to pre-filter the series or to use a different set of critical values as in the DF/ADF test for I(1) vs. I(0) processes.

The rest of the paper is structured as follows. Sections 2 analyzes each of the three cases described above, when the value of d needed to make the FDF test feasible is assumed to be fixed, as in the DF case where the value of d under the alternative is $d = 0$. Emphasis is placed on the differences with the standard results for the DF test and on identifying those key parameters that may lead to poor finite-sample behaviour of the limiting distributions. Each section contains detailed Monte-Carlo evidence. Further, we also analyze how the previous results change when d needs to be pre-estimated to make the FDF test feasible. As in DGM, pre-estimation of d through a $T^{1/2}$-consistent estimator yields asymptotic normality in all the three cases. Section 3 discusses some empirical applications of the previous tests. Finally, Section 4 draws some concluding remarks.

Proofs of theorems and lemmae are collected in an Appendix.

In the sequel, the definition of a $FI(d)$ process that we will adopt is that of an (asymptotically) stationary process, when $d < 0.5$ and that of a non-stationary (truncated) process, when $d \geq 0.5$. Those definitions are similar to those used in, e.g., Robinson (1994) and Tanaka (1999) and are summarized in Appendix A of DGM (2002). Moreover, the following conventional notation is adopted throughout the paper: $\Gamma(.)$ denotes the gamma function, $\{\pi_i(d)\}$ represents the sequence of coefficients associated to the expansion of Δ^d in powers of L and are defined as

$$\pi_i(d) = \frac{\Gamma(i-d)}{\Gamma(-d)\,\Gamma(i+1)}.$$

The indicator function is denoted by $1_{(.)}$ and I_n is the identity matrix of order n. $B(.)$ represents standard Brownian motion (BM), whereas $W_d(.)$ and $B_d(.)$ are standard fractional Brownian motions (FBM) corresponding to the limit distributions of the standardized partial sums of stationary and asymptotically stationary (truncated) $FI(d)$ processes, respectively, as defined in Marinucci and Robinson (1999). Finally, $\xrightarrow{w}$ and $\xrightarrow{p}$ denote weak convergence and convergence in probability, respectively.

DETERMINISTIC TERMS IN THE DGP AND IN THE REGRESSION MODEL

DF developed a procedure for testing the null hypothesis that a given series y_t has a unit root with or without drift versus the alternative that the series is a stationary I(0) process about a deterministic trend. Here we consider the same null hypothesis but we enlarge the set of possible alternatives by considering that the series is a FI(d) process, with $d < 1$, possibly around a constant or a constant and a linear time trend. More explicitly, we will consider the following DGP's under the null hypothesis

$$DGP\ 1: y_t = y_{t-1} + \varepsilon_t,\ t \geq 1, \tag{1}$$

and

$$DGP\ 2\text{: } y_t = \alpha + y_{t-1} + \varepsilon_t,\ t \geq 1, \tag{2}$$

according to the case where the process is a random walk without or with drift respectively. In order to test this hypothesis, two regression models will be estimated, one including a constant term and another including a constant term and a time trend:

$$\textit{Regression Model 1}: \Delta y_t = \alpha + \phi \Delta^{d} y_{t-1} + a_t, \tag{3}$$

or

$$\textit{Regression Model 2}: \Delta y_t = \alpha + \delta t + \phi \Delta^{d} y_{t-1} + a_t, \tag{4}$$

where a_t is an $I(0)$ process. If $a_t = \varepsilon_t$, ε_t being an *i.i.d.* process, then y_t is a driftless random walk when $\alpha = \delta = \phi = 0$. Conversely, when $\phi < 0$ and $\delta = 0$, Regression Model 1 can be expressed, under the alternative hypothesis that $\phi < 0$, as,

$$\left(\Delta^{1-d} - \phi L\right) \Delta^d y_t = \alpha + \varepsilon_t.$$

As proved by DGM, the polynomial $\Pi(z) = \left((1-z)^{1-d} - \phi z\right)$ has absolutely summable coefficients and, under the restriction $-2^{1-d} < \phi < 0$, all its roots are outside the unit circle. Therefore Regression Model 1 can be re-written as:

$$\Delta^d y_t = C(L)(\alpha + \varepsilon_t),$$

where $C(L) = \Pi(L)^{-1}$, with $C(0) = 1$ and $0 < C(1) = -\phi^{-1} < \infty$. Since we are considering truncated processes, this implies that:

$$\Delta^d y_t = \alpha \sum_{i=0}^{t-1} c_i + C(L)\varepsilon_t,$$

where $lim \sum_{i=0}^{t-1} c_i = C(1) < \infty$. Hence, $\Delta^d y_t$ is (asymptotically) an $I(0)$ process with a non-zero drift, given by $\alpha^* = \alpha C(1)$.

Likewise if $\delta \neq 0$, then the corresponding reformulation of Regression Model 2 becomes

$$\left(\Delta^{1-d} - \phi L\right) \Delta^d y_t = \alpha + \delta t + \varepsilon_t,$$

implying that,

$$\begin{aligned} \Delta^d y_t &= \alpha \sum_{i=0}^{t-1} c_i + \delta \sum_{i=0}^{t-1} c_i(t-i) + C(L)\varepsilon_t \\ &= \alpha \sum_{i=0}^{t-1} c_i - \delta \left(\sum_{i=0}^{t-1} i c_i\right) + \left(\sum_{i=0}^{t-1} c_i\right) \delta t + C(L)\varepsilon_t. \end{aligned} \tag{5}$$

The order of magnitude of each of the three components of $\Delta^d y_t$ is as follows. The first term, as stated before, is $O(1)$ since it tends to a bounded constant $\alpha^* = \alpha C(1)$. The second term $\sum_{i=0}^{t-1} ic_i$ is $O\left(t^d\right)$. To check the latter, notice that $\sum_{i=0}^{t-1} ic_i$ is the sum of the first $(t-1)$ terms of $C'(1)$. This last expression is given by,

$$C'(z) = \frac{-(1-d)(1-z)^{-d} + \phi}{\left((1-z)^{1-d} - \phi z\right)^2},$$

When evaluated in $z = 1$, the denominator is a bounded quantity. As for the numerator

$$(1-z)^{-d}\,|_{z=1}\, 1_{(t>0)} = O\left(t^d\right),$$

since the coefficients associated to the expansion of this polynomial, $\pi_i(-d)$ are equivalent to i^{-1+d} for large i. This implies that $C'(1) = O\left(t^d\right)$. Then, let us define the sequence $\{\varphi_1, ..., \varphi_t, ...\}$, where

$$\varphi_t = \left(\sum_{i=0}^{t-1} ic_i\right)/t,$$

so that $\sum_{i=0}^{t-1} ic_i = \varphi_t t$ for each t. It is clear that the limit of the sequence $\{\varphi_t\}$, φ, is zero for all $d < 1$. Hence, the process y_t can be rewritten as:

$$\Delta^d y_t = \alpha^* + \delta^* t + C(L)\varepsilon_t + o(1)$$

where $\alpha^* = \alpha C(1)$, $\delta^* = \delta C(1)$. This implies that the process y_t is (asymptotically) a $FI(d)$ process with a constant and a linear time trend. This structure is very similar to the one obtained in the DF case for which it is found that under the alternative hypothesis the process y_t is given by

$$y_t = \alpha + \delta t + (1-\rho L)^{-1}\varepsilon_t.$$

In the following subsections we study the behaviour of the test statistics associated to the above mentioned regression models under the null hypothesis of a random walk with and without drift.

Case I: Deterministic terms included in the regression. DGP is a driftless random walk.

In this subsection the following DGP is considered:

$$DGP\ 1: y_t = y_{t-1} + \varepsilon_t,\ t \geq 1, \tag{6}$$

whereas Regression Models 1 and 2 in equations (3) and (4), respectively, are being estimated.

In order to obtain the limiting distributions of $\hat{\phi}_{ols}$ and t_ϕ, the following lemmae, proved by DGM in order to justify the use of truncated $FI(d)$ processes in the FDF test, are reproduced for completeness

Lemma 1 *Let $\{\varepsilon_t\}$ be a sequence of zero-mean i.i.d. random variables with variance σ_ε^2 such that $E\left|\varepsilon_t^4\right| < \infty$ and consider the following linear processes*

$$\Delta^d x_t = \varepsilon_t \qquad d \in [-0.5, 0.5),$$

$$\Delta^d x_t^* = \varepsilon_t 1_{(t>0)} \qquad d \in [-0.5, 0.5),$$

and

$$z_t^* = \sum_{i=1}^{t} x_i^*.$$

Then, the following processes verify:

- *if $-0.5 < d < 0.5$,*

$$T^{-1} \sum_{t=1}^{T} (x_t - x_t^*) = o_p(1), \tag{7}$$

$$T^{-1} \sum_{t=1}^{T} \left(x_t^2 - x_t^{*^2}\right) = o_p(1), \tag{8}$$

and

$$T^{-1} \sum_{t=1}^{T} \left(x_t x_{t+k} - x_t^* x_{t+k}^*\right) = o_p(1). \tag{9}$$

- *if $d = -0.5$,*

$$(T \log T)^{-1} \sum_{t=1}^{T} z_t^{*^2} \xrightarrow{p} \frac{\sigma^2}{\pi}. \tag{10}$$

- *if $-0.5 < d < 0.5$,*

$$T^{-2(1+d)} \sum_{t=1}^{T} z_t^{*^2} \xrightarrow{w} \int_0^1 B_d^2(r)\, dr. \tag{11}$$

Lemma 2 *Let ε_t, x_t and x_t^* and z_t^* be defined as in Lemma 1. Then the following processes are martingale differences and verify*

- *if* $0 < d < 0.5$,

$$T^{-1/2}\sum_{t=2}^{T} x_{t-1}\varepsilon_t \xrightarrow{w} N\left(0, \sigma^4 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}\right), \tag{12}$$

$$T^{-1/2}\sum_{t=2}^{T} x^*_{t-1}\varepsilon_t \xrightarrow{w} N\left(0, \sigma^4 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}\right). \tag{13}$$

- *if* $d = -0.5$,

$$(T\log T)^{-1/2}\sum_{t=2}^{T} z^*_{t-1}\varepsilon_t \xrightarrow{w} N\left(0, \frac{\sigma^4}{\pi}\right) \tag{14}$$

- *if* $-0.5 < d < 0$,

$$T^{-(1+d)}\sum_{t=2}^{T} z_{t-1}\varepsilon_t \xrightarrow{w} \sigma^2 \int_0^1 B_d(r)\,dB(r), \tag{15}$$

$$T^{-(1+d)}\sum_{t=2}^{T} z^*_{t-1}\varepsilon_t \xrightarrow{w} \sigma^2 \int_0^1 B_d(r)\,dB(r). \tag{16}$$

Upon the use of the previous lemmae, the next theorems state the consistency of the estimators of ϕ in Regression Models 1 and 2 and present the asymptotic distributions of these estimators and their corresponding $t - ratios$. As in the simple FDF case with no deterministic components, a different asymptotic behaviour is found to hold depending on what value of d is used to run the regression. For the values of d in the stationary range $(0 \leq d < 0.5)$, the limiting distributions are nonstandard and the statistics converge to the demeaned and detrended version of the FDF distribution (as it is the case in the standard DF case). By contrast, for the remaining values of d in the nonstationary range $(0.5 \leq d < 1)$, the asymptotic distributions are standard and resemble, after demeaning and detrending, those obtained for the simple FDF case with no deterministic components. The intuition for this result, as explained in DGM, relies upon the proximity between the hypothesized value of d and its value under the null $(d = 1)$. When both values of d are close (i.e., when d belongs to the nonstationary range) asymptotic normality follows, whereas when they are far away from each other (i.e., when d belongs to the stationary range) the limiting distributions are nonstandard.

Theorem 1 *Under the null hypothesis that y_t is a random walk, the OLS coefficient associated to ϕ in* (3), $\hat{\phi}^{\mu}_{ols}$, *is a consistent estimator of $\phi = 0$ and converges to its true value at a rate T^{1-d} when $0 \leq d < 0.5$, $(T\log T)^{1/2}$ when $d = 0.5$, and at the standard rate $T^{1/2}$*

when $0.5 < d < 1$. *Its asymptotic distribution is given by*

$$T^{1-d}\hat{\phi}^{\mu}_{ols} \xrightarrow{w} \frac{\int_0^1 B^{\mu}_{-d}(r)\, dB(r)}{\int_0^1 \left(B^{\mu}_{-d}(r)\right)^2 dr} \qquad \text{if } 0 \leq d < 0.5, \tag{17}$$

$$(T\log T)^{1/2}\hat{\phi}^{\mu}_{ols} \xrightarrow{w} N(0,\pi) \qquad \text{if } d = 0.5, \tag{18}$$

and

$$T^{1/2}\hat{\phi}^{\mu}_{ols} \xrightarrow{w} N\left(0, \frac{\Gamma^2(d)}{\Gamma(2d-1)}\right) \qquad \text{if } 0.5 < d < 1, \tag{19}$$

where $B^{\mu}_{-d}(s)$ *is a demeaned fractional browian motion defined as* $B^{\mu}_{-d}(s) = B_{-d}(s) - \int_0^1 B_{-d}(r)\, dr$.

Theorem 2 *Under the null hypothesis that* y_t *is a random walk, the asymptotic distribution of the* $t-statistic$ *of* $\hat{\phi}^{\mu}_{ols}$ *in regression* (3) *is given by*

$$t^{\mu}_{\hat{\phi}_{ols}} \xrightarrow{w} \frac{\int_0^1 B^{\mu}_{-d}(r)\, dB(r)}{\left(\int_0^1 \left(B^{\mu}_{-d}(r)\right)^2 dr\right)^{1/2}} \qquad \text{if } 0 \leq d < 0.5, \tag{20}$$

and

$$t^{\mu}_{\hat{\phi}_{ols}} \xrightarrow{w} N(0,1) \qquad \text{if } 0.5 \leq d < 1. \tag{21}$$

Theorem 3 *Under the null hypothesis that* y_t *is a random walk,* $\hat{\phi}^{\tau}_{ols}$*, the OLS coefficient associated to* ϕ *in* (4) *is a consistent estimator of* $\phi = 0$ *and converges to its true value at a rate* T^{1-d} *when* $0 \leq d < 0.5$, $(T\log T)^{0.5}$ *when* $d = 0.5$, *and at the standard rate* $T^{1/2}$ *when* $0.5 < d < 1$. *Its asymptotic distribution is given by*

$$T^{1-d}\hat{\phi}^{\tau}_{ols} \xrightarrow{w} \frac{\int_0^1 B^{\tau}_{-d}(r)\, dB(r)}{\int_0^1 \left(B^{\tau}_{-d}(r)\right)^2 dr} \qquad \text{if } 0 \leq d < 0.5, \tag{22}$$

$$(T\log T)^{1/2}\hat{\phi}^{\tau}_{ols} \xrightarrow{w} N(0,\pi) \qquad \text{if } d = 0.5, \tag{23}$$

and

$$T^{1/2}\hat{\phi}^{\tau}_{ols} \xrightarrow{w} N\left(0, \frac{\Gamma^2(d)}{\Gamma(2d-1)}\right) \qquad \text{if } 0.5 < d < 1. \tag{24}$$

where $B^{\tau}_{-d}(s)$ *is a detrended fractional brownian motion defined as* $B^{\tau}_{-d}(s) = B_{-d}(s) + (6s-4)\int_0^1 B_{-d}(r)\, dr - (12s-6)\int_0^1 rB_{-d}(r)\, dr$.

Theorem 4 *Under the null hypothesis that y_t is a random walk, the asymptotic distribution of the $t-statistic$ of $\hat{\phi}_{ols}^{\tau}$ in regression* (4) *is given by*

$$t_{\hat{\phi}_{ols}} \xrightarrow{w} \frac{\int_0^1 B_{-d}^{\tau}(r)\,dB(r)}{\left(\int_0^1 \left(B_{-d}^{\tau}(r)\right)^2 dr\right)^{1/2}} \qquad if\ 0 \leq d < 0.5, \tag{25}$$

and

$$t_{\hat{\phi}_{ols}} \xrightarrow{w} N(0,1) \qquad if\ 0.5 \leq d < 1.$$

To check how the previous asymptotic results perform in finite samples, Table 2 reports the empirical critical values of the t_ϕ for different significance levels and different values of d. The results are based on a Monte-Carlo study with a number of replications $N = 5,000$ of DGP 1 where $\sigma_\varepsilon = 1$. The critical values for $d \in [0, 0.5)$ are clearly different from those corresponding to a one-sided test using a standardized normal distribution (-1.28, -1.64 and -2.33, respectively, for the three significance levels reported below). By contrast, when $d \in [0.5, 1)$, the critical values resemble much more those of a $N(0,1)$ distribution, particularly for values of $d > 0.6$ and samples sizes T$\geq$ 400.

TABLE 2

Empirical Critical Values

DGP: $\Delta y_t = \varepsilon_t$; **Regression:** $\Delta y_t = \alpha + \phi\Delta\ ^d y_{t-1} + a_t$									
T	$T = 100$			$T = 400$			$T = 1000$		
d / sig.lev.	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.0	-2.58	-2.89	-3.51	-2.57	-2.89	-3.45	-2.57	-2.86	-3.43
0.2	-2.31	-2.61	-3.29	-2.31	-2.60	-3.29	-2.77	-3.09	-3.67
0.4	-1.99	-2.31	-2.94	-1.99	-2.30	-2.95	-2.27	-2.60	-3.28
0.5	-1.77	-2.13	-2.77	-1.69	-2.13	-2.77	-1.68	-2.06	-2.87
0.6	-1.49	-1.80	-2.59	-1.51	-1.73	-2.45	-1.40	-1.85	-2.55
0.8	-1.42	-1.73	-2.41	-1.48	-1.83	-2.39	-1.33	-1.72	-2.44
0.9	-1.37	-1.68	-2.38	-1.27	-1.69	-2.35	-1.25	-1.66	-2.32

Table 3, in turn, reports the empirical size (95% significance level) of t_ϕ for the case where the asymptotic distribution is normal, using a 95% critical value of -1.64. Using the standard formula ($var(p) = p(1-p)/N$ with $p = 0.05$ and $N = 5,000$) to produce a confidence interval of the rejection rate of $[0.044, 0.056]$, it can be observed that the $N(0,1)$ tends to slightly over-reject the null for $T = 100$ and $0.5 \leq d < 0.8$, yet it fares reasonably well for larger sample sizes.

TABLE 3.

Empirical Size (Nominal Size: 5%)

DGP: $\Delta\ y_t = \varepsilon_t$; **Regression:** $\Delta y_t = \alpha + \phi\Delta\ ^d y_{t-1} + a_t$			
d / T.	T=100	T=400	T=1000
0.5	10.42%	7.14%	5.48%
0.6	8.10%	5.80%	4.96%
0.8	5.56%	5.47%	5.12%
0.9	5.44%	5.20%	5.10%

As for power, Table 4 reports the rejection rates of the FDF test under the alternative hypothesis. The DGP in this case is a $FI(d)$ process with no deterministic components. The critical values used to compute power are those tabulated in Table 2 for $d \in (0, 0.5)$, when the asymptotic distribution is non-standard, and the nominal critical values given the standardized normal (i.e., $-$ 1.64) for $d \in [0.5, 1)$. The rejection rates are 100% in most cases, except when $d > 0.8$ where samples of size $T = 400$ are needed to achieve almost full rejection of the false null hypothesis of a unit root. Use of the empirical critical values for the range $d \in [0.5, 1)$ instead of the nominal ones, not reported for the sake of brevity, does not qualitatively change the main results, lowering the rejection rates from 100% to about 93% in the smaller sample sizes when $d = 0.5$ or $d = 0.8$.

TABLE 4

Power (95%)

DGP: $\Delta\ ^d y_t = \varepsilon_t$; **Regression:** $\Delta y_t = \alpha + \phi\Delta\ ^d y_{t-1} + a_t$			
d	$T = 100$	$T = 400$	$T = 1000$
0.0	100%	100%	100%
0.2	100%	100%	100%
0.4	100%	100%	100%
0.5	99.9%	100%	100%
0.6	99.6%	100%	100%
0.8	65.6%	99.7%	100%
0.9	27.0%	63.7%	94.7%

Case II: Constant term included in the regression. DGP is a random walk with drift.

We now study the case considered by West (1988) in the $I(1)$ *vs* $I(0)$ framework, namely that both DGP and model share the same deterministic component. For simplicity, as in West (1988), we restrict the analysis to the presence of a drift although similar results obtain when both DGP and model share a drift and a linear trend. Hence:

$$\textit{DGP 2}: \ y_t = \alpha + y_{t-1} + \varepsilon_t, \ t \geq 1 \tag{26}$$

$$\textit{Regression Model 1}: \ \Delta y_t = \alpha + \phi \Delta^d y_{t-1} + a_t \tag{27}$$

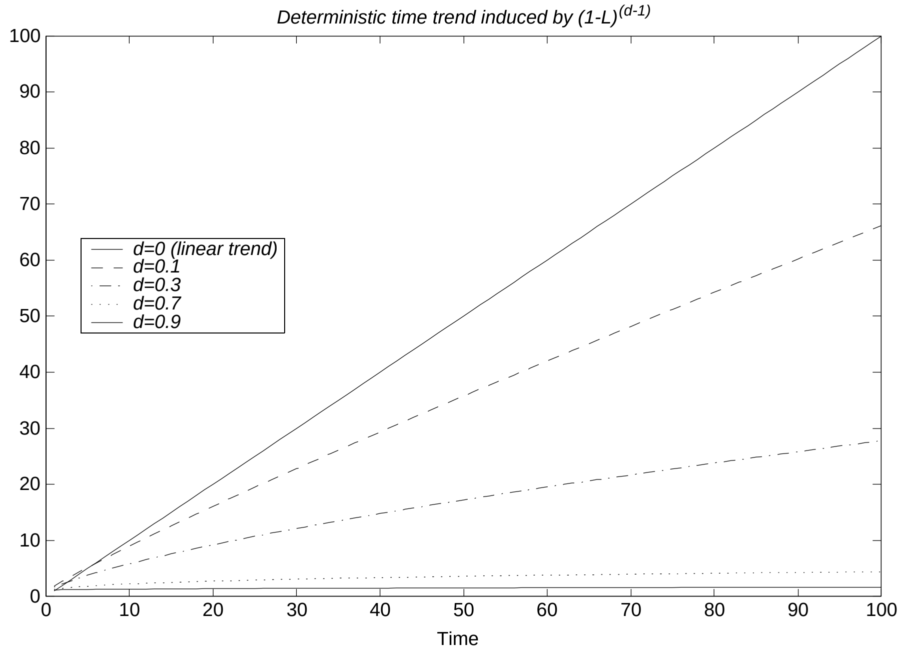
Before exploring the asymptotic properties of the test statistics associated to (27), it is useful to analyze the nature of the process $\Delta^d y_t$ under the null hypothesis. Under (26), it can be rewritten as the sum of a deterministic term and a purely stochastic $FI\,(1-d)$ process:

$$\Delta^d y_t = \Delta^{d-1}\alpha + \Delta^{d-1}\varepsilon_t. \tag{28}$$

The first component in the RHS of (28) is a deterministic time trend given by

$$\Delta^{d-1}\alpha = \alpha \sum_{i=0}^{t-1} \pi_i\,(d-1)$$

where the coefficients $\{\pi_i\,(d-1)\}$ come from the expansion of $(1-L)^{d-1}$ in powers of L. When $d=0$, (the DF case), $\pi_i\,(d-1) = 1$ for all i, which implies $\sum_{i=0}^{t-1} \pi_i\,(d-1) = t$. When $d=1$, $\pi_0\,(d-1) = 1$ and the remaining $\pi_i\,(d-1) = 0$ for all $i > 0$, which implies that $\sum_{i=0}^{t-1} \pi_i\,(d-1) = 1$ in this case. For all intermediate values of $d \in (0,1)$, $\sum_{i=0}^{t-1} \pi_i\,(d-1)$ represents and increasing time trend bounded by these two extreme cases. It turns out that it is easy to prove that the latter trend is of order $O(T^{1-d})$ since, by Stirling ´s approximation, we get that $\pi_i\,(d-1) = \Gamma(i+1-d)/[\Gamma(1-d)\Gamma(i+1)] \sim i^{-d}$. Hence, the sum from 1 to T of those terms will yield the previous order of magnitude. Note that $d=1$ implies $O(1)$ whereas $d=0$ implies $O(T)$ in accord with the previous discussion of the two extreme cases. As for the intermediate cases, Figure 1 plots a range of time trends generated with different values of $d \in [0,1)$.



Since for any value of $d \in [0,1)$, the term $\alpha\Delta^{\ d-1}$ induces a time trend, albeit a linear one only for $d=0$, the process $\Delta^{\ d} y_t$ is always non-stationary for any value of d in that range. This behavior contrasts with what happens in the case where y_t is a drifless random walk, where $\Delta^{\ d} y_t$ happens to be an (asymptotically) stationary process when $d \in [0.5, 1)$. This fact leads to the following asymptotic results.

Lemma 3 *Let* $\{\varepsilon_t\}$ *be a sequence of zero-mean i.i.d. random variables with variance* σ^2 *such that* $E\left|\varepsilon_t^4\right| < \infty$ *and let* y_t *be a random walk with drift defined as in* (26) *and consider the filtered process*

$$z_t = \Delta^d y_t = \Delta^{d-1}\alpha 1_{(t>0)} + \Delta^{d-1}\varepsilon_t 1_{(t>0)} \qquad d \in [0,1),$$

Then:

1. *if* $0 \leq d < 1$
$$T^{-(2-d)} \sum_{t=1}^{T} \Delta^d y_t \xrightarrow{p} \frac{\alpha}{\Gamma(3-\ d)},$$

2. *if* $0 \leq d < 1$
$$T^{-(3-2d)} \sum_{t=1}^{T} \left(\Delta^d y_t\right)^2 \xrightarrow{p} \frac{\alpha^2}{\left(\Gamma(2-d)\right)^2 (3-2d)}. \qquad (29)$$

Lemma 4 *Let ε_t and y_t be defined as in Lemma 3. Then the following processes are martingale differences and verify that if* $0 \leq d < 1$,

$$T^{-(3/2-\ d)} \sum_{t=2}^{T} \Delta^d y_{t-1} \varepsilon_t \xrightarrow{w} N\left(0, \frac{\sigma^2 \alpha^2}{\Gamma(2-d)^2 (3-2d)}\right). \tag{30}$$

The results in the next couple of theorems parallel those found by West (1988) in the DF case. As a consequence of the inclusion of a drift in the DGP, the asymptotic distributions of $\hat{\phi}_{ols}$ and t_ϕ are normal for all values of d.

Theorem 5 *Under the null hypothesis that y_t is a random walk with drift, $\hat{\phi}^{\mu}_{ols}$ is a consistent estimator of $\phi = 0$ and converges to its true value at a rate $T^{3/2-\ d}$ for $0 \leq d < 1$. Its asymptotic distribution is given by*

$$\begin{pmatrix} T^{1/2}\left(\hat{\alpha}_{ols} - \alpha\right) \\ T^{3/2-\ d} \hat{\phi}^{\mu}_{ols} \end{pmatrix} \xrightarrow{w} N\left(0, \sigma^2 Q_2^{-1}\right), \tag{31}$$

where

$$Q_2 = \begin{pmatrix} 1 & \frac{\alpha}{\Gamma(3-\ d)} \\ \frac{\alpha}{\Gamma(3-\ d)} & \frac{\alpha^2}{\Gamma(2-\ d)^2(3-2\ d)} \end{pmatrix}.$$

.

Theorem 6 *Under the null hypothesis that y_t is a random walk with drift, the asymptotic distribution of the $t-$statistic of $\hat{\phi}^{\mu}_{ols}$ in regression* (27) *is given by*

$$t_{\hat{\phi}_{ols}} \xrightarrow{w} N(0,1). \tag{32}$$

Table 5a and 5b report the empirical sizes for two alternative values of α, namely, $\alpha = 0.5$ and 5.0. Distinguishing between low a high values of the drift in the DGP relative to the variance of the error term ($\sigma_\varepsilon = 1$ in our case) turns out to be important since, as has been pointed out by Hylleberg and Mizon (1989), the orders of magnitude of the variability of the deterministic and the stochastic components of y_{t-1} in the DF framework are $O_p(\alpha^2 T^3)$ and $O_p(\sigma_\varepsilon^2 T^2)$, respectively, where the leading coefficients α^2 and σ_ε^2 have been included in the $O_p(.)$ terms for analytical convenience. In effect, note that if the squared drift, α^2, is very small relative to the variance, σ_ε^2, the leading term will be scaled by a very small number. This implies that, in finite samples, the stochastic component may dominate the behaviour of the distribution of t$_\phi$ in such a way that it will resemble that of the DF test when a constant term is included in the model and the DGP is a driftless random walk. In our setup, the orders of magnitude of the variability of the deterministic and stochastic components of $\Delta^d y_{t-1}$ are $O_p(\alpha^2 T^{3-2d})$ for the former, and $O_p(\sigma_\varepsilon^2 T^{2(1-d)}), O_p(\sigma_\varepsilon^2 T \log T)$ and $O_p(\sigma_\varepsilon^2 T)$ for the latter, depending on whether $d \in (0, 0.5)$, $d = 0.5$ or $d \in (0.5, 1)$ (see proof of Lemma 3 in the Appendix). As it can be observed below , a low value of α distorts somewhat the size, yet the distortions are not that large, particularly for $d \geq 0.5$. By contrast, when $\alpha = 5$, the empirical sizes match almost perfectly the nominal 5%. An intuitive explanation of why the size distortions in Tables 5.a and 5.b tend to decline with the value of d is that when $d \in [0.5, 1)$, the stochastic component also behaves closely to a normal distribution so that the size of α is less relevant. Notice also that a comparison of the orders of magnitude of the deterministic and stochastic components of $\Delta^d y_{t-1}$, for large T and $\sigma_\varepsilon^2 = 1$, implies that the former will dominate the latter when $\alpha > T^{-1/2}$ for $d \in (0, 0.5)$, $\alpha > (\log T)^{1/2} T^{-1}$ for $d = 0.5$, and when $\alpha > T^{-(1-d)}$ for $d \in (0.5, 1)$. These results extend those presented in Hylleberg and Mizon (1989) for the case $d = 0$, for which they found that the deterministic term will dominate when $\alpha > T^{-1/2}$, implying that small values of α will generate stronger size distortions in finite samples.

TABLE 5a.

Empirical Size 5%; α=0.5

DGP: $\Delta y_t = \alpha + \varepsilon_t$; **Regression:** $\Delta y_t = \alpha + \phi \Delta^{d} y_{t-1} + a_t$			
d / T	T=100	T=400	T=1000
0.0	0.011	0.071	0.063
0.2	0.144	0.093	0.066
0.4	0.112	0.091	0.075
0.5	0.109	0.087	0.080
0.6	0.091	0.082	0.083
0.8	0.083	0.070	0.066
0.9	0.071	0.061	0.058

TABLE 5b.

EMPIRICAL SIZE 5%;α=5

DGP: $\Delta y_t = \alpha + \varepsilon_t$; **Regression:** $\Delta y_t = \alpha + \phi \Delta^{\ d} y_{t-1} + a_t$			
d / T	T=100	T=400	T=1000
0.0	0.059	0.053	0.051
0.2	0.054	0.056	0.048
0.4	0.053	0.055	0.052
0.5	0.052	0.053	0.050
0.6	0.052	0.053	0.049
0.8	0.053	0.052	0.049
0.9	0.052	0.051	0.050

Finally, it should be pointed out that the results by West have been extended by Lubian (1999, Theorem 3.3) to the case where the standard DF test with a constant term in the model is applied to a $FI(1+d)$ process with $-\frac{1}{2} < d < \frac{1}{2}$, finding that $\mathrm{T}^{\frac{3}{2}-d}(\widehat{\rho}_{ols} - 1)$ tends to a normal distribution. It can be easily checked that, when $d = 0$, West´s results are recovered both in Lubians´s analysis and ours.

Case III: Constant term and time trend included in the regression. DGP is random walk with or without drift.

We now turn to examine the case where a linear trend is included in the model. As mentioned above, the presence of a linear trend in the regression allows one to achieve an invariant test in the DF framework. However, as shown below, this will not be the case once $FI(d)$ processes are allowed for. Our setup is as follows:

$$\textit{DGP 2}\text{: } y_t = \alpha + y_{t-1} + \varepsilon_t \tag{33}$$

$$\textit{Regression Model 2}: \Delta y_t = \alpha + \delta t + \phi \Delta^{\ d} y_{t-1} + a_t \tag{34}$$

To explain the different behaviour of the FDF test vis-á-vis the DF test in this case, it is convenient to recall why the introduction of a trend in the DF regression achieves invariance of the DF test. If the process is a random walk (with or without a drift) and the DF auxiliary regression contains a drift and a linear trend, the regressors t and y_{t-1} turn out to be colinear in large samples isolating in this manner the purely stochastic component of y_{t-1}.

Hence, the distributions are nonstandard, like in Case I above (replacing demeaned BM by detrended BM) when no deterministic components were included in the DGP. Nevertheless, this is not the case for $\Delta\, ^d y_{t-1}$ for any value of $d \neq 0$, since t and $\Delta\, ^d y_{t-1}$ are no longer colinear under the null hypothesis. This is so since the order of magnitude of the variability of the deterministic component of $\Delta\, ^d y_{t-1}$ is $O_p(\alpha^2 T^{3-2d})$ while that of the linear trend is $O(\sigma_\varepsilon^2 T^3)$. Hence, like in Case II, given that the stochastic component of $\Delta^d y_{t-1}$ is dominated by the smooth time trend represented by $\Delta^{d-1}\alpha$, asymptotic normality holds in the limit for all values of $d \in (0,1)$. Also note, that in agreement with the discussion of Case II and in contrast to what happens in the DF test, the size of the drift under the null again matters since the leading term is scaled by α. This dependence implies that, in finite samples, the test may suffer from lack of invariance with respect to the size of the drift relative to the variance of the error term, as was the case in the previous section. A possible solution to recover invariance would be to replace the linear trend in Regression Model 2 by the truncated smooth function of time $\Delta^{d-1} = \sum_{i=0}^{t-1} \pi_i\,(d-1)$. However, as in the DF framework, a FDF test constructed in this way would have a nonstandard limiting distribution for values of $d \in [0, 0.5)$ and standard for $d \in [0.5, 1)$.

Making use of the following lemma, asymptotic results can be derived for this case.

Lemma 5 *Let $\{\varepsilon_t\}$ be a sequence of zero-mean i.i.d. random variables with variance σ^2 such that $E\left|\varepsilon_t^4\right| < \infty$ and y_t be the random walk with drift defined as in* (33). *Then, if* $0 < d < 1$

$$\sum_{t=2}^{T} t\Delta^d y_{t-1} = O_p\left(T^{3-d}\right). \tag{35}$$

The next two theorems derive the relevant asymptotic distributions, stressing the discontinuity that there exists between the DF case ($d = 0$) and the other values of d under the alternative ($0 < d < 1$).

Theorem 7 *Under the null hypothesis that y_t is a random walk with drift as defined in* (33), *$\hat{\phi}_{ols}$ computed in regression* (34) *is a consistent estimator of $\phi = 0$ and converges to its true value at a rate T when $d = 0$, and $T^{3/2-d}$ when $0 < d < 1$. Its asymptotic distribution is given by*

1. *If $d = 0$, the DF distribution of $\hat{\phi}_{ols}$ for Model 2.*
2. *If $0 < d < 1$*

$$\begin{pmatrix} T^{1/2}\left(\hat{\alpha}_{ols} - \alpha\right) \\ T^{3/2}\hat{\delta}_{ols} \\ T^{3/2-\ d}\hat{\phi}_{ols} \end{pmatrix} \xrightarrow{w} N\left(0, \sigma^2 Q_3^{-1}\right), \tag{36}$$

where,

$$Q_3 = \begin{pmatrix} 1 & 1/2 & \frac{\alpha}{\Gamma(3-d)} \\ 1/2 & 1/3 & \frac{\alpha}{\Gamma(2-d)(3-d)} \\ \frac{\alpha}{\Gamma(3-d)} & \frac{\alpha}{\Gamma(2-d)(3-d)} & \frac{\alpha^2}{\Gamma(2-d)^2(3-2d)} \end{pmatrix}.$$

Theorem 8 *Under the null hypothesis that y_t is a random walk with drift as defined in* (33)*, the asymptotic distribution of the $t-$statistic of $\hat{\phi}^{\tau}_{ols}$ computed in regression* (34) *is given by*

1. *If $d = 0$, the Dickey-Fuller distribution of t_ϕ in Model 2.*
2. *If $0 < d < 1$*

$$t_{\hat{\phi}_{ols}} \xrightarrow{w} N(0,1). \tag{37}$$

As in Case I, Tables 6a and 6b reports the empirical size of the FDF test in this setup, using a nominal critical value of -1.64 . Since the test is not invariant to the value of α, sizes are reported again for the two values of α considered above. For $\alpha = 5.0$, the empirical sizes almost mimic the nominal size while for $\alpha = 0.5$, size distortions are very serious for $d \leq 0.5$. The intuitive reason is similar to the one conveyed above in the discussion of Case II.

TABLE 6a.

Empirical Size 95%; α=0.5

DGP: $\Delta y_t = \alpha + \varepsilon_t$; **Regression:** $\Delta y_t = \alpha + \delta t + \phi\Delta^d y_{t-1} + a_t$			
d / T.	T=100	T=400	T=1000
0.2	0.431	0.371	0.309
0.4	0.311	0.265	0.222
0.5	0.243	0.190	0.168
0.6	0.093	0.083	0.075
0.8	0.089	0.085	0.072
0.9	0.075	0.068	0.056

TABLE 6b.

EMPIRICAL SIZE 95%; α=5

DGP: $\Delta y_t = \alpha + \varepsilon_t$; **Regression:** $\Delta y_t = \alpha + \delta t + \phi \Delta^{\,d} y_{t-1} + a_t$			
d / T	T=100	T=400	T=1000
0.2	0.054	0.056	0.048
0.4	0.058	0.055	0.054
0.5	0.052	0.059	0.054
0.6	0.066	0.056	0.049
0.8	0.064	0.053	0.049
0.9	0.061	0.051	0.051

Deterministic Components in the FDF test with estimated d.

So far we have considered the case where the value of d needed to implement the FDF test is assumed to be *a priori* known. However, in many instances this turns out to be a too restrictive assumption, for instance, when a composite alternative is considered. Thus, the more realistic case, where the value of d needs to be estimated, is addressed in this section. As in the simple FDF framework analyzed in DGM, it is found that provided a $T^{1/2}-$consistent estimator of d is used, the asymptotic distribution of t_ϕ is $N(0,1)$ in all of the three cases analyzed above. This result is very convenient since it allows to use standard critical values in all possible situations.

More formally, it is assumed that y_t is a random walk without or with drift and the FDF auxiliary regression is run including a constant or a constant and a time trend, respectively. Since d is unknown, a $T^{1/2}$-consistent estimator of $d \leq 1$, $\hat{d}_T$, is used. Since the value of d used to implement the test should be strictly smaller than 1, the following trimming rule is applied in the choice of the estimator:

$$\hat{d} = \begin{cases} \hat{d}_T, \text{ if } \hat{d}_T < 1 - c \\ 1 - c, \text{ if } \hat{d}_T \geq 1 - c \end{cases} \tag{38}$$

where $c > 0$ is a (fixed) value in a neighborhood of zero, such that $(1 - c)$ is sufficiently close to unity. The following theorem sums up the results obtained in this case.

Theorem 9 *Under the null hypothesis that y_t is a random walk without or with a drift, the t-statistic associated to the OLS estimate of ϕ in either of the following regressions*

$$\Delta y_t = \alpha + \phi \Delta^{\hat{d}} y_{t-1} + a_t, \tag{39}$$

or

$$\Delta y_t = \alpha + \delta t + \phi \Delta^{\hat{d}} y_{t-1} + a_t, \tag{40}$$

where d has been chosen according to the criterion function defined in (38)*, is asymptotically distributed as:*

$$t^{\mu}_{\hat{\phi}_{ols}}\left(\hat{d}\right) \xrightarrow{w} N\left(0,1\right),$$

and

$$t^{\tau}_{\hat{\phi}_{ols}}\left(\hat{d}\right) \xrightarrow{w} N\left(0,1\right),$$

where $t^{\mu}_{\hat{\phi}_{ols}}\left(\hat{d}\right)$ *and* $t^{\tau}_{\hat{\phi}_{ols}}\left(\hat{d}\right)$ *are computed from regressions* (39) *and* (40,) *respectively.*

As discussed in DMG, among the different estimation procedures available in the time domain which yield which yield $T^{1/2}-$consistent estimates of d in the permissible range, the ML estimators derived by Beran (1995) and Tanaka (1999) or the Minimum Distance estimators derived by Galbraith and Zinde-Walsh (2001) and Mayoral (2002) can be used.

A simple strategy to test for the value of d in the presence of deterministic components

In view of the above results, we propose a natural strategy to test for a value of d in the presence of deterministic components. We will focus on the most realistic situation where d is estimated. Before commenting on this testing strategy, it is important to stress that an interesting consequence from our analysis is that, in contrast to the use of the DF test for $H_0 : d = 1$ when deterministic components are present, there is no need to use new critical values relative to the case where no deterministic components are considered. Furthermore, in our framework, all the critical values come from standard distributions. These two features transform the problem of determining the right deterministic components into the standard issue of variable selection. Another important advantage of our strategy, as will be discussed in the next section, is that we do not need to pre-filter the data by filters like $(1-L)^{1/2}$ to apply an estimation method only valid for $|d| < 1/2$, or to remove a linear trend by means of the filter $(1-L)$.

Our proposed testing strategy for $H_0 : d = 1$ vs. $H_1 : 0 < d < 1$ will take as starting point the general *Regression Model 2* in (34) and is based on the t-ratio of $\widehat{\phi}$ along the following steps. First, if the null is rejected, then the process is not $I(1)$ and the testing strategy will stop. If the null is not rejected, then we can use critical values from the $N(0,1)$ distribution to test whether the coefficient δ on t is significant. If it is significant, we stop. Otherwise, we estimate *Regression Model 1* in (25) and follow again the same strategy. In sum, our proposed strategy is easy to apply and turns out to be much simpler than those used in applied work, as the next section illustrates.

EMPIRICAL ILLUSTRATION

An interesting application of the theoretical results applied above is to examine whether the time-series of GDP per capita of OECD countries behave as $FI(d)$ processes with $0.5 < d < 1$. These are series which are clearly trending upwards and therefore provide nice examples of the role of deterministic terms in the use of the FDF test. As pointed out in an interesting paper by Michelacci and Zaffaroni (2000) such a long- memory behaviour could well explain the seemingly contradictory results obtained in the literature on growth and convergence that a unit root cannot be rejected in (the log of) those series and yet a 2% rate convergence rate to a steady-state level, approximated by a linear trend, is typically found in most empirical exercises of this kind (see Barro and Sala i Martín, 1995 and Jones, 1995). The explanation offered by these authors to the previous puzzle relies upon two well-known results in the literature on long-memory processes, namely, that standard unit root tests have low power against values of d in the nonstationary range ($0.5 < d < 1$), and that for all values of $d \in [0, 1)$ there is "mean reversion", in the sense that shocks do not have permanent effects . Using Maddison´s (1995) data set of annual GDP per capita series for 16 OECD countries during the period 1870 - 1994 (125 observations) and a log-periodogram estimator of d due to Robinson (1995), they find that in most countries the order of fractional integration is within the prespecified range, validating in this way their explanation of the puzzle. Since that estimation procedure is restricted to the range of $FI(d)$ processes with finite variance $|d| < 1/2$, the authors proceed by first detrending the data and then applying the truncated filter $(1 - L)^{1/2}$ to the residuals, discarding the first 10 observations.

The previous results have been recently criticized by Silverberg and Verspagen (2001) on the grounds of the use of the $(1-L)^{1/2}$ filter and of Robinson´s semi-parametric estimation procedure, which suffers from serious small-sample bias. Instead, they propose the use of the first-difference filter, $(1-L)$, to remove the trend and of Sowell´s (1992) parametric ML estimator of ARFIMA models to tackle short-memory contamination in the estimation of d. Using those alternative procedure they find, in stark contrast to Michelacci and Zaffaroni ´s results, that d tends to be either not significantly different from unity or significantly above unity for most countries in an extended sample of 25 countries.

To shed light on this controversy we apply the FDF test developed for Case III to the logged GDP p.c. of a subset of 10 of the main OECD countries which are listed in Table

7, where the estimated intercept and stanrd deviation of the residuals in the regression $\Delta y_t = \alpha + u_t$ is reported. As can be observed the ratio of the estimated mean to the standard deviation is small, around 0.4, implying that the use of a $N(0,1)$ distribution may yield misleading results particularly for values of d in the stationary range, as shown in Table 6a for T=100. However, when the ADF and the Phillips-Perron (P-P) unit root tests were computed using a constant and a time trend in the regression model, the $I(1)$ null hypothesis could not be rejected. By contrast, the KPPS test, which takes $I(0)$ as the null, yielded overall rejection confirming the high persistence of the series. Thus there are clear signs that the first-difference series have a drift and that it is likely that they are nonstationary.

TABLE 7

ESTIMATES OF α AND σ_u

Country	Mean	St. D. Res.
Australia	0.012	0.044
Canada	0.0195	0.055
Denmark	0.018	0.088
France	0.018	0.069
Germany	0.018	0.082
Italy	0.019	0.07
Netherlands	0.015	0.075
UK	0.013	0.0327
USA	0.017	0.056
Spain	0.019	0.045

Since there were clear signs of autocorrelation in u_t, an AFDF test with intercept and linear trend was applied to the series. The number of lags of the dependent variable was chosen according to the AIC criterion with a maximum lag of length $k = 4$, since $T = 125$ (95 for Spain) and $T^{1/3} = 5$. Pre-estimation of d using Sowell´s (1992) ML parametric approach for various ARFIMA (p,d,q) specifications of the first-differenced data, with p and q up to four lags, allows one to select a value of d for each country on the basis of the AIC criterion. The reported values of d, presented in Table 8, add unity to the obtained estimates. The preferred models were also estimated using Mayoral ´s (2001) MD estimation approach, with the series in levels, yielding the pre-estimates of d presented in Table 9. Both sets of estimates tend to provide very similar results. In general, the estimated values of d tend to be below unity and above one-half. Thus, in view that the size distortions of the FDF test in Case III are not too important for $d \in (0.5, 1)$ and that the ADF and P-P tests have reasonable power against stationary $FI(d)$ processes, i.e., $d \in [0, 0.5)$,we tested for the null of $d = 1$ against the sequence of alternative hypotheses $d = 0.6, 0.7, 0.8$ and 0.9, each at a time, by using the 95%-critical value of a standardized normal, i.e. -1.64. Given the strong rejection of the null in most cases, our results seem to be more favorable to Michelacci and Zaffaroni ´findings that to Silverberg and Verspagen´s

TABLE 8

ESTIMATES OF d (ML estimation in first differences)

Country	$(0, d, 0)$	$(1, d, 0)$.	$(1, d, 1)$	$(0, d, 1)$
Australia	♠1.03^{*} -417	-	1.05 -415	1.01 -415
Canada	1.13 -366	♠ 0.44^{*} -373	0.43^{*} -371	1.04 -367
Denmark	0.58^{*} -272	♠0.72^{*} -273.7	0.75 -271	0.78 -273.5
France	1.08 -306	0.76 -308	0.94 -310	♠ 0.82 -311
Germany	1.15 -268	0.65 -278	0.65 -282	♠ 0.80^{*} -283
Italy	0.99 -294	0.85 -294	0.87 -296	♠0.81^{*} -2.97
Netherlands	0.89 -285	0.72 -2.87	-0.89 -288	♠0.77^{*} -289
UK	1.16 -497	♠ 0.71 -505	0.96 -500	0.71 -503
USA	1.13 -361	♠0.73^{*} -374.4	0.60^{*} -374.2	0.80 -367
Spain	♠ 0.98 314	0.94 316	0.71 317	0.95 316

Note: (♠)denotes the preferred ARFIMA model according to the AIC criterion whereas $(*)$ denotes rejection of the null hypothesis of a unit root

TABLE 9

ESTIMATES OF d (MD for the preferred model)

Country	d	model
Australia	1.003	$(0, d, 0)$
Canada	0.50	$(1, d, 0)$
Denmark	0.71	$(1, d, 0)$
France	0.77	$(0, d, 1)$
Germany	0.81	$(0, d, 1)$
Italy	0.82	$(0, d, 1)$
Netherlands	0.77	$(0, d, 1)$
UK	0.60	$(1, d, 0)$
USA	0.78	$(0, d, 0)$
Spain	0.83	$(1, d, 0)$

TABLE 10

AFDF TEST AGAINST FI(d)

Country | d	0.9	0.8	0.7	0.6
Australia	-0.18	-1.97*	-2.04*	-2.09*
Canada	-2.68*	-2.76*	-2.85*	-2.95*
Denmark	-2.99*	-3.09*	-3.19*	-5.81*
France	-2.77*	-2.80*	-2.83*	-2.84*
Germany	-2.83*	-2.93*	-3.01*	-3.11*
Italy	-2.54*	-2.53*	-2.53*	-2.53*
Netherlands	-1.95*	-2.04*	-2.13*	-2.33*
UK	-3.54*	-3.53*	-3.50*	-3.44*
USA	-3.30*	-3.43*	-3.56*	-3.70*
Spain	-0.44	-0.54	-0.66	-0.79

Note: (*) denotes 5%- rejection of the null hypothesis of a unit root.

CONCLUSIONS

The basic result of this paper is that allowing for deterministic terms (a constant /a linear time trend) in the DGP and in the auxiliary regression used to implement the FDF test for $I(1)$ *vs.* $FI(d)$, $d \in [0, 1)$ processes (with d fixed under the alternative), does not change the main asymptotic results derived for the DF test in testing for $I(1)$ *vs.* $I(0)$ series, with two exceptions. *First*, if $d \in [0.5, 1)$ asymptotic normality holds. *Second*, when a linear trend is included in the model to capture a non-zero drift under the null, the DF has a nonstandard distribution while the FDF is asymptotically normal for all $d \in (0, 1)$. However, in finite samples, the use of standard critical values and confidence intervals has to be taken some doses of caution when $d < 0.5$ and when the drift is small relative to the variance of the first-differenced series. Nonetheless, if d is pre-estimated using a $T^{1/2}-$consistent estimator, all possible forms of the FDF test turn out to have standard asymptotic distributions. Since estimation of d turns out to be the most realistic case in applied work, our results provide a simple testing strategy to test for $d = 1$ against $0 < d < 1$ based on starting from *Regression Model 2* and testing for the significance of the coefficients on the deterministic components whenever the null hypothesis is not rejected. This testing strategy turns out to be much simpler than those typically used in applied work and only entails the use of asymptotically normally-distributed test statistics .

Useful extensions of the present paper´s setup that are under current investigation by the authors include allowing for structural breaks and testing for cointegration between two FI(d) series which have a non-zero drift and where a constant term or a linear trend is included in the regression model.

REFERENCES

Barro, R. J. and X. Sala i Martín (1995) *Economic Growth* .McGraw-Hill, New York.

Beran, J (1995): "Maximum likelihood estimation of the differencing parameter for invertible and short and long memory autoregressive integrated moving average models", *Journal of the Royal Statistical Society*, 57, 659-672.

Dickey, D. A. and W.A. Fuller (1981): "Likelihood ratio tests for autoregressive time series with a unit root," *Econometrica,* 49, 1057-1072.

Dolado, J., Gonzalo, J. and L. Mayoral (2002)," A fractional Dickey-Fuller test for unit roots", *Econometrica*, 2002, forthcoming.

Dolado, J. and F. Marmol (2001): " Some asymptotic inference results for multivariate long-memory processes". Mimeo, under revision in *The Econometrics Journal.*

Galbraith, J.W. and V. Zinde-Walsh (1997): "Time domain methods for the estimation of fractionally-integrated time series models," Mimeo.

Hall, P. and C.C. Heyde (1980): *Martingale Limit Theory and its Applications*, New York: Academic Press.

Helland, I. S. (1982): "Central limit theorems for martingales with discrete or continuous time", *Scandinavian Journal of Statistics*, 9, *Journal*, 94.

Hylleberg, S. and G. Mizon (1989):" A note on the distribution of the Least Squares estimator of a random walk with drift", *Economics Letters*, 29, 225-230.

Jones, C.(1995) "Time Series Tests of Endogenous Growth Models" *Quarterly Journal of Economics*, 110, 495-525.

Lubian, D. (1999) "Long-memory errors in time series regressions with a unit root" *Journal of Time Series,* 20, 565-577.

Maddison, A. (1995): Monitoring the World Economy, 1820-1992, Paris: OECD.

Marinucci, D., and P. Robinson(1999): "Alternative forms of Brownian motion," *Journal of Statistical Planning and Inference*, 80, 11-122.

Mayoral L., (2002), A new minimum distance estimation procedure of ARFIMA processes. Mimeo.

Michelacci, C. and P. Zaffaroni, P.(2000), "Fractional Beta Convergence", *Journal of Monetary Economics*, 45, 129-153.

Robinson, P.M (1994): "Efficient tests of nonstationary hypotheses," *Journal of the American Statistical Association, 89,* 1420-1437.

Robinson, P. (1995), "Log-periodogram of Time Series with Long-Range Dependence", *Annals of Statistics*, 23, 1048-1072.

Said, S. Dickey, D (1984): "Testing for unit roots in autoregressive moving average models of unknown order," *Biometrika*, 71, 599-608.

Silverberg, G. and B. Verspagen (2001), " A Note on Michelacci and Zaffaroni, Long Memory and Time Series of Economic Growth", University of Maastricht, mimeo

Sowell , F.B.(1992) : "Maximum likelihood estimation of stationary univariate fractionally-integrated time-series models," *Journal of Econometrics, 53*, 165-188.

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

Theil, H., (1971), Principles of Econometrics (Wiley, New York).

West, K.D. (1988) "Asymptotic normality when regressors have a unit root", *Econometrica*, 56,1397-1417.

APPENDIX

Proof of Theorem 1

The OLS estimates of $(\alpha, \phi)'$ are given by

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\phi}_{ols} \end{pmatrix} = \begin{pmatrix} T & \sum_{t=2}^{T} \Delta^{d} y_{t-1} \\ \sum_{t=2}^{T} \Delta^{d} y_{t-1} & \sum_{t=2}^{T} \left(\Delta^{d} y_{t-1}\right)^{2} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=2}^{T} \Delta y_{t} \\ \sum_{t=2}^{T} \Delta y_{t} \Delta^{d} y_{t-1} \end{pmatrix}. \tag{41}$$

We distinguish three cases according to the value of d.

I. First case: $0 \leq d < 0.5$. In this case, $\Delta^{\ d} y_{t-1}$ is a nonstationary $FI\,(1-\ d)$ and the following convergences hold:

$$T^{3/2-\ d} \sum_{t=2}^{T} \Delta^{\ d} y_{t-1} \xrightarrow{w} \sigma \int_{0}^{1} B_{-d}\,(r)\, dr,$$

and

$$T^{2(1-\ d)} \sum_{t=2}^{T} \left(\Delta^{d} y_{t-1}\right)^{2} \xrightarrow{w} \sigma^{2} \int_{0}^{1} B_{-\ d}^{2}\,(r)\, dr.$$

This implies that:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\phi}_{ols} \end{pmatrix} = \begin{pmatrix} O\,(T) & O_{p}\left(T^{3/2-\ d}\right) \\ O_{p}\left(T^{3/2-\ d}\right) & O_{p}\left(T^{2(1-\ d)}\right) \end{pmatrix}^{-1} \begin{pmatrix} O_{p}\left(T^{1/2}\right) \\ O_{p}\left(T^{1-\ d}\right) \end{pmatrix}. \tag{42}$$

Define the scaling matrix:

$$\Upsilon_{T} = \begin{pmatrix} T^{1/2} & 0 \\ 0 & T^{1-\ d} \end{pmatrix}. \tag{43}$$

Then:

$$\Upsilon_{T} \begin{pmatrix} \hat{\alpha} \\ \hat{\phi}_{ols} \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & T^{-3/2+\ d} \sum_{t=1}^{T} \Delta^{d} y_{t-1} \\ T^{-3/2+\ d} \sum_{t=1}^{T} \Delta^{d} y_{t-1} & T^{2(1-\ d)} \sum_{t=1}^{T} \left(\Delta^{d} y_{t-1}\right)^{2} \end{pmatrix}^{-1} \begin{pmatrix} T^{-1/2} \sum_{t=1}^{T} \Delta y_{t} \\ T^{-(1-\ d)} \sum_{t=1}^{T} \Delta y_{t} \Delta^{d} y_{t-1} \end{pmatrix}. \tag{44}$$

Which in turn implies that

$$\Upsilon_{T} \begin{pmatrix} \hat{\alpha} \\ \hat{\phi}_{ols} \end{pmatrix} \xrightarrow{w} \begin{pmatrix} 1 & \sigma \int_{0}^{1} B_{-\ d}\,(r)\, dr \\ \sigma \int_{0}^{1} B_{-\ d}\,(r)\, dr & \sigma^{2} \int_{0}^{1} B_{-\ d}^{2}\,(r)\, dr \end{pmatrix}^{-1} \begin{pmatrix} \sigma B(1) \\ \sigma^{2} \int_{0}^{1} B_{-\ d}\,(r)\, dB(r) \end{pmatrix}.$$

Taking into account:

$$\begin{pmatrix} 1 & \sigma \int_{0}^{1} B_{-\ d}\,(r)\, dr \\ \sigma \int_{0}^{1} B_{-\ d}\,(r)\, dr & \sigma^{2} \int_{0}^{1} B_{-\ d}^{2}\,(r)\, dr \end{pmatrix}^{-1} =$$

$$= \frac{1}{\sigma^2 \int_0^1 B_{-\ d}^2 (r)\, dr - \left(\sigma \int_0^1 B_{-\ d} (r)\, dr\right)^2} \begin{pmatrix} \sigma^2 \int_0^1 B_{-\ d}^2 (r)\, dr & -\sigma \int_0^1 B_{-\ d} (r)\, dr \\ -\sigma \int_0^1 B_{-\ d} (r)\, dr & 1 \end{pmatrix},$$

then,

$$T^{1-\ d} \hat{\phi}_{ols} \xrightarrow{w} \frac{\int_0^1 B_{-\ d} (r)\, dB(r) - B(1) \int_0^1 B_{-\ d} (r)\, dr}{\int_0^1 B_{-\ d}^2 (r)\, dr - \left(\int_0^1 B_{-\ d} (r)\, dr\right)^2}. \tag{45}$$

Defining $B_{-d}^{\mu} (s) = B_{-d} (s) - \int_0^1 B_{-d} (r)\, dr$, it is straight forward to check that,

$$T^{1-\ d} \hat{\phi}_{ols} \xrightarrow{w} \frac{\int_0^1 B_{-d}^{\mu} (r)\, dB(r)}{\int_0^1 \left(B_{-\ d}^{\mu} (r)\right)^2 dr}.$$

II. Second case: $d = 0.5$

Define the weighting matrix:

$$\Upsilon_T = \begin{pmatrix} T^{1/2} & 0 \\ 0 & (T \log T)^{1/2} \end{pmatrix}. \tag{46}$$

Then,

$$\begin{aligned} \Upsilon_T \begin{pmatrix} \hat{\alpha} \\ \hat{\phi}_{ols} \end{pmatrix} &= \begin{pmatrix} 1 & T^{-1} (\log T)^{-1/2} \sum_{t=1}^T \Delta^d y_{t-1} \\ T^{-1} (\log T)^{-1/2} \sum_{t=1}^T \Delta^d y_{t-1} & (T \log T)^{-1} \sum_{t=1}^T \left(\Delta^d y_{t-1}\right)^2 \end{pmatrix}^{-1} \times \\ & \begin{pmatrix} T^{1/2} \sum_{t=1}^T \Delta y_t \\ (T \log T)^{1/2} \sum_{t=1}^T \Delta y_t \Delta^d y_{t-1} \end{pmatrix}. \end{aligned} \tag{47}$$

Since

$$\begin{pmatrix} 1 & T^{-1} (\log T)^{-1/2} \sum_{t=1}^T \Delta^d y_{t-1} \\ T^{-1} (\log T)^{-1/2} \sum_{t=1}^T \Delta^d y_{t-1} & (T \log T)^{-1} \sum_{t=1}^T \left(\Delta^d y_{t-1}\right)^2 \end{pmatrix} \xrightarrow{p} \begin{pmatrix} 1 & 0 \\ 0 & \frac{\sigma^2}{\pi} \end{pmatrix},$$

and

$$\begin{pmatrix} T^{-1/2} \sum_{t=1}^T \varepsilon_t \\ (T \log T)^{-1/2} \sum_{t=1}^T \varepsilon_t \Delta^d y_{t-1} \end{pmatrix} \xrightarrow{w} N \left(0, \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & \frac{\sigma^2}{\pi} \end{pmatrix}\right),$$

then

$$\Upsilon_T \begin{pmatrix} \hat{\alpha} \\ \hat{\phi}_{ols} \end{pmatrix} \xrightarrow{w} N \left(0, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \pi \end{pmatrix}\right).$$

III. Third case: $0.5 < d < 1$.

Since in this case the process $\Delta^d y_{t-1}$ is stationary, then:

$$T^{1/2} \begin{pmatrix} \hat{\alpha} \\ \hat{\phi}_{ols} \end{pmatrix} \xrightarrow{w} \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2 \sum_{i=1}^{\infty} \pi_i^2 (1-d) \end{pmatrix}^{-1} \begin{pmatrix} \sigma B(1) \\ \sigma^2 (\sum_{i=1}^{\infty} \pi_i^2 (1-d))^{1/2} B(1) \end{pmatrix}.$$

Therefore,

$$T^{1/2}\begin{pmatrix}\hat{\alpha}\\ \hat{\phi}_{ols}\end{pmatrix}\xrightarrow{w}N_2\left(0,\begin{pmatrix}\sigma & 0\\ 0 & \frac{\Gamma^2(d)}{\Gamma(2d-1)}\end{pmatrix}\right).\blacksquare \tag{48}$$

Proof of Theorem 2

I. First case: $0<d<0.5$

The OLS $t-test$ of the null hypothesis that $\phi=0$ is

$$t=\frac{\hat{\phi}_{ols}}{\hat{\sigma}_{\hat{\phi}_{ols}}},$$

where

$$\hat{\sigma}^2_{\hat{\phi}_{ols}}=s_T^2\begin{pmatrix}0 & 1\end{pmatrix}\begin{pmatrix}T & \sum_{t=1}^T\Delta^d y_{t-1}\\ \sum_{t=1}^T\Delta^d y_{t-1} & \sum_{t=1}^T\left(\Delta^d y_{t-1}\right)^2\end{pmatrix}^{-1}\begin{pmatrix}0\\ 1\end{pmatrix}, \tag{49}$$

and

$$s_T^2=(T-2)^{-1}\sum_{t=1}^T\left(\Delta y_t-\hat{\alpha}_{ols}-\hat{\phi}_{ols}\Delta^d y_{t-1}\right).$$

Multiplying (49) by $T^{2(1-\ d)}$ and following Hamilton (1994) p. 493 we get

$$T^{2(1-\ d)}\hat{\sigma}^2_{\hat{\phi}_{ols}}=s_T^2\begin{pmatrix}0 & 1\end{pmatrix}\Upsilon_T\begin{pmatrix}T & \sum_{t=1}^T\Delta^d y_{t-1}\\ \sum_{t=1}^T\Delta^d y_{t-1} & \sum_{t=1}^T\left(\Delta^d y_{t-1}\right)^2\end{pmatrix}^{-1}\Upsilon_T\begin{pmatrix}0\\ 1\end{pmatrix}, \tag{50}$$

where the matrix Υ_T is defined as in (43). Since

$$\Upsilon_T\begin{pmatrix}T & \sum_{t=1}^T\Delta^d y_{t-1}\\ \sum_{t=1}^T\Delta^d y_{t-1} & \sum_{t=1}^T\left(\Delta^d y_{t-1}\right)^2\end{pmatrix}^{-1}\Upsilon_T$$

$$\xrightarrow{w}\begin{pmatrix}1 & 0\\ 0 & \sigma\end{pmatrix}^{-1}\begin{pmatrix}1 & \int B_{-\ d}(r)\,dr\\ \int B_{-\ d}(r)\,dr & \int\left(B_{-\ d}(r)\right)^2dr\end{pmatrix}^{-1}\begin{pmatrix}1 & 0\\ 0 & \sigma\end{pmatrix}^{-1},$$

then

$$T^{2(1-\ d)}\hat{\sigma}^2_{\hat{\phi}_{ols}}\xrightarrow{w}s_T^2\begin{pmatrix}0 & \sigma^{-1}\end{pmatrix}\begin{pmatrix}1 & \int B_{-\ d}(r)\,dr\\ \int B_{-\ d}(r)\,dr & \int\left(B_{-\ d}(r)\right)^2dr\end{pmatrix}^{-1}\begin{pmatrix}0\\ \sigma^{-1}\end{pmatrix}.$$

Given the consistency of $\hat{\alpha}_{ols}$ and $\hat{\phi}_{ols}$, it is easy to show that:

$$s_T^2\xrightarrow{p}\sigma^2.$$

Therefore:

$$T^{2(1-d)}\hat{\sigma}^2_{\hat{\phi}_{ols}} \xrightarrow{w} \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \int B_{-d}(r)\,dr \\ \int B_{-d}(r)\,dr & \int (B_{-d}(r))^2\,dr \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$= \frac{1}{\int (B_{-d}(r))^2\,dr - \left(\int B_{-d}(r)\,dr\right)^2}, \tag{51}$$

and this implies that:

$$t_{\hat{\phi}_{ola}} \xrightarrow{w} \frac{\int_0^1 B_{-d}(r)\,dB(r) - B(1)\int_0^1 B_{-d}(r)\,dr}{\left[\int_0^1 B^2_{-d}(r)\,dr - \left(\int_0^1 B_{-d}(r)\,dr\right)^2\right]^{1/2}},$$

or equivalently

$$t_{\hat{\phi}_{ola}} \xrightarrow{w} \frac{\int_0^1 B^{\mu}_{-d}(r)\,dB(r)}{\left[\int_0^1 \left(B^{\mu}_{-d}(r)\right)^2 dr\right]^{1/2}}.$$

2. $d = 0.5$.

Proceeding in the same way as before, is easy to check that in this case,

$$(T\log T)\,\hat{\sigma}^2_{\hat{\phi}_{ols}} = s_T^2 \begin{pmatrix} 0 & 1 \end{pmatrix} \Upsilon_T \begin{pmatrix} T & \sum_{t=1}^T \Delta^d y_{t-1} \\ \sum_{t=1}^T \Delta^d y_{t-1} & \sum_{t=1}^T \left(\Delta^d y_{t-1}\right)^2 \end{pmatrix}^{-1} \Upsilon_T \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

where in this case Υ_T is defined as in (46). Since s_T^2 is a consistent estimator of σ^2 then

$$(T\log T)\,\hat{\sigma}^2_{\hat{\phi}_{ols}} \xrightarrow{p} \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\pi} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \pi,$$

which implies the desired result.

3. $0.5 < d < 1$.

The proof of this part is analogous to the previous one just by taking the scaling matrix Υ_T equal to $T^{1/2} I_2$.

Proof of Theorem 3

The OLS estimates of $(\alpha, \delta, \phi)'$ are given by

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\delta} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} T & \sum t & \sum \Delta^d y_{t-1} \\ \sum t & \sum t^2 & \sum t\Delta^d y_{t-1} \\ \sum \Delta^d y_{t-1} & \sum t\Delta^d y_{t-1} & \sum (\Delta^d y_{t-1})^2 \end{pmatrix}^{-1} \times \begin{pmatrix} \sum \varepsilon_t \\ \sum t\varepsilon_t \\ \sum \Delta^d y_{t-1}\varepsilon_t \end{pmatrix}. \tag{52}$$

I. First case: $0 \leq d < 0.5$. In this case, $\Delta^{\ d} y_{t-1}$ is a nonstationary $FI\,(1- \ d)$ and the following convergence hold (see Dolado and Mármol, 2001):

$$T^{5/2- \ d} \sum t\Delta^{\ d} y_{t-1} \xrightarrow{w} \sigma \int_0^1 rB_{- \ d}\,(r)\,dr. \tag{53}$$

This implies that:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\delta} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} O_p\,(T) & O\left(T^2\right) & O_p\left(T^{3/2- \ d}\right) \\ O(T^2) & O\left(T^3\right) & O_p\left(T^{5/2- \ d}\right) \\ O_p\left(T^{3/2- \ d}\right) & O_p\left(T^{5/2- \ d}\right) & O_p\left(T^{2(1- \ d)}\right) \end{pmatrix}^{-1} \times \begin{pmatrix} O_p\left(T^{1/2}\right) \\ O_p\left(T^{3/2}\right) \\ O_p\left(T^{1- \ d}\right) \end{pmatrix}. \tag{54}$$

Define the scaling matrix:

$$\Upsilon_T = \begin{pmatrix} T^{1/2} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & T^{1- \ d} \end{pmatrix}. \tag{55}$$

Then,

$$\Upsilon_T \begin{pmatrix} \hat{\alpha} \\ \hat{\delta} \\ \hat{\phi} \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & T^{-2}\sum t & T^{-(3/2- \ d)} \sum \Delta^{\ d} y_{t-1} \\ T^{-2}\left(\sum t\right) & T^{2-3}\left(\sum t^2\right) & T^{-(5/2- \ d)} \sum t\Delta^{\ d} y_{t-1} \\ T^{-(3/2- \ d)} \sum \Delta^{\ d} y_{t-1} & T^{-(5/2- \ d)} \sum t\Delta^{\ d} y_{t-1} & T^{-2(1- \ d)} \sum (\Delta^{\ d} y_{t-1}) \end{pmatrix}^{-1} \times \begin{pmatrix} T^{-1/2}\sum \varepsilon_t \\ T^{-3/2}\sum t\varepsilon_t \\ T^{-(1- \ d)} \sum \Delta^{\ d} y_{t-1}\varepsilon_t \end{pmatrix}, \tag{56}$$

which implies,

$$\Upsilon_T \begin{pmatrix} \hat{\alpha} \\ \hat{\delta} \\ \hat{\phi} \end{pmatrix} \xrightarrow{w} \begin{pmatrix} 1 & \frac{1}{2} & \sigma \int_0^1 B_{- \ d}\,(r)\,dr \\ \frac{1}{2} & \frac{1}{3} & \sigma \int_0^1 rB_{- \ d}\,(r)\,dr \\ \sigma \int_0^1 B_{- \ d}\,(r)\,dr & \sigma \int_0^1 rB_{- \ d}\,(r)\,dr & \sigma^2 \int_0^1 B^2_{-d}\,(r)\,dr \end{pmatrix}^{-1} \times \begin{pmatrix} \sigma B(1) \\ \sigma B(1) - \sigma \int_0^1 B\,(r)\,dr \\ \sigma^2 \int_0^1 B_{- \ d}\,(r)\,dB(r) \end{pmatrix}. \tag{57}$$

Defining

$$B_{-d}^{\tau}(s) = B_{-d}(s) + (6s-4)\int_0^1 B_{-d}(r)\,dr - (12s-6)\int_0^1 rB_{-d}(r)\,dr,$$

it is straight forward to show that

$$\hat{\phi} \xrightarrow{w} \frac{\int_0^1 B_{-d}^{\tau}(r)\,dB(r)}{\left(\int_0^1 \left(B_{-d}^{\tau}(r)\right)^2\right)^{1/2}}.$$

II. Second case: $d = 0.5$

Define the weighting matrix:

$$\Upsilon_T = \begin{pmatrix} T^{1/2} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & (T\log T)^{1/2} \end{pmatrix}. \tag{58}$$

Then,

$$\Upsilon_T \begin{pmatrix} \hat{\alpha} \\ \hat{\delta} \\ \hat{\phi} \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & T^{-2}\sum t & T^{-1}(\log T)^{-1/2}\sum \Delta^{\ d} y_{t-1} \\ T^{-2}\sum t & T^{-3}\left(\sum t^2\right) & (\log T)^{-1/2} T^{-2}\sum t\Delta^{\ d} y_{t-1} \\ T^{-1}(\log T)^{-1/2}\sum \Delta^{\ d} y_{t-1} & (\log T)^{-1/2} T^{-2}\sum t\Delta^{\ d} y_{t-1} & (T\log T)^{-1}\sum(\Delta^{\ d} y_{t-1})^2 \end{pmatrix}$$
$$\times \begin{pmatrix} T^{-1/2}\sum \varepsilon_t \\ T^{-3/2}\sum t\varepsilon_t \\ (T\log T)^{-1/2}\sum \Delta^{\ d} y_{t-1}\varepsilon_t \end{pmatrix}. \tag{59}$$

Since

$$\begin{pmatrix} 1 & T^{-2}\sum t & T^{-1}(\log T)^{-1/2}\sum \Delta^{\ d} y_{t-1} \\ T^{-2}\left(\sum t\right) & T^{-3}\left(\sum t^2\right) & (\log T)^{-1/2} T^{-2}\sum t\Delta^{\ d} y_{t-1} \\ T^{-1}(\log T)^{-1/2}\sum \Delta^{\ d} y_{t-1} & (\log T)^{-1/2} T^{-2}\sum t\Delta^{\ d} y_{t-1} & (T\log T)^{-1}\sum(\Delta^{\ d} y_{t-1})^2 \end{pmatrix}^{-1}$$

$$\xrightarrow{p} \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{\sigma^2}{\pi} \end{pmatrix}^{-1}, \tag{60}$$

and

$$\begin{pmatrix} T^{-1/2}\sum\varepsilon_t \\ T^{-3/2}\sum t\varepsilon_t \\ (T\log T)^{-1/2}\sum\Delta\ ^d y_{t-1}\varepsilon_t \end{pmatrix} \xrightarrow{w} N\left(0,\sigma^2\begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{\sigma^2}{\pi} \end{pmatrix}\right).$$

Then

$$\Upsilon_T\begin{pmatrix} \hat{\alpha} \\ \hat{\delta} \\ \hat{\phi} \end{pmatrix} \xrightarrow{w} N\left(0,\sigma^2\begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{\sigma^2}{\pi} \end{pmatrix}^{-1}\right) = N\left(0,\begin{pmatrix} 4\sigma^2 & -6\sigma^2 & 0 \\ 0 & 12\sigma^2 & 0 \\ -6\sigma^2 & 0 & \pi \end{pmatrix}\right). \tag{61}$$

III. Third case: $0.5 < d < 1$.

In this case the process $\Delta^d y_{t-1}$ is stationary, then define:

$$\Upsilon_T = \begin{pmatrix} T^{1/2} & 0 & 0 \\ 0 & T^{1/2} & 0 \\ 0 & 0 & T^{3/2} \end{pmatrix}. \tag{62}$$

Taking into account that

$$\begin{pmatrix} 1 & T^{-2}\sum t & T^{-1}\sum\Delta\ ^d y_{t-1} \\ T^{-2}\left(\sum t\right) & T^{-3}\left(\sum t^2\right) & T^{-2}\sum t\Delta\ ^d y_{t-1} \\ T^{-1}\sum\Delta\ ^d y_{t-1} & T^{-2}\sum t\Delta\ ^d y_{t-1} & T^{-1}\sum(\Delta\ ^d y_{t-1})^2 \end{pmatrix}^{-1} \xrightarrow{p} \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & 0 & \sigma^2\sum_{i=1}^{\infty}\pi_i^2\,(1-d) \end{pmatrix}^{-1},$$

and that

$$\begin{pmatrix} T^{-1/2}\sum\varepsilon_t \\ T^{-3/2}\sum t\varepsilon_t \\ (T\log T)^{-1/2}\sum\Delta^d y_{t-1}\varepsilon_t \end{pmatrix} \xrightarrow{w} N\left(0,\sigma^2\begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & 0 & \sigma^2(\sum_{i=1}^{\infty}\pi_i^2\,(1-d))^{1/2} \end{pmatrix}\right),$$

then

$$\Upsilon_T\begin{pmatrix} \hat{\alpha} \\ \hat{\delta} \\ \hat{\phi} \end{pmatrix} \xrightarrow{w} N\left(0,\sigma^2\begin{pmatrix} 4\sigma^2 & -6\sigma^2 & 0 \\ -6\sigma^2 & 12\sigma^2 & 0 \\ 0 & 0 & \frac{\Gamma^2(d)}{\Gamma(2d-1)} \end{pmatrix}\right).\blacksquare \tag{63}$$

Proof of Theorem 4

The proof of this theorem is analogous to that in Theorem 2 and therefore it is omitted.

Proof of Lemma 3

Consider the process y_t defined according to (26). The differenced process $\Delta\ ^d y_t$ can be rewritten as:

$$\Delta^d y_t = \Delta^{d-1}\alpha 1_{(t>0)} + \Delta^{d-1}\varepsilon_t 1_{(t>0)} = \alpha\sum_{i=0}^{t-1}\pi_i\,(d-1) + \sum_{i=0}^{t-1}\pi_i\,(d-1)\,\varepsilon_{t-i}. \tag{64}$$

1. Expression (64) implies,

$$\sum_{t=1}^{T}\Delta^{d}y_{t}=\alpha\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_{i}\left(d-1\right)\right)+\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_{i}\left(d-1\right)\varepsilon_{t-i}\right) \tag{65}$$

The first term in the RHS of (65) is $O_{p}\left(T^{2-d}\right)$ since

$$\begin{aligned} T^{-(2-d)}\lim_{T\rightarrow\infty}\alpha\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_{i}\left(d-1\right)\right) &= T^{-(2-d)}\frac{\alpha}{\Gamma\left(1-\ d\right)\left(1-d\right)}\lim_{T\rightarrow\infty}\sum_{t=1}^{T}t^{1-\ d} \\ &= \frac{\alpha}{\Gamma\left(1-d\right)\left(1-d\right)\left(2-d\right)} \\ &= \frac{\alpha}{\Gamma\left(3-d\right)}. \end{aligned} \tag{66}$$

The second term in the RHS of (65) is the sum over time of a $FI\left(1-d\right)$ process and therefore is $O_{p}\left(T^{3/2-d}\right)$, $O_{p}\left(T\left(\log T\right)^{1/2}\right)$ or $O_{p}\left(T\right)$ according to whether $d\in\left[0,0.5\right)$, $d=0.5$ and $d\in\left(0.5,1\right)$ respectively and therefore, converges to zero when divided by $T^{2-\ d}$.

2. The process $\sum_{t=1}^{T}\left(\Delta y_{t}\right)^{2}$ is given by

$$\begin{aligned} \sum_{t=1}^{T}\left(\Delta\ ^{d}y_{t}\right)^{2} &= \alpha^{2}\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_{i}\left(\ d-1\right)\right)^{2}+\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_{i}\left(\ d-1\right)\varepsilon_{t}\right)^{2} \\ &\quad +2\alpha\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_{i}\left(\ d-1\right)\right)\Delta\ ^{d-1}\varepsilon_{t}. \end{aligned} \tag{67}$$

The first term in the RHS of (67) is $O_{p}\left(T^{3-2d}\right)$ and converges to

$$T^{-(3-2\ d)}\lim_{T\rightarrow\infty}\alpha^{2}\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_{i}\left(\ d-1\right)\right)^{2}=\frac{\alpha^{2}}{\Gamma^{2}\left(2-\ d\right)\left(3-2\ d\right)}. \tag{68}$$

The second term in (67) is $\mathrm{O}_{p}\left(T^{2(1-\ d)}\right)$, $O_{p}\left(T\log T\right)$ or $O_{p}\left(T\right)$ according to whether $d\in\left[0,0.5\right)$, $d=0.5$ or $d\in\left(0.5,1\right)$ respectively. Finally, the third term in the RHS of (67) is $O_{p}\left(T^{5/2-d}\right)$ (see Dolado and Marmol, 2001), which implies that again the first term is the leading one and the whole expression is $O_{p}\left(T^{3-2d}\right)$ being its limit that of (68) .■

Proof of Lemma 4.

Notice that

$$\sum_{t=2}^{T}\Delta^{d}y_{t-1}\varepsilon_{t}=\alpha\sum_{t=2}^{T}\left(\sum_{i=0}^{t-1}\pi_{i}\left(d-1\right)\right)\varepsilon_{t}+\sum_{t=2}^{T}\left(\sum_{i=0}^{t-1}\pi_{i}\left(d-1\right)\varepsilon_{t-i}\right)\varepsilon_{t}. \tag{69}$$

The first term in the RHS of expression (69) is a martingale difference sequence and verifies a CLT for this type of processes. To show that this is the case, it is necessary to check that the sequence $\{\left(\sum_{i=0}^{t-1}\pi_i(d-1)\right)\varepsilon_t\}$ verifies the conditions of the standard Central Limit Theorem (CLT) for martingale difference sequences (m.d.s.) (see Hall and Heyde, 1980, Chapter 3 and Helland, 1982). These conditions are: i) the sequence is a m.d.s., ii) the sum of the conditional variances tends to unity and iii) the Lindeberg condition (LC) holds.

Define:

$$\tilde{\varepsilon}_t = \sigma^{-1}\varepsilon_t, \tag{70}$$

$$\tilde{x}_t = \left((T^{(2-d)}\frac{\alpha^2}{\Gamma^2(2-d)(3-2d)}\right)^{-1/2}\sum_{i=0}^{t-1}\pi_i(1-d), \tag{71}$$

and

$$X_{T.t} = T^{-1/2}\tilde{x}_{t-1}\tilde{\varepsilon}_t. \tag{72}$$

Let $F_{T.t}$ be an array of $\sigma-$fields such that $F_{T.t-1}\subset F_{T.t}$. Condition (i) is fulfilled since

$$T^{-1/2}E(\tilde{x}_{t-1}\tilde{\varepsilon}_t|F_{T.t-1}) = T^{-1/2}\tilde{x}_{t-1}E(\tilde{\varepsilon}_t|F_{T.t-1}) = 0. \tag{73}$$

since $\tilde{\varepsilon}_t$ is m.d.s. Regarding condition (ii), we have

$$\begin{aligned} T^{-1}\sum_{t=2}^{T}Var(\tilde{x}_{t-1}\tilde{\varepsilon}_t|F_{T.t-1}) &= T^{-1}\sum_{t=2}^{T}\tilde{x}_{t-1}^2E\left(\tilde{\varepsilon}_t^2|F_{T.t-1}\right) - \tilde{x}_{t-1}^2E(\tilde{\varepsilon}_t|F_{T.t-1})^2 \\ &= T^{-1}\sum\tilde{x}_{t-1}^2 \xrightarrow{p} 1. \end{aligned} \tag{74}$$

Finally, condition (iii) holds since

$$\sum_{t=1}^{T}E\left(|X_{T.t}|^2 I\{|X_{T.t}|>\varrho|\}\right) = E\left(|\tilde{x}_{t-1}\tilde{\varepsilon}_t|^2 I_{\{|\tilde{x}_{t-1}\tilde{\varepsilon}_t|>T^{1/2}\varrho|\}}\right)\to 0,\ \text{for all } \varrho>0. \tag{75}$$

Conditions $(73), (74)$ and (75) jointly imply the desired result. The proof for the truncated process x_t^* is similar. Condition (i) holds since x_{t-1}^* and ε_t are independent. Condition (ii) holds since $T^{-1}\left(\sum(\tilde{x}_{t-1}^*)^2 - \sum(\tilde{x}_{t-1})^2\right) = o_p(1)$ (see Lemma 1) implies $T^{-1}\sum(\tilde{x}_{t-1}^*)^2 \xrightarrow{p} 1$. Lastly, a sufficient condition for condition (iii) is Liapunov's condition, $1/T^2\sum_{t=1}^{T}E\left(\tilde{x}_{t-1}^4\tilde{\varepsilon}_t^4\right)\to 0$. To prove this, consider

$$\begin{aligned} \frac{1}{T^2}\sum_{t=1}^{T}E\left(\tilde{x}_{t-1}^4\tilde{\varepsilon}_t^4\right) &= \frac{1}{T^2}\mu_4\sum_{t=1}^{T}\tilde{x}_{t-1}^4 = \\ & \frac{1}{T^{6-2d}}\mu_4\left(\frac{\alpha^2}{\Gamma^2(2-d)(3-2d)}\right)^{-2}\sum_{t=1}^{T-2}\left(\sum_{i=0}^{t-1}\pi_i(1-d)\right)^4. \end{aligned} \tag{76}$$

Noticing that $\pi_i(d-1) = i^{-d}$, it is easy to check that $\sum_{t=1}^{T-2}\left(\sum_{i=0}^{t-1}\pi_i(1-d)\right)^4$ is $O_p\left(T^{5-4d}\right)$ which implies that (76) tends to zero for all $d > -1/2$.

The second term of the RHS of (69) is $O_p\left(T^{1-d}\right)$ if $0 \leq d < 0.5$, $O_p\left((T\log T)^{1/2}\right)$ if $d = 0.5$ and $O_p\left(T^{0.5}\right)$ when $0.5 < d < 1$ (see lemma 1) and therefore vanishes when divided by $T^{3/2-d}$.■

Proof of Theorem 5

Define the following scaling matrix,

$$\Upsilon_T = \begin{pmatrix} T^{1/2} & 0 \\ 0 & T^{3/2-d} \end{pmatrix}, \tag{77}$$

and consider

$$\Upsilon_T \begin{pmatrix} \hat{\alpha}-\alpha \\ \hat{\phi}_{ols} \end{pmatrix} = \Upsilon_T \begin{pmatrix} T & \sum_{t=2}^{T}\Delta^d y_{t-1} \\ \sum_{t=2}^{T}\Delta^d y_{t-1} & \sum_{t=2}^{T}\left(\Delta^d y_{t-1}\right)^2 \end{pmatrix}^{-1} \Upsilon_T \Upsilon_T^{-1} \begin{pmatrix} \sum_{t=1}^{T}\varepsilon_t \\ \sum_{t=2}^{T}\varepsilon_t\Delta^d y_{t-1} \end{pmatrix}, \tag{78}$$

$$\Upsilon_T \begin{pmatrix} \hat{\alpha}-\alpha \\ \hat{\phi}_{ols} \end{pmatrix} = \begin{pmatrix} 1 & \frac{\alpha}{\Gamma(1-d)(1-d)} \\ \frac{\alpha}{\Gamma(1-d)(1-d)} & \frac{\alpha^2}{\Gamma(1-d)^2(1-d)^2(3-2d)} \end{pmatrix}^{-1} \Upsilon_T^{-1} \begin{pmatrix} \sum_{t=1}^{T}\varepsilon_t \\ \sum_{t=1}^{T}\varepsilon_t\Delta^d y_{t-1} \end{pmatrix} + o_p(1). \tag{79}$$

Noticing that

$$\Upsilon_T^{-1} \begin{pmatrix} \sum_{t=1}^{T}\varepsilon_t \\ \sum_{t=1}^{T}\varepsilon_t\Delta^d y_{t-1} \end{pmatrix} \xrightarrow{w} N\left(0, \sigma^2 Q_2\right),$$

where

$$Q_2 = \begin{pmatrix} 1 & \frac{\alpha}{\Gamma(3-d)} \\ \frac{\alpha}{\Gamma(3-d)} & \frac{\alpha^2}{\Gamma(2-d)^2(3-2d)} \end{pmatrix},$$

we obtain the desired result,

$$\Upsilon_T \begin{pmatrix} \hat{\alpha}-\alpha \\ \hat{\phi}_{ols} \end{pmatrix} \xrightarrow{w} N\left(0, \sigma^2 Q_2^{-1} Q_2 Q_2^{-1}\right) = N\left(0, \sigma^2 Q_2^{-1}\right). \tag{80}$$

Proof of Theorem 6

The proof of Theorem 6 is similar to that of Theorem 2, just by considering the scaling matrix defined in (77) and applying the results in Lemma 3 and therefore is omitted.■

Proof of Lemma 5

The LHS of (35) can be rewritten as

$$\sum t\Delta^d y_{t-1} = \alpha \sum t \left(\sum_{i=0}^{t-1} \pi_i (d-1) \right) + \sum t x_{t-1} \tag{81}$$

where x_t is a pure $FI(1-d)$ with no deterministic components. The first term of the RHS of (81) is completely deterministic and its limit is given by,

$$\lim_{T\to\infty} T^{-(3-d)} \alpha \sum t \left(\sum_{i=0}^{t-1} \pi_i (d-1) \right) = \lim_{T\to\infty} \frac{T^{-(3-d)}}{\Gamma(2-d)} \sum t^{2-d} = \frac{1}{\Gamma(2-d)(3-d)}.$$

The second term of the RHS of (81) is $O_p\left(T^{5/2-d}\right)$ (see Dolado and Mármol, 2001) and therefore the first term dominates.■

Proof of Theorem 7

1.The first part of the Theorem is the standard result for the Dickey-Fuller case (see DF, 1981). When $0 < d < 1$, let us define the scaling matrix:

$$\Upsilon_T = \begin{pmatrix} T^{1/2} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & T^{3/2-d} \end{pmatrix}, \tag{82}$$

then,

$$\Upsilon_T \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\delta} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} 1 & T^{-2}\left(\sum t\right) & T^{2-d}\left(\sum \Delta^d y_{t-1}\right) \\ T^{-2}\left(\sum t\right) & T^{3}\left(\sum t^2\right) & T^{3-d}\left(\sum t\Delta^d y_{t-1}\right) \\ T^{2-d}\left(\sum \Delta^d y_{t-1}\right) & T^{3-d}\left(\sum t\Delta^d y_{t-1}\right) & T^{3-2d}\left(\sum (\Delta^d y_{t-1})^2\right) \end{pmatrix}^{-1}$$

$$\times \Upsilon_T^{-1} \begin{pmatrix} \sum \varepsilon_t \\ \sum t\varepsilon_t \\ \sum \Delta^d y_{t-1}\varepsilon_t \end{pmatrix}. \tag{83}$$

The first term in the RHS of (**??**) converges in probability to:

$$\begin{pmatrix} 1 & 1/2 & \frac{\alpha}{\Gamma(3-d)} \\ 1/2 & 1/3 & \frac{\alpha}{\Gamma(2-d)(3-d)} \\ \frac{\alpha}{\Gamma(3-d)} & \frac{\alpha}{\Gamma(2-d)(3-d)} & \frac{\alpha^2}{\Gamma(2-d)^2(3-2d)} \end{pmatrix}, \tag{84}$$

and the second term converges weakly to

$$\begin{pmatrix} T^{-1/2}\sum \varepsilon_t \\ T^{3/2}\sum t\varepsilon_t \\ T^{3/2-d}\sum \Delta^d y_{t-1}\varepsilon_t \end{pmatrix} \xrightarrow{w} N\left(0, \sigma^2 Q_3\right). \tag{85}$$

Expressions (84) and (85) imply the desired result.■

Proof of Theorem 8

Again, the proof of this theorem is omitted since it is similar of that of Theorem 2.■

Proof of Theorem 9

When d is chosen according to the criterion function in (38), and since $\hat{d}_T$ is a consistent estimator of d $(=1)$, it follows that $\hat{d} \xrightarrow{p} 1-c$. Applying the mean value theorem (MVT) on $t^{\mu}_{\phi_{ols}}$ around the point $(1-c)$, it is obtained that

$$t^{\mu}_{\phi_{ols}}\left(\hat{d}\right) = t^{\mu}_{\phi_{ols}}(1-c) + \frac{\partial t^{\mu}_{\phi_{ols}}\left(\breve{d}\right)}{\partial d}\left(\hat{d}-(1-c)\right)$$

where $\breve{d}$ is an intermediate point between $\hat{d}$ and $(1-c)$. This implies that in order to prove $\left(t^{\mu}_{\phi_{ols}}\left(\hat{d}\right) - t^{\mu}_{\phi_{ols}}(1-c)\right) = o_p(1)$ it has to be shown that $\frac{\partial t^{\mu}_{\phi_{ols}}(\breve{d})}{\partial d}\left(\hat{d}-(1-c)\right) = o_p(1)$. Since $\breve{d} \in \left(\hat{d}, 1-c\right)$ then, $\breve{d} \xrightarrow{p} (1-c)$. Also, taking into account that $\left(\hat{d}-(1-c)\right) = O_p\left(T^{-1/2}\right)$, it is just needed to prove that $\frac{\partial t^{\mu}_{\phi_{ols}}(\breve{d})}{\partial d} = o_p\left(T^{1/2}\right)$. The $t-ratio$ is given by

$$t^{\mu}_{\phi_{ols}}\left(\hat{d}\right) = \frac{\sum \Delta y_t \Delta^{\hat{d}} y_{t-1} - T(\overline{\Delta^{\hat{d}} y}\,\overline{\Delta y})}{\hat{\sigma}\left(\sum\left(\Delta^{\hat{d}} y_{t-1}\right)^2 - T^{-1}\left(\sum \Delta^{\hat{d}} y_{t-1}\right)^2\right)^{1/2}}.$$

Let consider first the case where y_t is a random walk without drift under the null hypothesis. Under the later hypothesis, the $t-statistic$ can be rewritten as

$$t^{\mu}_{\phi_{ols}}\left(\hat{d}\right) = \frac{\sum \varepsilon_t \Delta^{\hat{d}-1}\varepsilon_{t-1}1_{(t-1>0)} - \overline{\varepsilon}\sum \Delta^{\hat{d}-1}\varepsilon_{t-1}}{\hat{\sigma}\left(\sum\left(\Delta^{\hat{d}-1}\varepsilon_{t-1}\right)^2 - T^{-1}\left(\sum \Delta^{\hat{d}-1}\varepsilon_{t-1}\right)^2\right)^{1/2}}. \tag{86}$$

The first derivative of $\partial t^{\mu}_{\phi_{ols}}(d)$, evaluated at $d=(1-c)$, is given by

$$\left.\frac{\partial t^{\mu}_{\phi_{ols}}(d)}{\partial d}\right|_{d=1-c} = \left.\frac{(\partial A_1/\partial d)\,A_2 - A_1\partial A_2/\partial d}{\hat{\sigma}^2\left(\sum\left(\Delta^{d-1}\varepsilon_{t-1}\right)^2 - T^{-1}\left(\sum \Delta^{d-1}\varepsilon_{t-1}\right)^2\right)}\right|_{d=1-c},$$

where A_1 and A_2 are the numerator and denominator of (86) respectively evaluated in $d=1-c$. Taking into account that (see DGM, 2001 for details)

$$\frac{\partial A_1}{\partial d} = \sum \varepsilon_t\left(\log(1-L)\,\Delta^{-c}\varepsilon_{t-1}1_{(t-1>0)}\right) - \overline{\varepsilon}\left(\sum \log(1-L)\,\Delta^{-c}\varepsilon_{t-1}1_{(t-1>0)}\right) = o_p(T),$$

$$\begin{aligned}
\frac{\partial A_2}{\partial d} &= \hat{\sigma}\frac{\left(\sum \log\left(1-L\right)\Delta^{-c}\varepsilon_{t-1}1_{(t-1>0)}\right) - T^{-1}\left(\sum \Delta^{d-1}\varepsilon_{t-1}\right)\left(\sum \log\left(1-L\right)\Delta^{-c}\varepsilon_{t-1}1_{(t-1>0)}\right)}{\left(\sum\left(\Delta^{d-1}\varepsilon_{t-1}\right)^2 - T^{-1}\left(\sum \Delta^{d-1}\varepsilon_{t-1}\right)^2\right)^{1/2}} \\
&\quad + \frac{\partial \hat{\sigma}}{\partial d}\left(\sum\left(\Delta^{d-1}\varepsilon_{t-1}\right)^2 - T^{-1}\left(\sum \Delta^{d-1}\varepsilon_{t-1}\right)^2\right)^{1/2} \\
&= O_p\left(1\right)\frac{o_p\left(T\right) + o_p\left(1\right)o_p\left(T\right)}{O_p\left(T^{1/2}\right)} + o_p\left(1\right)O_p(T^{1/2}) = o_p\left(T^{1/2}\right),
\end{aligned}$$

$$A_1 = o_p\left(T\right),$$

and

$$A_2 = O_p\left(T^{1/2}\right),$$

it follows that,

$$\left.\frac{\partial t^{\mu}_{\phi_{ols}}\left(d\right)}{\partial d}\right|_{d=1-c} = \frac{o_p\left(T\right)O_p\left(T^{1/2}\right) - o_p\left(T\right)o_p\left(T^{1/2}\right)}{O_p\left(T\right)} = o_p\left(T^{1/2}\right),$$

which implies the desired result.

The case where y_t is a random walk with a drift different from zero, is similar. In this case, the t statistic under the null hypothesis can be rewritten as

$$t^{\mu}_{\phi_{ols}}\left(\hat{d}\right) = \frac{\sum \varepsilon_t(\Delta^{\hat{d}-1}\alpha + \Delta^{\hat{d}-1}\varepsilon_{t-1})1_{(t-1>0)} - \overline{\varepsilon}\sum\left(\Delta^{\hat{d}-1}\alpha + \Delta^{\hat{d}-1}\varepsilon_{t-1}\right)}{\hat{\sigma}\left(\sum\left(\alpha\Delta^{\hat{d}-1} + \Delta^{\hat{d}-1}\varepsilon_{t-1}\right)^2 - T^{-1}\left(\sum \alpha\Delta^{\hat{d}-1} + \Delta^{\hat{d}-1}\varepsilon_{t-1}\right)^2\right)^{1/2}}. \tag{87}$$

The first derivative, evaluated at $d = (1-c)$, is given by

$$\left.\frac{\partial t^{\mu}_{\phi_{ols}}\left(d\right)}{\partial d}\right|_{d=1-c} = \left.\frac{\left(\partial A_3/\partial d\right)A_4 - A_3\partial A_4/\partial d}{\hat{\sigma}^2\left(\sum\left(\alpha\Delta^{d-1} + \Delta^{d-1}\varepsilon_{t-1}\right)^2 - T^{-1}\left(\sum \alpha\Delta^{d-1} + \Delta^{d-1}\varepsilon_{t-1}\right)^2\right)}\right|_{d=1-c},$$

where A_3 and A_4 are the numerator and denominator of (86) respectively evaluated in $d = 1-c$. Taking into account that (see DGM, 2001 for details)

$$\begin{aligned}
\frac{\partial A_3}{\partial d} &= \sum(\varepsilon_t\left(\log\left(1-L\right)(\Delta^{-c}\alpha + \Delta^{-c}\varepsilon_{t-1})1_{(t-1>0)}\right) \\
&\quad -(\overline{\varepsilon})\left(\sum \log\left(1-L\right)\left(\Delta^{-c}\alpha + \Delta^{-c}\varepsilon_{t-1}\right)1_{(t-1>0)}\right) \\
&= O_p\left(T^{1/2+c}\log T\right),
\end{aligned} \tag{88}$$

$$
\begin{aligned}
\frac{\partial A_4}{\partial d} &= \hat{\sigma}\frac{\sum(\Delta^{-c}\alpha+\Delta^{-c}\varepsilon_{t-1})\left(\log\left(1-L\right)\left(\Delta^{-c}\alpha+\Delta^{-c}\varepsilon_{t-1}\right)1_{(t-1>0)}\right)\right)}{\left(\sum\left(\alpha\Delta^{d-1}+\Delta^{d-1}\varepsilon_{t-1}\right)^2-T^{-1}\left(\sum\alpha\Delta^{d-1}+\Delta^{d-1}\varepsilon_{t-1}\right)^2\right)^{1/2}} \\
&\quad -\frac{T^{-1}\left(\sum(\Delta^{-c}\alpha+\Delta^{-c}\varepsilon_{t-1})\right)\left(\sum\log\left(1-L\right)\left(\Delta^{-c}\alpha+\Delta^{-c}\varepsilon_{t-1}\right)1_{(t-1>0)}\right)}{\left(\sum\left(\alpha\Delta^{d-1}+\Delta^{d-1}\varepsilon_{t-1}\right)^2-T^{-1}\left(\sum\alpha\Delta^{d-1}+\Delta^{d-1}\varepsilon_{t-1}\right)^2\right)^{1/2}} \\
&\quad +\frac{\partial\hat{\sigma}}{\partial d}\left(\sum\left(+\Delta^{d-1}\varepsilon_{t-1}\right)^2-T^{-1}\left(\sum\alpha\Delta^{d-1}+\Delta^{d-1}\varepsilon_{t-1}\right)^2\right)^{1/2} \\
&= O_p\left(1\right)\frac{O_p\left(T^{2c+1}\log T\right)}{O_p\left(T^{1/2}\right)}=O_p\left(T^{1/2+2c}\log T\right), \qquad (89)
\end{aligned}
$$

$$
A_3=O_p\left(T^{1/2+c}\right),
$$

and

$$
A_4=O_p\left(T^{1/2+c}\right),
$$

it follows that,

$$
\left.\frac{\partial t^{\mu}_{\phi_{ols}}\left(d\right)}{\partial d}\right|_{d=1-c}=\frac{O_p\left(T^{1/2+2c}\log T\right)O_p\left(T^{1/2+c}\right)}{O_p\left(T^{1+2c}\right)}=O_p\left(T^{c}\log T\right)=o_p\left(T^{1/2}\right). \qquad (90)
$$

The proof for the case where a deterministic trend is included is similar, although algebraically more complex, and therefore it is omitted.