Centre de Referència en Economia Analítica

Barcelona Economics Working Paper Series

Working Paper nº 29

Testing for a Unit Root Againts Fractional Alternatives in the Presence of a Maintained Trend

Juan J. Dolado, Jesús Gonzalo and Laura Mayoral

December, 2003

TESTING FOR A UNIT ROOT AGAINST FRACTIONAL ALTERNATIVES IN THE PRESENCE OF A MAINTAINED TREND

By Juan J. Dolado[a], Jesus Gonzalo[a], and Laura Mayoral[b] *
[a]Dept. of Economics, Universidad Carlos III de Madrid.
[b]Dept. of Economics, Universitat Pompeu Fabra.

Barcelona Economics WP #29

December 10, 2003

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 $d \in [0,1)$. Invariant tests to the presence of a drift under the null of $I(1)$ are derived. In common with the standard DF approach in the $I(1)$ $vs.$ $I(0)$ framework, we also examine the consequences of including a constant and /or a linear trend in the regression model when there is a drift under the null. A simple testing strategy entailing only asymptotically normally-distributed tests is proposed. Finally, an empirical application is provided where the FDF test allowing for deterministic components is used to test for long-memory in the per capita GDP of several OCDE countries.

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

INTRODUCTION

The goal of this paper is to extend an existing statistical procedure for detecting a unit root against mean-reverting fractional alternatives in time series which may exhibit a trending behavior or/and a non-zero initial conditions. In particular we focus on the test proposed

*Corresponding E-mail:jgonzalo@est-econ.uc3m.es. We are grateful to Claudio Michelacci and Bart Verspagen for making the data available, and to Francesc Mármol and participants in seminars at Ente Luigi Einaudi (Rome), ESEM 2002 (Venice) and CREST (Paris) for useful comments to earlier drafts of the paper. The usual disclaimer applies. Laura Mayoral wants to thank the support of the Barcelona Economics Program of CREA.

by Dolado, Gonzalo and Mayoral (2002, DGM henceforth) who have generalized the traditional Dickey-Fuller (DF henceforth) test of $I(1)$ against $I(0)$ processes to the broader framework of testing for a unit root against long-range dependence. Relying upon the DF approach, the underlying idea is to test for statistical significance of the coefficient on the regressor in a potentially unbalanced regression where the regressand is the filtered series under the null and the regressor is the lagged value of filtered series under the alternative hypothesis.

The advantages of this test, in parallel with the DF approach, rely on its simplicity and on its good performance in finite samples, both in terms of size and power. Specifically, when compared to other well-known tests for long memory, like the Lagrange Multiplier (LM) tests developed by Robinson (1994) and Tanaka (1999), the FDF test presents the advantage of not requiring the correct specification of a parametric model. For this reason, although the FDF test is not the asymptotically UIMP test (see Tanaka, 1999) under a sequence of alternatives approaching the null at the $\mathrm{T}^{-1/2}$ rate in a parametric model with gaussian errors, it fares very well in terms of power relative to both parametric and semiparametric tests in the frequency and time domains, and even better than the UIMP test when errors are non-gaussian, as discussed at length in DGM (2002).[1]

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 considering the OLS regression $\Delta y_t = \phi\ \Delta^d y_{t-1} + \varepsilon_t$, where ε_t is an i.i.d. disturbance, $\Delta = (1 - L)$ and L is the lag operator. To operationalise the FDF test, 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 $\hat{\phi}_{ols}$, $t_\phi(d)$, 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 all its roots outside the unit circle. The distribution of $t_\phi(d)$ depends on whether d is assumed (arbitrarily) pre-fixed (when considering a simple alternative) or estimated (when considering a composite alternative), and the distance $1-d$. When d is prefixed as in the standard DF case (where $d = 0$), the asymptotic distribution of the $t_\phi(d)$ is a $N(0, 1)$ variate when $0.5 \leq d < 1$, whilst it is nonstandard, i.e., a functional of Fractional Brownian motion (fBM henceforth), when $0 \leq d < 0.5$. In particular, for $d = 0$, $t_\phi(d)$ follows the well-known DF distribution, otherwise the critical values become less negative than the standard DF case as $d \to 0.5$. By contrast, whenever d is pre-estimated using any

[1] As shown in DGM (2002), the proposed test has also better power properties than those based on a direct estimation of d in semiparametric or parametric models since the former often yield large confidence intervals whilst the precision of the latter hinges on the correct specification of the model.

(trimmed) $T^{1/2}$-consistent estimator[2], $\widehat{d}$, of $d \in [0, 1)$, the asymptotic distribution of $t_\phi(\widehat{d})$ becomes pivotal and is always $N(0,1)$ for any value of d within the pre-specified range.[3]

Following the development of unit root tests in the past, where the initial canonical AR(1) model was subsequently augmented with further deterministic components (including linear, nonlinear and broken trends), our goal in this paper is to investigate how the limiting distribution of the FDF test changes when some deterministic components are considered in the DGP and in the maintained hypothesis. In particular, we will focus on the role of a *drift* and/or a *linear trend* in such a hypothesis. After all, many (macro) economic time series have a trending behavior in their levels 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 and a linear time 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 time trend in the maintained model allows one to achieve an invariant test to the presence of a drift in the true data generation process. When dealing with $FI(d)$ processes, two approaches have been considered in the literature to cater for deterministic components, both in the context of LM tests for the null of $d = d_0$. Some authors, like e.g. Robinson (1994), consider an additive setup where $y_t = \mu(t) + FI(d)$, with $\mu(t)$ being a vector of deterministic functions like a constant or a time trend, so that $E(\Delta^d y_t) = \Delta^d \mu(t)$. In this setup, our *first contribution* in this paper is to derive the corresponding (numerically) invariant test w.r.t. $\mu(t)$ for a unit root against a fractional one when $\mu(t) = \alpha + \beta t$ under the alternative hypothesis of an $FI(d)$ process. As will be shown below, invariance of the FDF test to the values of α and β is achieved by including the nonlinear trend $\Delta^d \mu(t)$ in the maintained hypothesis. When d is pre-fixed, the asymptotic distributions of the FDF test differ according to $0 \leq d < 0.5$ and $0.5 \leq d < 1$. However, if d is estimated with a (trimmed) $T^{1/2}$-consistent estimator, then the asymptotic distribution of the invariant FDF test is always $N(0,1)$.

By contrast, other authors, like e.g. Breitung and Hassler (2002), assume an innovative setup whereby $\Delta^d y_t = \mu(t) + I(0)$ so that $E(\Delta^d y_t) = \mu(t)$. In line with this approach, our *second contribution* deals with the case where only a constant and /or linear trend is included in the regression model, like in the standard DF approach to test $I(1)$ vs. $I(0)$

[2] A trimming such as the one proposed in DGM (2002, formula (33)) may be necessary in small samples to avoid estimates of d above 1.

[3] The intuition for 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 trimmed $T^{1/2}$-consistent estimator) asymptotic normality follows under the null hypothesis, whereas when they distant (i.e., when d belongs to the stationary range) the limiting distributions are nonstandard.

in the presence of deterministic components. In this case, the test is not similar since the value of the statistics depend upon the value of the deterministic components included in the DGP. When d is pre-fixed, three results arise. First, when the null is a driftless random walk, the FDF test with a constant and/or linear trend is non-standard for $0 \leq d < 0.5$ and asymptotically normal when $0.5 < d < 1$. Second, the result derived by West (1988) about the limiting $N(0,1)$ distribution of the DF test when both the DGP and the model share the same deterministic terms also holds for the FDF test, although the size of the true coefficients on those components relative to the variance of the disturbance, may lead to serious size distortions in finite samples. And third, when the null is a random walk with drift and a linear trend is included in the maintained hypothesis, in contrast to the non-standard limiting distribution of the DF test, a limiting $N(0,1)$ distribution holds for the FDF test, subject to similar size distortions in finite samples as before. Finally, as with the invariant FDF test, estimation of d using a (trimmed) $\mathrm{T}^{1/2}$-consistent procedure always yields an asymptotic $N(0,1)$ distribution. Hence, although the test is not numerically invariant to the values of the deterministic components in the DGP, asymptotically it is similar when an estimate of d is used, since in this case the asymptotic distributions of the test-statistics are always normal regardless the values of those latter components.

In view of these properties, our results provide a simple testing strategy of $H_0 : d = 1$ vs. $H_0 : 0 < d < 1$ when d is estimated to implement the FDF test in the presence of deterministic terms since only $N(0,1)$ critical values need to be used.

At this stage it should be pointed out that, in recent work, Lobato and Velasco (2003) have addressed the issue of optimality of the FDF test where the DGP is a pure $FI(d)$ process with no deterministic components and found that $T^{1/2}$- consistency in the estimation of d can be relaxed to $T^{1/4}\log(T)$- consistency. Since this condition holds for many semiparametric estimators with an appropriate choice of the bandwidth parameter (see Velasco, 1999) the range of estimators that can be used to implement the FDF test is much larger. However, investigating how this generalization extends to the presence of deterministic components exceeds the scope of this paper. Thus, in the sequel we will restrict our results to $T^{1/2}$- consistency although we conjecture that, under weaker conditions, their results may still hold.

Lastly, we wish to stress that, despite focusing on the case where the error term in the DGP 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 (augmented) DF case (ADF henceforth). In this respect, DGM (2002, 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). This procedure gives rise to the (augmented) FDF test (ADF henceforth) 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.

The rest of the paper is structured as follows. Sections 2 analyzes the derivation of invariant FDF tests when the null hypothesis is a random walk with or without drift. Section 3 focuses on the case where only a constant and/or a linear trend is used in the regression model emphasizing the differences with the standard results for the DF test and identifying those key parameters that may lead to poor finite-sample behavior 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. Section 4 discusses some empirical applications of the previous tests. Finally, Section 5 draws some concluding remarks.

Proofs of theorems and lemmae are gathered in Appendix 1 whereas sets of non-standard critical values for the FDF test with pre-fixed $d \in [0, 0.5)$ appear in Appendix 2.

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; $W_d(.)$ and $B(.)$ represent standard fBM corresponding to the limit distributions of the standardized partial sums of asymptotically stationary (truncated) $FI(d)$ processes as defined in Marinucci and Robinson (1999) and standard BM, respectively. Finally, $\xrightarrow{w}$ and $\xrightarrow{p}$ denote weak convergence and convergence in probability, respectively.

DEFINITION OF THE INVARIANT FDF TEST

Employing the methodology in DGM (2002), we assume, like in Robinson (1994), that the process y_t is generated as the sum of a constant and a linear time trend, and an integrated component u_t,

$$y_t = \alpha + \beta t + u_t, \tag{1}$$

with

$$u_t = \frac{\varepsilon_t 1_{(t>0)}}{\Delta - \phi \Delta^d L}, \tag{2}$$

where, for simplicity, ε_t is assumed to be an *i.i.d.* error term. Our interest is in $\mathrm{H}_0 : \phi = 0$ versus $\mathrm{H}_1 : \phi < 0$.

The null and alternative hypotheses can be rewritten as

$$H_0 : \Delta\left(y_t - \alpha - \beta t\right) = \varepsilon_t, \tag{3}$$

versus

$$H_1 : \left(\Delta - \phi\Delta^d L\right)\left(y_t - \alpha - \beta t\right) = \varepsilon_t, \tag{4}$$

where α represents the constant term, capturing the initial condition y_0, while β captures the trending behavior of the process in levels. Both can take are any real value (including zero). In DGM (2002) it is shown that $\left(\Delta - \phi\Delta^d L\right) = \Pi(L)\Delta^d$, where $\Pi(L) = (\Delta^{1-d} - \phi L)$ has all its roots outside the unit circle if $-2\phi^{1-d} < 0$, and verifies $\Pi(0) = 1$ and $\Pi(1) = -\phi$. Thus, under H_1, denoting $C(L) = \Pi(L)^{-1}$ with $C(0) = 1$ and $C(1) = -1/\phi$, y_t is governed by the process $\Delta^d\left(y_t - \alpha - \beta t\right) = C(L)\varepsilon_t$ with $C(L)$ having its roots outside the unit circle as well.

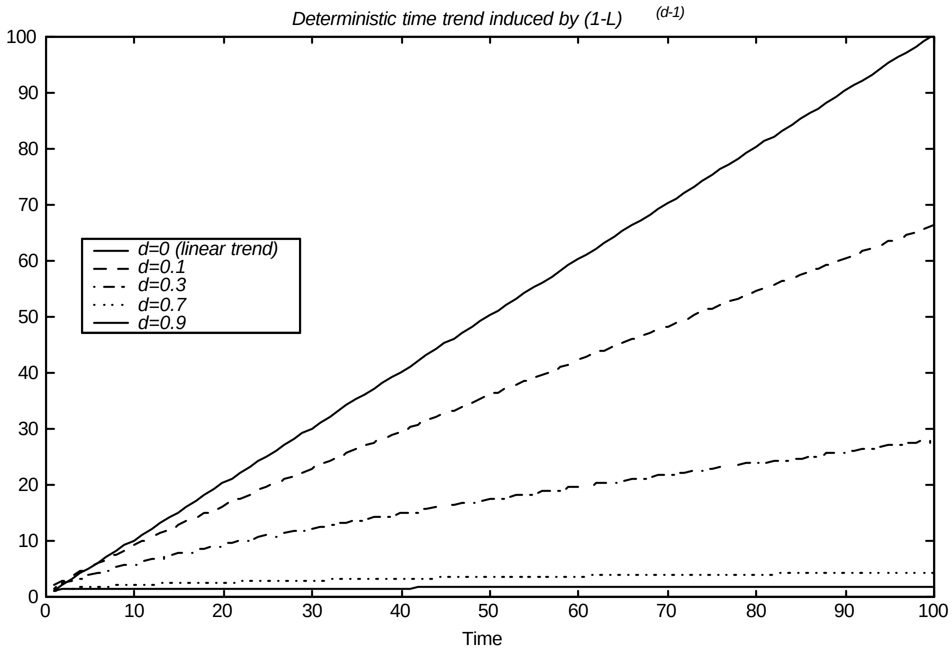
Premultiplying (1) by the polynomial $(\Delta - \phi\Delta^d L)$ we get the following regression model (denoted as *RM,* henceforth) as the maintained hypothesis

$$RM\ 1\text{:}\ \Delta y_t = \phi\Delta^d y_{t-1} + \beta - \phi\alpha\tau_{t-1}(d) - \phi\beta\tau_{t-1}(d-1) + \varepsilon_t, \tag{5}$$

with

$$\tau_t(\delta) = \sum_{i=0}^{t-1} \pi_i(\delta),$$

where the coefficients $\pi_i(\delta)$ come from the binomial expansion of $(1-L)^{\delta}$ in powers of L. Note that $\Delta^d t = \Delta^d \Delta^{-1} 1_{\{t>0\}}$ so that, in accord with the notation used above, such a trend is labelled in the sequel as $\tau_{t-1}(d-1)$. Both nonlinear time trends capture the trending behavior of the series under the alternative. Notice that the DF case for $d = 0$ is embedded in this setup since $\tau_{t-1}(0) = 1$ and $\tau_{t-1}(-1) = t-1$, giving rise to a constant and a linear time trend in the maintained hypothesis. As for the intermediate cases, Figure 1 plots a range of time trends generated with different values of $d \in [0, 1)$.



The test of $H_0 : I(1)$ relies upon $\phi = 0$ in model (5). Thus, when H_0 is true, the process becomes

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

whereas, under H_1, it is a $FI(d)$ process with a linear time trend like in (4).

If the presence of the linear trend in the level of the series is discarded from the outset (as e.g. in interest rates or in exchange rates) then $\beta = 0$ in (1), giving rise to

$$DGP\ 2 \colon \Delta y_t = \varepsilon_t,\ t \geq 1. \tag{7}$$

with the corresponding regression becoming

$$RM\ 2 :\ \Delta y_t = \phi \Delta^d y_{t-1} + \alpha_1 \tau_{t-1}(d) + \varepsilon_t. \tag{8}$$

As in the traditional DF framework, it is not difficult to prove that the t-ratio on the OLS estimator of ϕ in either (5) or (8), denoted as $t_{\hat{\phi}^{\tau}_{ols}}$ and $t_{\hat{\phi}^{\mu}_{ols}}$, respectively, is numerically invariant to the (unknown) value of α and β. In the following theorem, the asymptotic properties of the test under the null hypothesis are presented.

Theorem 1 *Under the null hypothesis that y_t is generated by* DGP 1 (DGP 2)*, the OLS coefficient associated to ϕ in* RM 1, $\hat{\phi}^{\tau}_{ols}$, *(to ϕ in* RM 2, $\hat{\phi}^{\mu}_{ols}$, *when $\beta = 0$) is a consistent estimator of $\phi = 0$ and converges to its true value ($\phi = 0$) at a rate T^{1-d} when $0 \leq d < 0.5$ and at the standard rate $T^{1/2}$ when $0.5 < d < 1$. The asymptotic distributions of the associated $t-ratios$, $t_{\hat{\phi}^{\tau}_{ols}}$ and $t_{\hat{\phi}^{\mu}_{ols}}$ are given by*

$$t_{\hat{\phi}^{i}_{ols}} \xrightarrow{w} \Lambda_i(d) \qquad \text{if } 0 < d < 0.5,\ \text{for } i = \{\mu, \tau\},$$

and,

$$t_{\hat{\phi}^i_{ols}} \xrightarrow{w} N(0,1) \qquad if\, 0.5 < d < 1, for\ i = \{\mu, \tau\},$$

where $\Lambda_i(d), i = \{\mu, \tau\}$ are functionals of fBM (see Appendix 1) that depend on d but not on other parameters of the model.

If rather than assuming that the integration order under the alternative hypothesis is an (arbitrary) pre-fixed value, d is estimated then *RM 1* would be as follows

$$\Delta y_t = \phi \Delta^{\widehat{d}_T} y_{t-1} + \alpha_1 + \alpha_2 \tau_{t-1}(\widehat{d}_T) + \alpha_3 \tau_{t-1}\left(\widehat{d}_T - 1\right) + \varepsilon_t, \tag{9}$$

where $\widehat{d}_T$ ($\widehat{d}$ in short-hand notation) is a (trimmed) $T^{1/2}-$consistent estimate of d.[4] If no drift is allowed under H_0, like in (8), then the model would be

$$\Delta y_t = \phi \Delta^{\widehat{d}} y_{t-1} + \alpha_1 \tau_{t-1}(\widehat{d}) + \varepsilon_t, \tag{10}$$

As discussed in DGM (2002), 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 (2003) can be used. Besides, as discussed in the Introduction, Lobato and Velasco (2003) have relaxed the assumption of $T^{1/2}$-consistency to $T^{1/4}\log(T)$- consistency of the estimator in a simple context, which imply that also semiparametric estimators of d could be used, although the following theory does not cover this case.

Theorem 2 *Let be a (trimmed) $T^{1/2}-$consistent estimator of d, $0 \leq d < 1$, such that $T^{1/2}(\widehat{d}_T - d) \xrightarrow{w} \xi$, where ξ is a non-degenerate distribution. Then, under the null hypothesis that y_t is a random walk with initial condition y_0, the asymptotic distribution of the $t-$ratios on the OLS coefficient associated to ϕ in (9) and in (10), $t_{\hat{\phi}^\tau_{ols}}(\widehat{d})$ and $t_{\hat{\phi}^\mu_{ols}}(\widehat{d})$ are given by*

$$t_{\hat{\phi}^i_{ols}}(\widehat{d}) \xrightarrow{w} N(0,1), \ i = \{\mu, \tau\}.$$

It is important to notice that use of the Frisch-Waugh Theorem and invariance to the presence of deterministic components in *RM 1* and *RM 2* imply that the test statistics can also be computed in a two-step procedure as follows. First, regress y_t on a linear trend (*RM 1)* or only on a constant term *(RM 2)* and obtain the residuals $\widehat{u}_t$. Second, construct the filtered series $\Delta\ \widehat{u}_t$ and $\Delta^d\ \widehat{u}_{t-1}$ and compute the t-ratio of the estimated coefficient in the regression of $\Delta\ \widehat{u}_t$ on $\Delta^d\ \widehat{u}_{t-1}$where d can be pre-fixed or estimated.

[4] Effectively, if $\widetilde{d}_T$ is a $\mathrm{T}^{1/2}$-consistent estimator of d, $\widehat{d}_T = \widetilde{d}_T$, if $\widetilde{d}_T < 1 - c$, and $\widehat{d}_T = 1 - c$, if $\widetilde{d}_T \geq 1$, where $c > 0$ is a (fixed) value in the neighborhood of zero that ensures that $\widehat{d}_T$ is strictly smaller than unity.

To check how the previous asymptotic results perform in finite samples, Tables A2 a,b (Appendix 2) report the empirical critical values of $t_{\hat{\phi}^{\tau}_{ols}}$ and $t_{\hat{\phi}^{\mu}_{ols}}$ in *RM 1* and *RM 2* for different significance levels and different values of d. The results are based on a Monte-Carlo study with a number of replications $N = 10,000$ of *DGP 2* (since the test is invariant to the value of α and β) where $\sigma_{\varepsilon} = 1$ and $T = 100, 400, 1000$. The critical values for $d \in [0, 0.5)$ are clearly different from those corresponding to a one-sided test using a standardized $N(0, 1)$ 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. As for the case where d is estimated using Mayoral´s (2003) Minimum Distance estimator which satisfies the requirements above, the empirical sizes at the 5% nominal level are 5.18, 5.12 and 4.98 for T= 100, 400 and 1000, respectively. As for power, Table 1 shows the rejection rates of the FDF test with estimated d test under the alternative. The most relevant finding is that there does not seem to be a power loss in finite samples when deterministic components are included relative to the case where they are not (see DGM, 2002, Table 5).

TABLE 1

POWER (SIZE CORRECTED). S.L. 5%

DGP 2: $\Delta y_t = \varepsilon_t$; *Estimated d*						
	RM 1			*RM* 2		
	$T = 100$	$T = 400$	$T = 1000$	$T = 100$	$T = 400$	$T = 1000$
0.9	23.7%	61.4%	94.5%	23.3%	62.7%	95.5%
0.7	84.9%	100%	100%	89.9%	100%	100%
0.6	97.4%	100%	100%	99.6%	100%	100%
0.4	100%	100%	100%	100%	100%	100%
0.2	100%	100%	100%	100%	100%	100%

PROPERTIES OF THE FDF TEST IN THE PRESENCE OF A CONSTANT/LINEAR TIME TREND IN DGP AND MODEL

In parallel with the conventional DF procedure to deal with deterministic components, we now consider the case where only a constant term and/or a linear time trend, rather than the nonlinear trends discussed above, are included in the regression model. Thus, in accord with the setup in Breitung and Hassler (2002), we deal now with a procedure for testing the null hypothesis that y_t has a unit root with or without drift versus the alternative 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 that y_t is generated according to

$$y_t = \frac{\beta + \delta t + \varepsilon_t 1_{(t>0)}}{\Delta - \phi \Delta^d L}, \tag{11}$$

where δ is a multiple of ϕ, say $\delta = k\phi$.[5] Then, under H_0: $\phi = 0$, becomes *DGP 1* above

$$y_t = \beta + y_{t-1} + \varepsilon_t,\ t \geq 1, \tag{12}$$

if $\beta \neq 0$, whereas it becomes *DGP 2* if $\beta = 0$, i.e.

$$y_t = y_{t-1} + \varepsilon_t,\ t \geq 1, \tag{13}$$

In order to test H_0, two regression models are considered, one including a constant term and another including a constant term and a linear time trend

$$RM\ 3 : \Delta y_t = \alpha_1 + \phi \Delta^{\ d} y_{t-1} + \varepsilon_t, \tag{14}$$

or

$$RM\ 4 : \Delta y_t = \alpha_1 + \alpha_2 t + \phi \Delta^{\ d} y_{t-1} + \varepsilon_t, \tag{15}$$

where y_t is a driftless random walk when $\alpha_1 = \alpha_2 = \phi = 0$. Conversely, under the alternative hypothesis that $\phi < 0$, when $\delta = 0$ (i.e., $k = 0$), (11) implies that *RM 3* can be expressed as

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

where $C(L)$ is defined above. Since we are considering truncated processes, this implies that:

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

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 $\beta^* = \beta C(1)$.

Likewise, if $\delta \neq 0$, then the corresponding reformulation of *RM 4* becomes

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

implying that

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

[5] In this fashion we exclude the possibility that the deterministic component of y_t under H_0:$\phi = 0$ may be a quadratic trend. Note that the constant in this case is denoted by β, so that, under the null, the DGP corresponds to (6).

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 this, notice that $\sum_{i=0}^{t-1} ic_i$ is the sum of the first $(t-1)$ terms of $C'(1)$ where

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

When evaluated at $z = 1$, 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, defining 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:

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

where $\beta^* = \beta C(1)$, $\delta^* = \delta C(1)$. This implies that the process y_t is (asymptotically) a $FI(d)$ process with a constant and a linear time trend where $\mathrm{E}(\Delta^d y_t) = \beta^* + \delta^* t$.

In the following subsections we study the behavior of the test statistics associated to the above mentioned regression models under the null hypothesis of a random walk with $(\beta = 0)$ and without $(\beta \neq 0)$ drift.

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

In this subsection, *DGP 2* is considered whereas *RM 3* and *RM 4* are estimated. The next theorems state the consistency of the estimators of ϕ in *RM 3* and *RM 4* and present the asymptotic distributions of these estimators and their corresponding $t-ratios$. As in the previous section, a different asymptotic behavior is found to hold depending on what pre-fixed value of d is used to implement the test.

Theorem 3 *Under the null hypothesis that y_t is generated according to* DGP 2*, the OLS coefficient associated to ϕ in* RM 3*, $\hat{\phi}^{\tau}_{ols}$, (to ϕ in RM 4, $\hat{\phi}^{\mu}_{ols}$, when $\delta = 0$) 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* $0.5 < d < 1$. *The asymptotic distributions of the associated* $t-ratios$, $t_{\hat{\phi}^{\tau}_{ols}}$ *and* $t_{\hat{\phi}}$, *are given by*

$$t_{\hat{\phi}^i_{ols}} \xrightarrow{w} \frac{\int_0^1 W^i_{-d}(r)\, dB(r)}{\left(\int_0^1 \left(W^i_{-d}(r)\right)^2 dr\right)^{1/2}} \qquad if\, 0 \le d < 0.5,\ i = \{\mu, \tau\}$$

and,

$$t_{\hat{\phi}^i_{ols}} \xrightarrow{w} N(0,1) \qquad if\ 0.5 < d < 1,\ for\ i = \{\mu, \tau\}.$$

where $W^{\mu}_{-d}(r)$ *is a demeaned fBM defined as* $W^{\mu}_{-d}(s) = W_{-d}(s) - \int_0^1 W_{-d}(r)\, dr$ *and* $W^{\tau}_{-d}(s)$ *is a detrended fractional Brownian motion defined as* $W^{\tau}_{-d}(s) = W_{-d}(s) + (6s-4)\int_0^1 W_{-d}(r)\, dr - (12s-6)\int_0^1 rW_{-d}(r)\, dr$.

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

We now turn to the interesting case considered by West (1988) in the $I(1)$ *vs.* $I(0)$ framework where both the DGP and the model share the same deterministic component. For simplicity, 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 y_t is assumed to be generated by *DGP 1* whilst the regression model considered is *RM 2.*

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

$$\Delta^d y_t = \Delta^{d-1}\beta + \Delta^{d-1}\varepsilon_t. \tag{16}$$

The first component in the RHS of (16) is the nonlinear trend $\Delta^{d-1}\beta = \beta\tau_t(d-1) = \sum_{i=0}^{t-1}\pi_i(d-1)$. As argued in Section 2, when $d = 0$, (the DF case), $\pi_i(d-1) = 1$ for all i, which implies $\tau_t(-1) = t$. By contrast, when $d = 1$, $\pi_0(d-1) = 1$ and the remaining $\pi_i(d-1) = 0$ for all $i > 0$, which implies that $\tau_t(0) = 1$. 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 $\pi_i(d-1) = \Gamma(i+1-d)/[\Gamma(1-d)\Gamma(i+1)]$ $\sim i^{-d}/\ \Gamma(1-d)$. 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. Since for any value of $d \in [0,1)$, the term $\beta\Delta^{\ d-1}$ 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 $d \in [0.5,1)$.

The results in the next theorem 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 $\hat{\phi}^{\mu}_{ols}$ and $t_{\hat{\phi}^{\mu}_{ols}}$ are normal for all values of d.

Theorem 4 *Under the null hypothesis that y_t is generated according to* DGP 1, *$\hat{\phi}^{\mu}_{ols}$ 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*

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

where

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

The asymptotic distribution of the $t-ratio$ of $\hat{\phi}^{\mu}_{ols}$ in RM 3 *is given by*

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

Tables 2a, b report the empirical sizes for two alternative values of β, namely $\beta = 0.5$ and 5.0 for the three sample sizes considered above. Distinguishing between low and 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 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 $O_p(\beta^2 T^3)$ and $O_p(\sigma_{\varepsilon}^2 T^2)$, respectively, where the leading coefficients β^2 and σ_{ε}^2 have been included in the $O_p(.)$ terms for analytical convenience. In effect, note that if the squared drift, β^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 behavior of the distribution of t_{ϕ} in such a way that it will resemble the distribution 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 $O_p(\beta^2 T^{3-2d})$ 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 1 in the Appendix). As it can be observed below , a low value of β distorts somewhat the 5% size for $T = 100, 400$, yet the distortions are not that large, particularly for $d \geq 0.5$. By contrast, when $\beta = 5$, the empirical sizes match almost perfectly the nominal 5%. An intuitive explanation of why the size distortions in Tables 3a, 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 β 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 $\beta > T^{-1/2}$ for $d \in (0, 0.5)$, $\beta > (\log T)^{1/2} T^{-1}$ for $d = 0.5$, and when $\beta > T^{-(1-d)}$ for $d \in (0.5, 1)$. Thus, 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 $\beta > T^{-1/2}$, implying that small values of β will generate stronger size distortions in finite samples.

TABLE 2a.

EMPIRICAL SIZE 5%; β=0.5

DGP: $\Delta y_t = \beta + \varepsilon_t$; **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 2b.

EMPIRICAL SIZE 5%;β=5

DGP: $\Delta y_t = \beta + \varepsilon_t$; **Regression:** $\Delta y_t = \beta + \phi \Delta^d y_{t-1} + a_t$

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 $\mathrm{T}^{\frac{3}{2}-d}(\widehat{\phi}_{ols} - 1)$ 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 drift.

Finally, we now examine the case where a linear trend is included in the maintained hypothesis, as in *RM 4*. 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 in (5), this will not be the case once $FI(d)$ processes are allowed for. Our setup is one where y_t is generated according to *DGP 1* and the regression model is *RM 4*.

To explain the different behavior of the FDF test vis-á-vis the DF test in this case, it is convenient to recall that the introduction of a trend in the DF regression achieves invariance of the DF test because the regressors t and y_{t-1} become colinear in large samples isolating in this manner the purely stochastic component of y_{t-1}. Nonetheless, 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 $O_p(\beta^2 T^{3-2d})$ while that of the linear trend is $O(\sigma_\varepsilon^2 T^3)$. Hence, like in Case II, the stochastic component of $\Delta^d y_{t-1}$ is dominated by the smooth time trend represented by $\Delta^{d-1}\beta$ leading to asymptotic normality for all values of $d \in (0,1)$. Also note that, in common with the discussion of Case II, the size of the drift under the null matters again since the leading term is scaled by β. Thus the test may suffer from lack of invariance with respect to the size of the drift relative to the variance of the error term in finite samples, as was the case in the previous section. Of course, the solution to recover invariance would be to replace *RM 4* by *RM 1* and use the corresponding critical values.

The theorem derives 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 5 *Under the null hypothesis that y_t is generated according to* DGP 1, $\hat{\phi}_{ols}^{\tau}$ *computed in* RM 4 *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 $0 < d < 1$. The asymptotic distribution of the $t-$ratio of $\hat{\phi}_{ols}^{\tau}$ computed in* RM 4 *is given by*

1. *If $d = 0$, the DF distribution of t_ϕ in* RM 4.
2. *If $0 < d < 1$*

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

As in Case II, Monte-Carlo simulations, not reported this time for the sake of brevity, show that, for $\beta = 5.0$, the empirical sizes almost mimic the nominal size. Conversely, for $\beta = 0.5$, size distortions are serious for $d \leq 0.5$.

Deterministic Components in the FDF test with estimated d.

As with the invariant tests described in Section 2, when d is estimated using a (trimmed) $T^{1/2}$-consistent estimator, the asymptotic distribution of t_ϕ is $N(0,1)$ in all of the three cases analyzed above. This result turns out to be very convenient since it allows to use standard critical values in all possible setups. The following theorem sums up the results obtained in this case.

Theorem 6 *Under the null hypothesis that y_t is a random walk without or with a drift, the $t-ratio$ associated to the $\widehat{\phi}_{ols}$ in* RM 3 *and* RM 4 *where $\hat{d}$ is a $T^{1/2}$-consistent (possibly trimmed) estimator of d, is asymptotically distributed as:*

$$t^i_{\hat{\phi}_{ols}}\left(\hat{d}\right) \xrightarrow{w} N(0,1)\,,\ \textit{for}\ i=\{\mu,\tau\}.$$

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

In view of the above results, a natural strategy arises to test for the null of $I(1)$ vs. $FI(d)$ in the presence of deterministic components when 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. $H_1: 0\leq d<1$ will take as starting point *RM 1* in (5) or in the two-step procedure using residuals, when it is assumed that $\mathrm{E}(\Delta^d y_t)=\Delta^d\mu(t)$, or *RM 4* in (15) when $\mathrm{E}(\Delta^d y_t)=\mu(t)$, 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 stops. If the null is not rejected, then we can use critical values from the $N(0,1)$ distribution to test whether the coefficient β is significant. If it is significant, we stop. Otherwise, we estimate *RM 2* or *RM 3* with $\beta=0$ 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 several 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 behavior 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 this 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 invariant FDF test developed in Section 2 to the logged GDP p.c. of a subset of ten of the main OECD countries which are listed in Table 3, where the estimated intercept and its standard deviation in the regression $\Delta y_t = \beta + u_t$ is reported. As can be observed the mean (average GDP p.c. growth rate) is always highly significant making it convenient to use *RM 1* or *RM 4* as the maintained hypothesis. Indeed, when the ADF and the Phillips-Perron (P-P) unit root tests (not reported) 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 3

ESTIMATES OF $\widehat{\beta}$ AND SD($\widehat{\beta}$)

Country	Mean	St. D.
Australia	0.012	0.004
Canada	0.0195	0.005
Denmark	0.018	0.008
France	0.018	0.006
Germany	0.018	0.007
Italy	0.019	0.006
Netherlands	0.015	0.006
UK	0.013	0.003
USA	0.017	0.005
Spain	0.019	0.005

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 in the preferred models, $\widehat{d}_{ML}$, presented in the first two columns of Table 4, add unity to the obtained estimates. Estimates were also obtained using Mayoral ´s (2003) MD estimation approach, with the series in levels, yielding the pre-estimates of d, $\widehat{d}_{ML}$, in the preferred models presented in the last two columns of Table 4. Both sets of estimates tend to provide similar results. In general, the estimated values of d belong to $(0.5, 1)$. 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 initially tested in *RM 4* 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. The first four columns of Table 5 show strong rejections of H_0: $d = 1$ in most cases. Likewise, for robustness, the last column reports the results of the FDF test in *RM 1* with estimated d, using the $\widehat{d}_{MD}$ estimates in Table 4 and a trimming value of $c = 0.05$ for Australia whose estimated d exceeds unity. Again, with the exception of Spain, we find strong rejections of the null. Thus, our results seem to favor nonstationary, albeit mean-reverting, values of d.[6]

TABLE 4

ESTIMATES OF d (ML and MD)

Country	$\widehat{d}_{ML}$	model	$\widehat{d}_{MD}$	model
Australia	1.003	$(0,d,0)$	1.03	$(0,d,0)$
Canada	0.50	$(1,d,0)$	0.44	$(1,d,0)$
Denmark	0.71	$(1,d,0)$	0.72	$(1,d,0)$
France	0.77	$(0,d,1)$	0.82	$(0,d,1)$
Germany	0.81	$(0,d,1)$	0.80	$(0,d,1)$
Italy	0.82	$(0,d,1)$	0.81	$(0,d,1)$
Netherlands	0.77	$(0,d,1)$	0.77	$(0,d,1)$
UK	0.60	$(1,d,0)$	0.71	$(1,d,0)$
USA	0.78	$(0,d,0)$	0.73	$(1,d,0)$
Spain	0.83	$(1,d,0)$	0.92	$(0,d,0)$

TABLE 5

AFDF TEST AGAINST FI(d)

Country | d	0.9	0.8	0.7	0.6	$\widehat{d}_{MD}$
Australia	-2.27*	-2.41*	-2.55	-2.67*	-2.54*
Canada	-2.78*	-2.87*	-2.95*	-3.05*	-4.21*
Denmark	-2.84*	-2.99*	-3.09*	-5.83*	-3.16*
France	-2.26*	-2.32*	-2.38*	-2.47*	-2.42*
Germany	-2.63*	-2.73*	-2.81*	-3.87*	-2.77*
Italy	-2.04*	-2.06*	-2.03	-2.05	-2.11*
Netherlands	-2.41*	-2.52*	-2.56*	-2.62*	-2.54*
UK	-2.31*	-2.34*	-2.36*	-2.36*	-2.41*
USA	-3.12*	-3.29*	-3.39*	-3.53*	-3.42*
Spain	-0.24	-0.39	-0.66	-0.79	-0.34

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

CONCLUSIONS

This paper has developed statistics for detecting the presence of a unit root in time-series data against the alternative of mean-reverting fractional processes allowing for deterministic terms, $\mu(t)$, (a constant /a linear time trend) in the DGP and in the auxiliary regression

[6] Use of the testing strategy described in Section 3 yields similar results.

used to implement the FDF test. Two setups have been considered. *First*, if it assumed under the alternative that $y_t = \mu(t) + FI(d)$, so that $E(\Delta^d y_t) = \Delta^d \mu(t)$ then including nonlinear trends of the form $\Delta^d \mu(t)$ in the regression model yields invariant test to the parameters defining $\mu(t)$. Alternatively, one can use the simple two-step procedure based on the residuals of the regression of y_t on a constant or a linear trend described above. This test has a non standard asymptotic distribution when d is (arbitrarily) pre-fixed in the range $(0, 0.5)$. However, asymptotic normality holds when $d \in (0.5, 1)$. *Second*, when assuming that $E(\Delta^d y_t) = \mu(t)$, so that a linear trend (not a nonlinear trend) is included in the model to capture a non-zero drift under the null, the FDF test is asymptotically normal for all $d \in (0, 1)$, in contrast to the nonstandard distribution achieved by the traditional DF test when $d = 0$. 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, and most importantly, if d is pre-estimated using a (trimmed) $T^{1/2}$-consistent estimator, as the ones discussed in DGM (2002), 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 $0 \leq d < 1$ based on starting from *RM 1* or *RM 4* 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

Akonom, J. and C. Gourieroux (1987): "A funtional central limit theorem for fractional processes", preprint, CEPREMAP.

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press.

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.

Breitung, J. and U. Hassler (2002) , "Inference on the Cointegration Rank in Fractionally Integrated Processes", *Journal of Econometrics,* 110, 167-185.

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*, 70, 1963-2006.

Dolado, J. and F. Marmol (2003): "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.

Lobato, I. and C. Velasco (2003): " Optimal Fractional Dickey-Fuller Tests for Unit Roots", Universidad Carlos III (mimeo).

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. (2003): "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. M. (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.

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

White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press.

APPENDIX 1

In order to prove Theorem 1, the following lema would be needed.

Lemma 1 *Let y_t be a random walk process defined as in* (7) *. Under the assumptions of Section 2, the following convergences follow:*

If $0 < d < 1$, *then*

1. $T^{-(1-d)} \sum_{i=2}^{T} \tau_{t-1}(d) \to \frac{1}{(d-1)\Gamma(-d)}$.

2. $T^{-(2-d)} \sum_{i=2}^{T} \tau_{t-1}(d-1) \to \frac{1}{\Gamma(3-d)}$.

3. $T^{-(1-2d)} \sum_{i=2}^{T} \tau_{t-1}^{2}(d) \to C_1(d) < \infty$, *for* $d \in (0, 0.5)$ *and*

   $\sum_{i=2}^{T} \tau_{t-1}^{2}(d) \to C_2(d) < \infty$, *for* $d \in (0.5, 1)$.

4. $T^{-(3-2d)} \sum_{i=2}^{T} \tau_{t-1}^{2}(d-1) \to \frac{1}{(3-2d)\Gamma^2(2-d)}$.

5. $T^{-(2-2d)} \sum_{i=2}^{T} \tau_{t-1}(d) \tau_{t-1}(d-1) \to C_3(d)$.

6. $\sum_{i=2}^{T} \tau_{t-1}(d) \varepsilon_t \xrightarrow{d} N\left(0, \sigma^2 C_1\right)$ *for* $d \in (0, 0.5)$.

   $T^{-(1/2-d)} \sum_{i=2}^{T} \tau_{t-1}(d) \varepsilon_t \xrightarrow{w} N\left(0, \sigma^2 C_2\right)$ *for* $d \in (0.5, 1)$.

7. $T^{-(3/2-d)} \sum_{i=2}^{T} \tau_{t-1}(d-1) \varepsilon_t \xrightarrow{w} N\left(0, \sigma^2 C_4\right)$, $C_4 = \frac{1}{(3-2d)\Gamma^2(2-d)}$.

8. $T^{-(1-d)} \sum_{i=2}^{T} \Delta^d y_{t-1} \varepsilon_t \xrightarrow{w} \sigma^2 \int_0^1 W_{-d}(r)\, dB(r)$ *for* $d \in (0, 0.5)$ *and*

   $T^{-0.5} \sum_{i=2}^{T} \Delta^d y_{t-1} \varepsilon_t \xrightarrow{w} \sigma^2 N\left(0, \frac{\Gamma(2d-1)}{\Gamma(d)}\right)$ *for* $d \in (0.5, 1)$.

9. $T^{-(3/2-2d)} \sum_{i=2}^{T} \tau_{t-1}(d) \Delta^d y_{t-1} \xrightarrow{w} \frac{1}{\Gamma(-d)(d-1)} \int_0^1 r^{-d} W_{-d}(d)\, dr$ *for* $d \in (0, 0.5)$, *and*

   $T^{-(1-d)} \sum_{i=2}^{T} \tau_{t-1}(d) \Delta^d y_{t-1} \xrightarrow{w} 0$ *for* $d \in (0.5, 1)$.

10. $T^{-(2-d)} \sum_{i=2}^{T} \tau_{t-1}(d-1) \Delta^d y_{t-1} \xrightarrow{p} 0$ *for* $d \in (0.5, 1)$

    $T^{-(5/2-2d)} \sum_{i=2}^{T} \tau_{t-1}(d-1) \Delta^d y_{t-1} \xrightarrow{w} \sigma^2 \int_0^1 r^{1-d} W_{-d}(r)\, dr$.

11. $T^{-1} \sum (\Delta^d y_{t-1})^2 \xrightarrow{p} Var(y)$ *if* $d \in (0.5, 1)$ *and*

    $T^{-2(1-d)} \sum (\Delta^d y_{t-1})^2 \xrightarrow{w} \sigma^2 \int_0^1 W_{-d}^2(r)\, dr$ *if* $d \in [0, 0.5)$,

12. $T^{-1} \sum_{i=2}^{T} \Delta^d y_{t-1} \xrightarrow{p} 0$ *if* $d \in (0.5, 1)$ *and*

    $T^{-(3/2-d)} \sum_{i=2}^{T} \Delta^d y_{t-1} \xrightarrow{w} \int_0^1 W_{-d}(r)\, dr$ *if* $d \in (0, 0.5)$.

Proof of Lemma 1

1. Notice that $\sum_{i=2}^{T}\tau_{t-1}(d)$ can alternatively be written as

$$\lim_{T\to\infty}\sum_{i=2}^{T}\tau_{t-1}(d)=\lim_{T\to\infty}T\pi_0(d)+(T-1)\pi_1(d)+... \tag{17}$$

and also note that $\sum_{i=0}^{\infty}\pi_i(d)=0$, then,

$$\lim_{T\to\infty}T^{-(1-d)}\sum_{i=2}^{T}\tau_{t-1}(d)=\lim_{T\to\infty}T^{-(1-d)}\sum_{i=0}^{T}(T-i))\pi_i$$

$$=T^{d}\lim_{T\to\infty}\sum_{i=0}^{T}\pi_i(d)-T^{-(1-d)}\lim_{T\to\infty}\sum_{i=1}^{T}i\pi_i(d) \tag{18}$$

$$=\frac{1}{\Gamma(-d)}\lim_{T\to\infty}T^{-(1-d)}\sum_{i=1}^{T}i^{-d}=\frac{-1}{\Gamma(-d)(1-d)}, \tag{19}$$

where the last equality follows from applying L´hopital´s rule to the first term of (18) and noticing that it tends to zero.

2. In this case,

$$\lim_{T\to\infty}T^{-(2-d)}\sum_{i=2}^{T}\tau_{t-1}(d-1)=\frac{1}{\Gamma(1-d)}\lim_{T\to\infty}\sum_{t=1}^{T}\sum_{i=1}^{t}i^{-d}=\frac{1}{\Gamma(3-d)}.$$

3. Since $\sum_{i=0}^{\infty}\pi_i(d)=0$, it is possible to write $\pi_0=-\sum_{i=1}^{\infty}\pi_i(d)$, $\pi_0+\pi_1=-\sum_{i=2}^{\infty}\pi_i(d)$, etc. Then,

$$\sum_{i=2}^{T}\tau_{t-1}^{2}(d)=-\sum_{j=1}^{T}\left(\sum_{i=j}^{\infty}\pi_i(d)\right)^2. \tag{20}$$

Since the coefficients $\{\pi_i(d)\}_{i=0}^{\infty}$ are such that $\pi_i\sim i^{-1-d}$, then $\left(\sum_{i=j}^{\infty}\pi_i(d)\right)^2=O\left(j^{-2d}\right)$ (see, Davidson (1994, $p.$ 32)). This implies that if $d\in(0.5,1)$, the quantity in (20) is summable and if $d\in(0,0.5)$ it is $O\left(T^{1-2d}\right)$.

4. The proof of this result is similar to the previous ones and therefore is omitted.

5. Idem.

6. These limits are a direct application of Corollary 5.25 p.130 in White (1984).

7. Idem

8. See DGM (2002) for the proofs of these results

9. The first result follows from point 1 in this lemma and the results in Dolado and Marmol (2003). The second follows from noting that $T^{-(1-d)}E\left(\sum_{i=2}^{T}\tau_{t-1}(d)\Delta^{d}y_{t-1})\right)=0$ and that $T^{-2(1-d)}E\left(\left(\sum_{i=2}^{T}\tau_{t-1}(d)\Delta^{d}y_{t-1})\right)^2\right)\to 0$.

10. Idem.

11. See DGM for the proof of this result.

12. Idem.■

Proof of Theorem 1

For simplicity, let us consider first *RM 2* defined in equation (8). Since the nature of the asymptotic distribution depends upon the value of d used to run the regression, two cases ought to be distinguished.

I. First case: $0 \leq d < 0.5$. Define the scaling matrix

$$\Upsilon_T = \begin{pmatrix} T^{1/2-d} & 0 \\ 0 & T^{1-d} \end{pmatrix}, \tag{21}$$

and taking into account the results in Lemma 1 we easily get

$$\begin{pmatrix} T^{1/2} & 0 \\ 0 & T^{1-d} \end{pmatrix}\begin{pmatrix} \hat{\alpha} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} C_1 & \frac{1}{\Gamma(-d)(1-d)}\sigma \int_0^1 r^{-d} W_{-d}(d)\, dr \\ \frac{1}{\Gamma(-d)(1-d)}\sigma \int_0^1 r^{-d} W_{-d}(d)\, dr & \sigma^2 \int_0^1 W_{-d}^2(r)\, dr \end{pmatrix}^{-1} \begin{pmatrix} \sigma N(0, C_1) \\ \sigma^2 \int_0^1 W_{-d}(r)\, dB(r) \end{pmatrix},$$

which implies that

$$t_{\hat{\phi}}^{\mu} \xrightarrow{d} \frac{C_1^{1/2}\left(\Gamma(-d)(1-d)\right)^{-1}\left[\int_0^1 W_{-d}(r)\, dB(r) - B(1)\int_0^1 r^{-d} W_{-d}(r)\, dr\right]}{\left[C_1 \int_0^1 W_{-d}^2(r)\, dr - \left(\frac{1}{\Gamma(-d)(1-d)}\int_0^1 r^{-d} W_{-d}(d)\, dr\right)^2\right]^{1/2}}$$

which is a functional of fractional brownian motions and other terms just depending on d.

II. Second case: $0.5 \leq d < 1$. Defining the scaling matrix

$$\Upsilon_T = \begin{pmatrix} 1 & 0 \\ 0 & T^{1-d} \end{pmatrix}$$

and taking into account the results of Lemma 1 is straightforward to check that $t_{\hat{\phi}}^{\mu} \sim N(0,1)$.

Consider now RM 1 as defined in (5). To see that the parameter ϕ is numerically invariant to any linear transformation in y_t, note that the regression (5) can be equivalently written as

$$\begin{aligned} \Delta y_t &= \phi \Delta^d (y_{t-1} - y_0 + \beta(t-1)) + \alpha_0 + (\alpha_1 + \phi y_0)\tau_{t-1}(d) + (\alpha_2 + \phi\beta)\tau_{t-1}(d-1) + \varepsilon_t \quad (22) \\ &= \phi^* \Delta^d \xi_{t-1} + \alpha_0^* + \alpha_1^* \tau_{t-1}(d) + \alpha_2^* \tau_{t-1}(d-1). \quad (23) \end{aligned}$$

where $\phi^* = \phi, \alpha_0^* = \alpha_0, \alpha_1^* = (\alpha_1 + \phi y_0)$ and $\alpha_2^* = (\alpha_2 + \phi\beta)$ and under the null hypothesis, ξ_t is a random walk without drift and with initial condition equal to zero. Following

Hamilton (1994, p.498), it is straightforward to see that the OLS estimator of ϕ and its associated t-statistic are numerically identical to the one that would be obtained if the original process was ξ_t in stead of y_t. Taking into account this invariance property, it is possible to consider without loss of generality that $y_0 = \beta = 0$. Then, the rest of the proof is similar to the previous one and, therefore, is omitted.

Proof of Theorem 2

When $\widehat{d}$ is chosen such that $\widehat{d} = \widehat{d}_T$ if $\widehat{d}_T < 1 - c$ and $\widehat{d} = 1 - c$ if $\widehat{d}_T \geq 1 - c$, where c is a (fixed) value in the neighborhood of zero, it is clear that $\hat{d} \xrightarrow{p} 1 - c$, since $\hat{d}_T$ is a consistent estimator of $d\ (= 1)$. Applying the mean value theorem (MVT) on $t^{\mu}_{\phi_{ols}}$ around the point $(1 - c)$ yields

$$t^{\mu}_{\phi_{ols}}\left(\hat{d}\right) = t^{\mu}_{\phi_{ols}}(1 - c) + \frac{\partial t^{\mu}_{\phi_{ols}}\left(\breve{d}\right)}{\partial d}\left(\hat{d} - (1 - c)\right), \tag{24}$$

where $\breve{d}$ is an intermediate point between $\hat{d}$ and $(1 - c)$. This implies that in order to prove that $\left(t^{\mu}_{\phi_{ols}}\left(\hat{d}\right) - t^{\mu}_{\phi_{ols}}(1 - c)\right) = o_p(1)$ it has to be shown that $\frac{\partial t^{\mu}_{\phi_{ols}}(\breve{d})}{\partial d}\left(\hat{d} - (1 - c)\right) = o_p(1)$. Notice that $\breve{d} \in \left(\hat{d}, 1 - c\right)$ and therefore, $\breve{d} \xrightarrow{p} (1 - c)$. In order to replace $\breve{d}$ in (24) by its probability limit, $1 - c$, it is needed to show that $\partial t^{\mu}_{\phi_{ols}}(d) / \partial d$ converges uniformly to a non-stochastic function in an open neighborhood of $(1 - c)$ (see Amemiya, 1985). Using the same strategy as in DGM (2002), it can be shown that $T^{-1/2}\partial t^{\mu}_{\phi_{ols}}(d) / \partial d$ converges pointwise to zero. The uniform convergence follows from the pointwise convergence and an equicontinuity argument deriving from the differenciability of $\partial t^{\mu}_{\phi_{ols}}(d) / \partial d$ with respect to d (cf. Davidson, (1994), p. 340, and Velasco and Robinson, 2000). The result follows just by noticing that $T^{1/2}\left(\hat{d} - (1 - c)\right)$ is $O_p(1)$ and therefore $\partial t^{\mu}_{\phi_{ols}}(d) / \partial d\left(\hat{d} - (1 - c)\right) = o_p(1)$.

The proof for the case where a deterministic trend is included can be constructed along the same lines.■

Proof of Theorem 3

Consider first *RM 3* in (14). 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 $FI(1 - d)$.

Define the scaling matrix:

$$\Upsilon_T = \begin{pmatrix} T^{1/2} & 0 \\ 0 & T^{1-d} \end{pmatrix}. \tag{25}$$

Then,

$$\Upsilon_T \begin{pmatrix} \hat{\alpha} \\ \hat{\phi}_{ols} \end{pmatrix} \xrightarrow{w} \begin{pmatrix} 1 & \sigma\int_0^1 W_{-d}(r)\,dr \\ \sigma\int_0^1 W_{-d}(r)\,dr & \sigma^2\int_0^1 W^2_{-d}(r)\,dr \end{pmatrix}^{-1} \begin{pmatrix} \sigma B(1) \\ \sigma^2\int_0^1 W_{-d}(r)\,dB(r) \end{pmatrix}.$$

and consequently,

$$T^{1-d}\hat{\phi}_{ols} \xrightarrow{w} \frac{\int_0^1 W_{-d}(r)\,dB(r) - B(1)\int_0^1 W_{-d}(r)\,dr}{\int_0^1 W^2_{-d}(r)\,dr - \left(\int_0^1 W_{-d}(r)\,dr\right)^2}. \tag{26}$$

Defining $W^{\mu}_{-d}(s) = W_{-d}(s) - \int_0^1 W_{-d}(r)\,dr$, it is straightforward to check that

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

Consider now the OLS $t-test$ of the null hypothesis that $\phi = 0$, $t^{\mu}_{\hat{\phi}_{ols}} = \hat{\phi}_{ols}/\hat{\sigma}_{\hat{\phi}_{ols}}$. Straigth forward calculations show that,

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

where

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

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

$$s^2_T \xrightarrow{p} \sigma^2.$$

Therefore

$$T^{2(1-\ d)}\hat{\sigma}^2_{\hat{\phi}_{ols}} \xrightarrow{w} \frac{1}{\int (W_{-\ d}(r))^2\,dr - \left(\int W_{-\ d}(r)\,dr\right)^2}, \tag{27}$$

and this implies that:

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

or equivalently

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

II. Second case: $d = 0.5$

Define the weighting matrix:

$$\Upsilon_T = \begin{pmatrix} T^{1/2} & 0 \\ 0 & (T\log T)^{1/2} \end{pmatrix}. \tag{28}$$

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

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

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

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

which implies the desired result.

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,

$$T^{1/2} \begin{pmatrix} \hat{\alpha} \\ \hat{\phi}_{ols} \end{pmatrix} \xrightarrow{w} N_2 \left(0, \begin{pmatrix} \sigma & 0 \\ 0 & \frac{\Gamma^2(d)}{\Gamma(2d-1)} \end{pmatrix}\right). \tag{29}$$

Now, taking as scaling matrix $\Upsilon_T = T^{1/2} I_2$, the distribution of the $t-statistic$ follows trivially.

Next, consider RM 4. Again, we need to differenciate different cases according to the value of d. The OLS-statistic is given as usual by,

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\delta} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} T & \sum t & \sum \Delta^{d} y_{t-1} \\ \sum t & \sum t^2 & \sum t \Delta^{d} y_{t-1} \\ \sum \Delta^{d} y_{t-1} & \sum t \Delta^{d} y_{t-1} & \sum (\Delta^{d} y_{t-1})^2 \end{pmatrix}^{-1} \times \begin{pmatrix} \sum \varepsilon_t \\ \sum t \varepsilon_t \\ \sum \Delta^{d} y_{t-1} \varepsilon_t \end{pmatrix}. \tag{30}$$

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, 2003):

$$T^{-(5/2-d)} \sum t \Delta^{d} y_{t-1} \xrightarrow{w} \sigma \int_0^1 r W_{-d}(r)\, dr \tag{31}$$

Define the scaling matrix:

$$\Upsilon_T = \begin{pmatrix} T^{1/2} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & T^{1-d} \end{pmatrix}. \tag{32}$$

then,

$$\Upsilon_T \begin{pmatrix} \hat{\alpha} \\ \hat{\delta} \\ \hat{\phi} \end{pmatrix} \xrightarrow{w} \begin{pmatrix} 1 & \frac{1}{2} & \sigma \int_0^1 {}_{-\ d}W\,(r)\,dr \\ \frac{1}{2} & \frac{1}{3} & \sigma \int_0^1 rW_{-\ d}\,(r)\,dr \\ \sigma \int_0^1 W_{-\ d}\,(r)\,dr & \sigma \int_0^1 rW_{-\ d}\,(r)\,dr & \sigma^2 \int_0^1 W_{-d}^2\,(r)\,dr \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \sigma B(1) \\ \sigma B(1) - \sigma \int_0^1 B\,(r)\,dr \\ \sigma^2 \int_0^1 W_{-\ d}\,(r)\,dB(r) \end{pmatrix}. \tag{33}$$

Defining

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

it is straight forward to show that

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

II. Second case: $d = 0.5$.

In this case, the scaling matrix is

$$\Upsilon_T = \begin{pmatrix} T^{1/2} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & (T\log T)^{1/2} \end{pmatrix}. \tag{34}$$

Since

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

$$\xrightarrow{p} \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{\sigma^2}{\pi} \end{pmatrix}^{-1}, \tag{35}$$

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

$$\Upsilon_T \begin{pmatrix} \hat{\alpha} \\ \hat{\delta} \\ \hat{\phi} \end{pmatrix} \xrightarrow{w} N\left(0, \sigma^2 \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{\sigma^2}{\pi} \end{pmatrix}^{-1}\right) = N\left(0, \begin{pmatrix} 4\sigma^2 & -6\sigma^2 & 0 \\ 0 & 12\sigma^2 & 0 \\ -6\sigma^2 & 0 & \pi \end{pmatrix}\right). \tag{36}$$

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

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

$$\Upsilon_T = \begin{pmatrix} T^{1/2} & 0 & 0 \\ 0 & T^{1/2} & 0 \\ 0 & 0 & T^{3/2} \end{pmatrix}. \tag{37}$$

Taking into account that

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

and that

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

then

$$\Upsilon_T \begin{pmatrix} \hat{\alpha} \\ \hat{\delta} \\ \hat{\phi} \end{pmatrix} \xrightarrow{w} N\left(0, \sigma^2 \begin{pmatrix} 4\sigma^2 & -6\sigma^2 & 0 \\ -6\sigma^2 & 12\sigma^2 & 0 \\ 0 & 0 & \frac{\Gamma^2(d)}{\Gamma(2d-1)} \end{pmatrix}\right). \blacksquare \tag{38}$$

For the sake of brevity, the proofs of the distributions of the $t-statistics$ are not reported although they can be easily obtained following the steps of the first part of this proof.$\blacksquare$

To prove Theorem 4, the following lemmas are needed.

Lemma 2 *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 generated according to* DGP 1 *and consider the filtered process*

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

Then:

1. *if* $0 \leq d < 1$

$$T^{-(2-d)}\sum_{t=1}^{T} \Delta^d y_t \xrightarrow{p} \frac{\beta}{\Gamma\left(3-\ d\right)},$$

2. *if* $0 \leq d < 1$

$$T^{-(3-2d)}\sum_{t=1}^{T} \left(\Delta^d y_t\right)^2 \xrightarrow{p} \frac{\beta^2}{\left(\Gamma\left(2-d\right)\right)^2\left(3-2d\right)}.$$

Lemma 3 *Let ε_t and y_t be defined as in Lemma 2. Then the process $\Delta^d y_{t-1}\varepsilon_t$ is a martingale differences and verify that if $0 \leq d < 1$,*

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

Proof of Lemma 2

Consider y_t to be random walk as in (7). The differenced process $\Delta^d y_t$ can then be rewritten as

$$\Delta^d y_t = \Delta^{d-1}\alpha 1_{(t>0)} + \Delta^{d-1}\varepsilon_t 1_{(t>0)} = \alpha\sum_{i=0}^{t-1}\pi_i\left(d-1\right) + \sum_{i=0}^{t-1}\pi_i\left(d-1\right)\varepsilon_{t-i}. \tag{39}$$

1. Expression (39) implies

$$\sum_{t=1}^{T}\Delta^d y_t = \alpha\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_i\left(d-1\right)\right) + \sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_i\left(d-1\right)\varepsilon_{t-i}\right) \tag{40}$$

The first term in the RHS of (40) is $O_p\left(T^{2-d}\right)$ since

$$\begin{aligned} T^{-(2-d)}\lim_{T\to\infty}\alpha\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_i\left(d-1\right)\right) &= T^{-(2-d)}\frac{\alpha}{\Gamma\left(1-\ d\right)\left(1-d\right)}\lim_{T\to\infty}\sum_{t=1}^{T}t^{1-\ d} \\ &= \frac{\alpha}{\Gamma\left(1-d\right)\left(1-d\right)(2-d)} \\ &= \frac{\alpha}{\Gamma\left(3-d\right)}. \end{aligned} \tag{41}$$

The second term in the RHS of (40) 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 $d \in [0, 0.5)$, $d = 0.5$ and $d \in (0.5, 1)$ 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

$$\begin{aligned} \sum_{t=1}^{T}\left(\Delta^{\ d} y_t\right)^2 &= \alpha^2\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_i\left(\ d-1\right)\right)^2 + \sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_i\left(\ d-1\right)\varepsilon_t\right)^2 \\ &\quad + 2\alpha\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_i\left(\ d-1\right)\right)\Delta^{\ d-1}\varepsilon_t. \end{aligned} \tag{42}$$

The first term in the RHS of (42) is $O_p\left(T^{3-2d}\right)$ and converges to

$$T^{-(3-2\ d)}\lim_{T\to\infty}\alpha^2\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_i\left(\ d-1\right)\right)^2 = \frac{\alpha^2}{\Gamma^2\left(2-\ d\right)\left(3-2\ d\right)}. \tag{43}$$

The second term in (42) 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 $d \in [0, 0.5)$, $d = 0.5$ or $d \in (0.5, 1)$ respectively. Finally, the third term in the RHS of (42) is $O_p\left(T^{5/2-d}\right)$ (see Dolado and Marmol, 2003), 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 (43).■

Proof of Lemma 3.

Notice that

$$\sum_{t=2}^{T}\Delta^d y_{t-1}\varepsilon_t=\alpha\sum_{t=2}^{T}\left(\sum_{i=0}^{t-1}\pi_i\left(d-1\right)\right)\varepsilon_t+\sum_{t=2}^{T}\left(\sum_{i=0}^{t-1}\pi_i\left(d-1\right)\varepsilon_{t-i}\right)\varepsilon_t. \tag{44}$$

The first term in the RHS of expression (44) 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 $\{\left(\sum_{i=0}^{t-1}\pi_i\left(d-1\right)\right)\varepsilon_t\}$ 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:

$$\tilde{\varepsilon}_t=\sigma^{-1}\varepsilon_t, \tag{45}$$

$$\tilde{x}_t=\left((T^{(2-d)}\frac{\alpha^2}{\Gamma^2\left(2-d\right)\left(3-2d\right)}\right)^{-1/2}\sum_{i=0}^{t-1}\pi_i\left(1-d\right), \tag{46}$$

and

$$X_{T.t}=T^{-1/2}\tilde{x}_{t-1}\tilde{\varepsilon}_t. \tag{47}$$

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

$$T^{-1/2}E\left(\tilde{x}_{t-1}\tilde{\varepsilon}_t|F_{T.t-1}\right)=T^{-1/2}\tilde{x}_{t-1}E\left(\tilde{\varepsilon}_t|F_{T.t-1}\right)=0. \tag{48}$$

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

$$T^{-1}\sum_{t=2}^{T}Var\left(\tilde{x}_{t-1}\tilde{\varepsilon}_t|F_{T.t-1}\right)=T^{-1}\sum_{t=2}^{T}\tilde{x}_{t-1}^2E\left(\tilde{\varepsilon}_t^2|F_{T.t-1}\right)-\tilde{x}_{t-1}^2E\left(\tilde{\varepsilon}_t|F_{T.t-1}\right)^2$$

$$=T^{-1}\sum\tilde{x}_{t-1}^2\xrightarrow{p}1. \tag{49}$$

Finally, condition (iii) holds since

$$\sum_{t=1}^{T}E\left(|X_{T.t}|^2\,I\left\{|X_{T.t}|>\varrho|\right\}\right)=E\left(|\tilde{x}_{t-1}\tilde{\varepsilon}_t|^2\,I_{\left\{|\tilde{x}_{t-1}\tilde{\varepsilon}_t|>T^{1/2}\varrho|\right\}}\right)\rightarrow0,\text{ for all }\varrho>0. \tag{50}$$

Conditions (48), (49) and (50) 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 $T^{-1}\left(\sum\left(\tilde{x}_{t-1}^*\right)^2-\sum\left(\tilde{x}_{t-1}\right)^2\right)=o_p\left(1\right)$ (see Lemma 1) implies $T^{-1}\sum\left(\tilde{x}_{t-1}^*\right)^2\xrightarrow{p}$

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

$$\frac{1}{T^2}\sum_{t=1}^{T} E\left(\tilde{x}_{t-1}^4 \tilde{\varepsilon}_t^4\right) = \frac{1}{T^2}\mu_4 \sum_{t=1}^{T} \tilde{x}_{t-1}^4 =$$

$$\frac{1}{T^{6-2d}}\mu_4 \left(\frac{\alpha^2}{\Gamma^2(2-d)(3-2d)}\right)^{-2} \sum_{t=1}^{T-2}\left(\sum_{i=0}^{t-1}\pi_i(1-d)\right)^4. \quad (51)$$

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 (51) tends to zero for all $d > -1/2$.

The second term of the RHS of (44) 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 4

Define the scaling matrix,

$$\Upsilon_T = \begin{pmatrix} T^{1/2} & 0 \\ 0 & T^{3/2-d} \end{pmatrix}, \quad (52)$$

and notice that

$$\Upsilon_T \begin{pmatrix} \hat{\alpha}-\alpha \\ \hat{\phi}_{ols} \end{pmatrix} = \begin{pmatrix} 1 & \frac{\alpha}{\Gamma(1-d)(1-d)} \\ \frac{\alpha}{\Gamma(1-d)(1-d)} & \frac{\alpha^2}{\Gamma(1-d)^2(1-d)^2(3-2d)} \end{pmatrix}^{-1} \Upsilon_T^{-1} \begin{pmatrix} \sum_{t=1}^{T}\varepsilon_t \\ \sum_{t=1}^{T}\varepsilon_t \Delta^d y_{t-1} \end{pmatrix} + o_p(1), \quad (53)$$

and

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

where

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

Then

$$\Upsilon_T \begin{pmatrix} \hat{\alpha}-\alpha \\ \hat{\phi}_{ols} \end{pmatrix} \xrightarrow{w} N\left(0, \sigma^2 Q_2^{-1} Q_2 Q_2^{-1}\right) = N\left(0, \sigma^2 Q_2^{-1}\right). \quad (54)$$

The distribution of the $t-statistic$ can be obtained from (54) using the same procedure as in Theorem 3.■

Lemma 4 *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 generated by* DGP 1. *Then, if $0 < d < 1$*

$$\sum_{t=2}^{T} t\Delta^d y_{t-1} = O_p\left(T^{3-d}\right). \quad (55)$$

Proof of Lemma 3

The LHS of (55) can be rewritten as

$$\sum t\Delta^d y_{t-1} = \beta \sum t \left(\sum_{i=0}^{t-1} \pi_i \left(d-1\right) \right) + \sum t x_{t-1}, \tag{56}$$

where x_t is a pure $FI\left(1-d\right)$ with no deterministic components. The first term of the RHS of (56) is completely deterministic and its limit is given by

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

The second term of the RHS of (56) is $O_p\left(T^{5/2-d}\right)$ (see Dolado and Mármol, 2003) and therefore the first term dominates.■

Proof of Theorem 5

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:

$$\Upsilon_T = \begin{pmatrix} T^{1/2} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & T^{3/2-d} \end{pmatrix}, \tag{57}$$

The first term of the OLS estimator, conveniently standardized, converges in probability to

$$\begin{pmatrix} 1 & 1/2 & \frac{\beta}{\Gamma(3-d)} \\ 1/2 & 1/3 & \frac{\beta}{\Gamma(2-d)(3-d)} \\ \frac{\beta}{\Gamma(3-d)} & \frac{\beta}{\Gamma(2-d)(3-d)} & \frac{\beta^2}{\Gamma(2-d)^2(3-2d)} \end{pmatrix}^{-1}, \tag{58}$$

and the second term converges weakly to

$$\begin{pmatrix} T^{-1/2}\sum \varepsilon_t \\ T^{3/2}\sum t\varepsilon_t \\ T^{3/2-d}\sum \Delta^d y_{t-1}\varepsilon_t \end{pmatrix} \xrightarrow{w} N\left(0, \sigma^2 Q_3\right). \tag{59}$$

Expressions (58) and (59) imply the desired result.■

Proof of Theorem 9

This proof can be constructed along the same lines as that of Theorem 2 and therefore it is omitted.■

APPENDIX 2

TABLE A2a

Critical Values

DGP 2: $\Delta y_t = \varepsilon_t$;RM 1:$\Delta y_t = \alpha_1 + \alpha_2\tau_{t-1}(d) + \alpha_3\tau_{t-1}(d-1) + \phi\Delta^{d_1}y_{t-1} + e_t$

T	$T = 100$			$T = 400$			$T = 1000$		
d_1 / sig.lev.	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.0 (DF c.v.)									
0.05	-3.277	-3.583	-4.211	-3.219	-3.524	-4.076	-3.094	-3.488	-4.006
0.10	-3.179	-3.478	-4.059	-3.116	-3.418	-4.021	-3.006	-3.360	-3.877
0.15	-3.036	-3.357	-3.985	-2.993	-3.325	-3.880	-2.931	-3.252	-3.759
0.20	-2.947	-3.157	-3.835	-2.869	-3.124	-3.769	-2.784	-3.101	-3.643
0.25	-2.792	-3.014	-3.765	-2.739	-2.993	-3.674	-2.731	-2.975	-3.548
0.30	-2.670	-2.895	-3.619	-2.597	-2.889	-3.504	-2.481	-2.882	-3.433
0.35	-2.576	-2.716	3.564	-2.468	-2.806	-3.398	-2.303	-2.781	-3.352
0.40	-2.469	-2.695	-3.432	-2.340	-2.653	-3.261	-2.214	-2.600	-3.247
0.45	-2.315	-2.586	-3.320	-2.226	-2.565	-3.229	-2.049	-2.441	-3.148
0.50	-2.202	-2.428	-3.183	-2.086	-2.402	-3.050	-1.974	-2.318	-2.978
0.55	-2.100	-2.282	-3.222	-1.847	-2.370	-3.021	-1.751	-2.279	-2.930
0.60	-2.009	-2.182	-3.001	-1.758	-2.116	-2.881	-1.621	-2.164	-2.994
0.65	-1.807	-2.102	-2.849	-1.666	-2.188	-2.811	-1.563	-1.981	-2.708
0.70	-1.753	-2.015	-2.757	-1.629	-2.056	-2.735	-1.524	-1.971	-2.673
0.75	-1.641	-1.962	-2.644	-1.568	-1.982	-2.630	-1.448	-1.969	-2.617
0.80	-1.563	-1.833	-2.564	-1.492	-1.902	-2.554	-1.376	-1.759	-2.501
0.85	-1.491	-1.750	-2.505	-1.341	-1.760	-2.495	-1.331	-1.758	-2.446
0.90	-1.441	-1.702	-2.437	-1.293	-1.750	-2.428	-1.292	-1.705	2.418
0.95	-1.381	-1.682	-2.388	-1.283	-1.710	-2.372	-1.279	-1.280	-2.331

TABLE A2b

Critical Values

DGP 2: $\Delta y_t = \varepsilon_t$;RM 2:$\Delta y_t = \alpha_1 \tau_{t-1}(d) + \phi \Delta^{d_1} y_{t-1} + e_t$									
T	$T = 100$			$T = 400$			$T = 1000$		
d_1 / sig.lev.	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.0 (DF c.v.)									
0.05	-2.508	-2.808	-3.508	-2.468	-2.751	-3.360	-2.516	-2.826	-3.383
0.10	-2.424	-2.762	-3.424	-2.406	-2.676	-3.276	-2.404	-2.703	-3.296
0.15	-2.311	-2.665	-3.311	-2.318	-2.641	-3.241	-2.334	-2.651	-3.160
0.20	-2.217	-2.542	-3.217	-2.214	-2.511	-3.111	-2.168	-2.497	-3.086
0.25	-2.099	-2.380	-3.099	-2.108	-2.419	-3.033	-2.104	-2.434	-3.055
0.30	-1.994	-2.344	-2.994	-1.951	-2.296	-2.940	-1.980	-2.296	-2.904
0.35	-1.885	-2.242	-2.885	-1.880	-2.190	-2.977	-1.816	-2.158	-2.777
0.40	-1.801	-2.1267	-2.801	-1.734	-2.070	-2.749	-1.677	-2.001	-2.625
0.45	-1.724	-2.082	-2.724	-1.640	-1.999	-2.687	-1.628	-1.974	-2.673
0.50	-1.623	-1.971	-2.643	-1.514	-1.886	-2.569	-1.537	-1.872	-2.575
0.55	-1.540	-1.913	-2.596	-1.486	-1.840	-2.541	-1.430	-1.781	-2.467
0.60	-1.456	-1.821	-2.525	-1.408	-1.743	-2.511	-1.366	-1.769	-2.423
0.65	-1.449	-1.811	-2.483	-1.370	-1.730	-2.448	-1.345	-1.751	-2.469
0.70	-1.422	-1.815	-2.439	-1.347	-1.746	-2.403	-1.314	-1.696	-2.393
0.75	-1.353	-1.793	-2.393	-1.347	-1.699	-2.386	-1.307	-1.676	-2.357
0.80	-1.341	-1.736	-2.371	-1.296	-1.681	-2.351	-1.336	-1.669	-2.342
0.85	-1.310	-1.694	-2.350	-1.290	-1.682	-2.339	-1.335	-1.673	-2.337
0.90	-1.298	-1.664	-2.347	-1.305	-1.651	-2.338	-1.324	-1.649	-2.343
0.95	-1.257	-1.654	-2.337	-1.266	-1.643	-2.406	-1.262	-1.642	-2.3339