A Fractional Dickey-Fuller Test.

Juan J. Dolado[a], Jesús Gonzalo[b] and Laura Mayoral[b*]

a. Dept. of Economics. *b*. Dept. of Statistics and Econometrics.

Universidad Carlos III de Madrid

Abstract

Most of the testing and estimation approaches of fractional unit roots in time series have been derived in the frequency domain. This article proposes Wald type tests in the time domain along the lines of the well-known Dickey-Fuller approach and extends the original framework of $I(1)$ versus $I(0)$ to the more general $FI(d_0)$ versus $FI(d_1)$ with $d_1 < d_0$. The tests are based on the OLS estimator or on the t-statistic associated to $\Delta^{d_1} y_{t-1}$ in a regression of $\Delta^{d_0} y_t$ on $\Delta^{d_1} y_{t-1}$ and, possibly, some lags of $\Delta^{d_0} y_t$. In particular, we focus on the case where $d_0 = 1$ and $0 \leq d_1 < 1$. The asymptotic distributions of the statistics are derived and it is shown that their properties depend on the distance between the null and the alternative hypotheses as well as on the nature of the process under the null hypothesis. When the integration order under the alternative hypothesis, d_1, is unknown a priori, a preestimation of d_1 is needed to implement the test. For this purpose we propose a minimum distance type estimator in the time domain which is easy to implement in this context and has good properties both asymptotically and in finite samples. Monte Carlo simulations support the asymptotic analytical results obtained in this paper and show that the proposed tests have better size and power properties than many of the available tests in the literature.

* This paper is part of the Ph.D. dissertation of Laura Mayoral, Department of Statistics and Econometrics, Universidad Carlos III de Madrid. The authors would like to thank Uwe Hassler, Francesc Mármol and Carlos Velasco for their very useful comments.

1. INTRODUCTION

Since the influential work of Nelson and Plosser (1982) many time series in macroeconomics and finance have been considered to be $I(1)$ processes. However, Kwiatkowski et al. (1992) have pointed out that taking the null hypothesis to be $I(1)$ rather than $I(0)$ can give rise to a bias in favor of the former, which could explain the common failure in rejecting the null of a unit root. Consequently, it has become a standard practice to perform tests of both null hypotheses, the $I(1)$ and the $I(0)$; cf. Phillips and Xiao (1998) for an updated survey of unit root testing approaches. By proceeding in this way, it is often found that both null hypotheses are rejected, suggesting that many series are not well represented as either $I(1)$ or $I(0)$. In view of this outcome, the class of fractionally integrated processes, denoted as FI(d), where the order of integration d is extended to be any real number, have proved to be very useful in capturing the persistence properties of many long memory processes; cf. Baillie (1996) for a recent survey on this topic.

In general, the above mentioned tests are consistent when the alternative is a FI(d) series but their power turns out to be quite low when this is the case (see Diebold and Rudebusch, 1991 and Lee and Schmidt, 1996). In particular, this lack of power has motivated the development of new testing approaches that take this type of alternative explicitly into consideration. There is already a vast literature on this subject which can be basically classified into two groups. First, there are Wald type tests which, by working under the alternative hypothesis, provide point estimates of the memory parameter and build confidence intervals around it. Secondly, there are Lagrange Multiplier (LM) tests where statistics are evaluated under the corresponding null hypothesis. Within the first group, there are a very large number of rather heterogeneous contributions: parametric and semiparametric methods of estimating d both in the frequency and in the time domain (see, among others, Geweke and Porter-Hudak, 1983, Fox and Taqqu, 1986, Sowell, 1992, Robinson 1992, etc.). However, most of them lack power when used for testing purposes. On the one hand, the semiparametric techniques present large confidence intervals that include too often the null hypothesis when the alternative is close to that hypothesis. On the other hand, although in general the parametric methods present narrower confidence intervals, the suitability of the point estimation depends on the correct specification of the model. Within the second group, Robinson (1994a) and Tanaka (1999) have proposed LM tests in the frequency and the time domain, respectively. A distinctive feature of both LM tests is that, in contrast to the classical unit root tests where asymptotic distributions are nonstandard and require case-by-case numerical tabulation, they have standard asymptotic distributions. In particular, Robinson (1994a) attributes this different limiting behavior to the use of an explicit autoregressive (AR) alternative in the classical unit root testing approach. Nonetheless, a possible shortcoming of the LM approach is that by working under the null hypothesis of a unit root it obviously does not yield any direct information about the correct order of differencing d when the null is rejected.

In order to overcome those drawbacks, we propose in this paper a simple Wald type test in the time domain that has acceptable power properties while as a by product, it does provide information about the values of the memory parameter under the alternative hypothesis. It turns out to be a generalization of the Dickey-Fuller (D-F) test, originally developed for the case $I(1)$ versus $I(0)$, to the more general case of $FI(d_0)$ versus $FI(d_1)$ with $d_1 < d_0$ and thus, we will refer to it as the Fractional Dickey-Fuller (FD-F) test. The test is based on the OLS estimator or on the t-statistic associated to $\Delta^{d_1} y_{t-1}$ in a regression of $\Delta^{d_0} y_t$ on $\Delta^{d_1} y_{t-1}$ and, possibly some lags of $\Delta^{d_0} y_t$.We shall show that, in this framework, the standard or nonstandard limiting behavior of the proposed test statistics does depend on both the distance between the null and the alternative hypotheses and on the nature of the process under the null.

Since the FD-F is a Wald type test, a value of d is needed under the alternative hypothesis to implement it. When a simple hypothesis is considered, i.e., $H_0 : d = d_0$ versus $H_1 : d = d_1$, this last value is used to run the test. When more general hypotheses are considered, namely $H_0 : d = d_0$ versus $H_1 : d < d_0$, a preestimation under the alternative hypothesis can be carried out according to any of the available methods in the literature and then, this estimated value can be used to run the above mentioned regression. Notwithstanding, in order to preserve simplicity in the implementation of the test, we propose a new minimum distance estimator of d in the time domain. It exploits the properties, under the alternative hypothesis, of the statistics used in the FD-F test. Since the same statistics are used, its computation does not require much additional effort. Further, it possesses good properties at the asymptotic level ($\sqrt{T}-$consistency and normality) as well as in finite samples.

The advantages of this type of formulation stem from several sources. First, by generalizing the simple and well known D-F framework, it keeps its simplicity as one of its key features. Secondly, relative to LM tests, it provides an alternative which is not based at any point on a known density for the errors and therefore offers potential for greater robustness. Thirdly, the D-F regression framework allows to test the null of $I(1)$ versus some interesting composite alternatives, like for instance the alternative of $I(d)$ plus a break in the mean or in any other parameters of the processes. To the best of our knowledge, this can not be easily done with the available tests in the literature. Fourthly, this approach provides a natural framework for estimating the memory parameter d in a easy way. Lastly, it provides equal or even better results in finite samples in terms of power and size than most of the tests discussed earlier.

The rest of the paper is structured as follows. In Section 2 the FD-F test is defined and its asymptotic properties are derived. Also, its finite sample behavior is studied via Monte-Carlo simulation and it is compared with that of other leading unit root tests available in the literature. For expository purposes, we just consider in this section fractional white noise processes, which integration order is taken to be known under the alternative. Section 3 extends the analysis to the more realistic case where d needs to be estimated; for this, a new minimum distance estimator is proposed and its properties are analyzed. Section 4, in turn, provides the corresponding results for more general FI processes. Section 5 discusses some empirical applications of the previous tests. Finally, Section 6 draws some concluding remarks.

Proofs of theorems and lemmae are gathered in Appendix 1. A set of critical values needed to implement the fractional D-F test is reported in Appendix 2.

The following conventional notation is adopted throughout the paper. L is the lag operator, $\Delta = (1-L)$, $\Gamma(.)$ 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:

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

Likewise, $\{\pi_i\}$ is the notation for the coefficients of the expansion $\Delta^{1-d_1^*}$. Summations, denoted by $\sum$, go from 1 to T unless otherwise noted. $B(.)$ and $B_\delta(.)$ are standard Brownian motion (BM) and standard fractional Brownian motion (FBM) respectively, the later defined as, e.g., by Beran (1994, p.56), $\xrightarrow{w}$ and $\xrightarrow{p}$ denote weak convergence and convergence in probability respectively.

2. FRACTIONAL DICKEY-FULLER TEST

In this section we develop the basis of the Fractional Dickey-Fuller (FD-F) test. Our test-statistic follows the D-F approach closely. Let us consider a $FI(d)$ process given by,

$$(1-\rho L)^{d_0-d_1}\Delta^{d_1} y_t = u_t \qquad u_t = 0, \quad \text{for all } t \leq 0, \tag{2}$$

where $|\rho| \leq 1$, $d_1 < d_0$, and u_t is a linear covariance stationary process with strictly positive and bounded spectral density at zero frequency. Notice that this formulation is not restrictive since any $FI(d)$ process can be written in this way. This is so because ρ can take any value within the unit circle including zero. For any $|\rho| < 1$, y_t is an $FI(d_1)$ process and becomes $FI(d_0)$ when $\rho = 1$. Therefore, the null hypothesis of $FI(d_0)$ corresponds to the case $\rho = 1$ versus the alternative of $\rho < 1$. By applying the Beveridge and Nelson (B-N) decomposition to the polynomial $(1-\rho L)^{d_0-d_1}$ and adding and subtracting $(1-\rho)^{d_0-d_1} L$ we obtain:

$$(1-\rho L)^{d_0-d_1} = (1-\rho)^{d_0-d_1} L + \Delta^{d_0-d_1}(\Phi^*(L) + \Delta^{d_1-d_0+1}(1-\rho)^{d_0-d_1}), \tag{3}$$

where $\Phi^* (L)$ is given by:

$$\Phi^* (L) = \frac{(1-\rho L)^{d_0-d_1} - (1-\rho)^{d_0-d_1}}{\Delta^{d_0-d_1}}.$$

Substituting (3) into (2) we obtain:

$$\Delta^{d_0} y_t = -(1-\rho)^{d_0-d_1} \Delta^{d_1} y_{t-1} + \left(1 - \left(\Phi^* (L) + \Delta^{d_1-d_0+1} (1-\rho)^{d_0-d_1}\right)\right) \Delta^{d_0} y_t + u_t. \tag{4}$$

Expression (4) can be rewritten to obtain the model[1]:

$$\Delta^{d_0} y_t = \phi \Delta^{d_1} y_{t-1} + a_t, \tag{5}$$

where $a_t = \left(1 - \left(\Phi^* (L) + \Delta^{d_1-d_0+1} (1-\rho)^{d_0-d_1}\right)\right) \Delta^{d_0} y_t + u_t$ and $\phi = -(1-\rho)^{d_0-d_1}$. $\phi = 0$ and $\phi < 0$ correspond to the $FI(d_0)$ null hypothesis and the $FI(d_1)$ alternative, respectively. Thus, as in the D-F case, a simple test based on the normalized OLS estimator of ϕ or on its associated t-statistic can be conducted. Notice that when $d_0 = 1$ and $d_1 = 0$ we are in the standard $I(1)$ versus $I(0)$ D-F framework. The treatment of a_t in the estimation of ϕ will be covered in more detail in the following sections.

For the sake of simplicity, in the sequel we will mainly develop the test of a process being $I(1)$ versus $FI(d_1)$, $0 \leq d_1 < 1$, which turns out to be the most popular case in the literature. Values of $d_1 < 0$ are not considered since standard unit root tests are very well behaved in this situation.

An important particular case is when $d_0 = 1$, $\rho = 0$ and u_t is a white noise process which corresponds to the *random walk* versus the *simple fractional white noise* hypothesis. In order to simplify the presentation, in the following subsection we present the test statistics and explore their asymptotic behavior for the case where u_t is an *i.i.d.* process.

[1] Note that since $\left.\Phi^* (L) + \Delta^{d_0-d_1+1} (1-\rho)^{d_1-d_0}\right|_{L=0} = 1$, a_t does not contain contemporaneous values of $\Delta^{d_1} y_t$. Also, notice that this term vanishes under the null hypothesis of $\phi = 0$.

2.1. The test and its asymptotic properties

Let us set $d_0 = 1$ and $u_t = \varepsilon_t$ in expression (2), where ε_t is a sequence of zero-mean *i.i.d.* random variables with unknown variance σ^2. Consider the simple regression of Δy_t on $\Delta^{d_1} y_{t-1}$ as in equation (5). The OLS estimator of ϕ and its t-value are given by the usual expressions:

$$\hat{\phi}_{ols} = \frac{\sum \Delta y_t \Delta^{d_1} y_{t-1}}{\sum \left(\Delta^{d_1} y_{t-1}\right)^2}, \tag{6}$$

$$t_{\hat{\phi}_{ols}} = \frac{\sum \Delta y_t \Delta^{d_1} y_{t-1}}{S_T \left(\sum \left(\Delta^{d_1} y_{t-1}\right)^2\right)^{1/2}}, \tag{7}$$

where S_T^2 is given by:

$$S_T^2 = \frac{\sum \left(\Delta y_t - \hat{\phi}_{ols} \Delta^{d_1} y_{t-1}\right)^2}{T}. \tag{8}$$

Notice that a_t presents serial correlation yet, unlike parallel approaches in the unit root literature, as in Phillips (1997) and Phillips-Perron (1988), no semiparametric corrections are needed to account for this serial correlation in the asymptotic distributions under the null hypothesis. This is due to the fact that under $H_0 : \rho = 1$, the term $\left(1 - \left(\Phi^*(L) + \Delta^{d_1}(1-\rho)^{1-d_1}\right)\right)\Delta y_t$ vanishes and a_t equals ε_t.

In order to obtain the asymptotic properties of $\hat{\phi}_{ols}$ and $t_{\hat{\phi}_{ols}}$ under the null hypothesis, we need the following auxiliary result.

Lemma 1 *Let ε_t be an i.i.d $(0,\sigma^2)$ sequence of random variables such that $E\left(\varepsilon_t^4\right) < \infty$ and consider the linear process:*

$$\Delta^{\delta} x_t = \varepsilon_t \qquad \delta \in (0,1)$$

The process $\{x_{t-1}\varepsilon_t\}$ is a martingale difference sequence and:

- *if* $0 < \delta < 0.5$:

$$T^{-1/2} \sum_{t=2}^{T} x_{t-1}\varepsilon_t \xrightarrow{w} N\left(0, \sigma^4 \frac{\Gamma(1-2\delta)}{\Gamma^2(1-\delta)}\right); \tag{9}$$

- *if* $\delta = 0.5$:

$$(T\log(T))^{-1/2}\sum_{t=2}^{T}x_{t-1}\varepsilon_t \xrightarrow{w} N\left(0, \frac{\sigma^4}{\pi}\right). \tag{10}$$

- *if* $0.5 < \delta < 1$:

$$T^{-\delta}\sum_{t=2}^{T}x_{t-1}\varepsilon_t \xrightarrow{w} \int_0^1 B_\delta(r)\,dB(r)$$

In view of Lemma 1, the following two theorems state the consistency and derive the asymptotic distributions of the suitably standardized OLS estimator of ϕ and its t-ratio under the null hypothesis of $I(1)$.

Theorem 1 *Under the null hypothesis of* $I(1)$, $\hat{\phi}_{ols}$ *is a consistent estimator of* $\phi = 0$ *and converges to the true value at a rate* T^{1-d_1} *when* $0 < d_1 < 0.5$, $(T\log T)^{0.5}$ *when* $d_1 = 0.5$ *and at the usual rate* $\sqrt{T}$ *when* $0.5 < d_1 < 1$. *Its asymptotic distribution is given by:*

$$T^{1-d_1}\hat{\phi}_{ols} \xrightarrow{w} \frac{\int_0^1 B_{1-d_1}(r)\,dB(r)}{\int_0^1 B_{1-d_1}^2(r)\,dr} \qquad \text{if } 0 \le d_1 < 0.5, \tag{11}$$

$$(T\log T)^{1/2}\hat{\phi}_{ols} \xrightarrow{w} N(0,\pi) \qquad \text{if } d_1 = 0.5, \tag{12}$$

$$T^{1/2}\hat{\phi}_{ols} \xrightarrow{w} N\left(0, \frac{\Gamma^2(d_1)}{\Gamma(2d_1-1)}\right) \qquad \text{if } 0.5 < d_1 < 1. \tag{13}$$

Theorem 2 *Under the null hypothesis of* $I(1)$, *the asymptotic distribution of* $t_{\hat{\phi}_{ols}}$ *in* (7) *is given by:*

$$t_{\hat{\phi}_{ols}} \xrightarrow{w} \frac{\int_0^1 B_{1-d_1}(r)\,dB(r)}{\left(\int_0^1 B_{1-d_1}^2(r)\,dr\right)^{1/2}} \qquad \text{if } 0 \le d_1 < 0.5,$$

$$t_{\hat{\phi}_{ols}} \xrightarrow{w} N(0,1) \qquad \text{if } 0.5 \le d_1 < 1$$

It is clear from Theorems 1 and 2 that the asymptotic behavior depends on the distance between the null and the alternative hypothesis. When the alternative is also a non stationary process ($0.5 \le d_1 < 1$), the asymptotic distributions are standard and, by contrast, they turn out to be functions of fractional Brownian motion, and therefore nonstandard, whenever the alternative is a stationary process. Therefore, what matters in determining the asymptotic behaviour is the distance in the sense described above, between both hypotheses, rather than the explicit autoregressive alternative used in the classical unit root approach, as it was suggested in Robinson (1994a). More generally, for the case of FI(d_0) versus FI(d_1), the asymptotic behaviour is determined by both the distance between the integration orders d_0 and d_1 and by the nature of the process under the null. If $d_0 > 0.5$, that is, when the process is nonstationary under the null, the distributions will be nonstandard whenever $d_0 - d_1 > 0.5$ and standard otherwise. However, when $d_0 < 0.5$, the distributions are always standard. Notice also that when $d_0 = 1$ and $d_1 = 0$, we recover again the super-consistency of $\hat{\phi}_{ols}$ and the asymptotic distributions of $t_{\hat{\phi}_{ols}}$ and $\hat{\phi}_{ols}$ correspond to the ones derived by Dickey and Fuller $(1979, 1981)$.

Next, let us consider the behavior of the test under the alternative hypothesis. It is a well-known fact that the standard D-F test is consistent under fractional alternatives (see Sowell, 1990, and Mármol, 1998, for the nonstationary and the stationary fractional alternative respectively). In the following theorem we prove a more general result which states the consistency of the FD-F test when the Data Generating Process (DGP) is a fractional white noise with an integration order d_1^* that could be possibly different of the order d_1 used to compute the t-statistic. Therefore, this theorem encompasses the results for the standard D-F test as a particular case when d_1, the integration order used in the regression, is set equal to 0, i.e. the regressor is $y_{t-1.}$

Theorem 3 *If the DGP is given by:*

$$\Delta^{d_1^*} y_t = \varepsilon_t, \qquad \varepsilon_t = 0 \quad \text{if } t \leq 0,\ d_1^* \in [0, 1) \tag{14}$$

the tests based on the OLS estimator in (6) *or on the t-statistic in* (7) *associated to* ϕ *in the regression* $\Delta y_t = \phi \Delta^{d_1} y_{t-1}$*, are consistent for any value of* $d_1 \in [0, 1)$.

Theorem 3 turns out to be extremely helpful since it guarantees the consistency of the proposed tests even when a wrong value of d_1 under the alternative is considered to implement the test, insofar as $d_1 \in [0, 1)$.

2.2. Finite sample performance of the FD-F test

In this section we examine the finite sample behavior of the proposed FD-F test by means of Monte Carlo simulations and compare it with a number of available unit root tests when the alternative hypothesis is a fractional process. The computations were carried out using MATLAB 5.1 for UNIX.

In order to study the size and the power of the proposed test, we have generated random walks and FI processes of order $d_1^* \in [0, 0.9)$ as defined in (14) respectively, with NID$(0, 1)$ disturbances and for samples of size T=100 and T=400. The number of replications is 1000 and the significance level is 5%. Table 1 reports the rejection frequencies at the 5% significance level corresponding to the case where the true DGP is a random walk. For the FD-F test, the t-statistic was computed with different values of d_1 and size is reported for these values[2]. Table 1 also reports the size of the standard D-F test, the Geweke and Porter-Hudak test (1983) and the LM tests of Tanaka (1999) and Robinson (1994*a*), denoted by GPH, TAN and ROB respectively. The number of periodogram ordinates included in the log-periodogram regression for the computation of the GPH estimator is set equal to $\sqrt{T}$, as it was originally suggested by the authors. Table 2 reports the power of these tests. For the FD-F test we have computed the t-statistic in (7), when the true order of integration d_1^* was used to compute its value[3]. When $d_1^* \in [0, 0.5)$, the numerical critical values for $T = 100$ and $= 400$ in Appendix 2 are used to compute the size-corrected power and for the remaining range of values of d_1^*, the nominal critical values given by a N$(0, 1)$ are considered.

[2] We just report the size for values of $d_1 \in [0.5, 0.9]$ for which critical values of a N(0,1) are used. For the remaining range, numerical critical values for $T = 1000$ (see Appendix 2) have been computed and used to calculate the size. The obtained figures are very similar to those reported and they are available upon request.

[3] Table 5 reports the power of the FD-F test when the parameter d_1 is estimated according to the method presented in Section 3. Remark that very similar figures are obtained in both cases.

Table 1. Size of unit root tests against FI alternatives.

	FD-F						D-F	GPH	TAN	ROB
$T = 100$	d_1	0.5	0.6	0.7	0.8	0.9	5%	4.4%	3.1%	9.7%
	$size$	5.7%	6.0%	5.4%	4.9%	5.2%				
$T = 400$	d_1	0.5	0.6	0.7	0.8	0.9	5%	4.6%	3.7%	7.9%
	$size$	5.6%	5.3%	5.1%	5.0%	5.3%				

Table 2. Power of the unit root tests against FI alternatives.

$T = 100$									
d_1^*	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
FD-F	100%	100%	100%	100%	100%	100%	93.2%	65.4%	26.9%
D-F	100%	100%	100%	100%	88%	71%	43.9%	17.9%	9.0%
GPH	92.6%	86.6%	74.6%	61.8%	51.6%	33.6%	21.6%	13.8%	6.6%
TAN	100%	100%	100%	100%	99.9%	98.6%	89.9%	57.6%	16.5%
ROB	100%	100%	100%	100%	99%	96.6%	74.9%	28.5%	16.2%

$T = 400$									
d_1^*	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
FD-F	100%	100%	100%	100%	100%	100%	100%	99.9%	63.1%
D-F	100%	100%	100%	100%	100%	97.8%	79.8%	43.9%	20.3%
GPH	100%	100%	98.8%	95.2%	84.4%	66.8%	44.0%	22.2%	10.8%
TAN	100%	100%	100%	100%	100%	100%	100%	99.8%	64.2%
ROB	100%	100%	100%	100%	100%	100%	100%	97.8%	59.9%

From Tables 1 and 2, it follows that FD-F performs very well both in terms of size and power. In terms of size of the tests, both LM tests presents size distortions in opposite directions, while other tests perform correctly. With respect to power, the FD-F test is superior in moderate samples ($T = 100$) and behaves similarly to

Tanaka's test for larger sample sizes. It is remarkable the low power of the GPH procedure that performs even worse than the standard DF test. As expected, as the sample size T grows, the power of the tests improves given their consistency.

Since the true value of $d_1 = d_1^*$, is unknown in practice, we have explored through simulations the robustness of the FD-F test against misspecifications in the value d_1 used to run regression (5). Table 3 reports the power of the test when incorrect values are used. More specifically, we have considered deviations of ± 0.1, ± 0.2 and ± 0.3 from d_1^*. Note that very similar rejection rates to those reported in Table 1, where the true d_1 was used to compute the power, are obtained. This totally agrees with the result in Theorem 3. It is worth noticing that the power tends to decrease when values of d_1 greater than d_1^* are employed, especially when $T = 100$ and $d_1^* > 0.7$. This is so because in this case we are considering alternative hypotheses which are closer to the null and therefore the test is less powerful. Summing up, the previous results suggest that the FD-F test is quite robust to misspecifications in the value of d_1 in finite samples and that the good power properties of the test do not depend much on the accuracy of the estimation of d_1.

Table 3. Power of the FD-F test under misspecification of the true d_1 .

$T = 100$										
$d_1 \backslash d_1^*$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$d_1^* - 0.1$	-	100%	100%	100%	100%	100%	99.5%	96.5%	67.5%	30.5%
$d_1^* - 0.2$	-	-	100%	100%	100%	100%	100%	97.0%	74.5%	32.5%
$d_1^* - 0.3$	-	-	-	100%	100%	100%	100%	98.0%	71.5%	32.0%
$d_1^* + 0.1$	100%	100%	100%	100%	100%	100%	98.0%	89.5%	56.5%	-
$d_1^* + 0.2$	100%	100%	100%	100%	100%	100%	95.5%	80.5%	-	-
$d_1^* + 0.3$	100%	100%	100%	100%	100%	100%	92.5%	-	-	-

$T = 400$										
d_1/d_1^*	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$d_1^* - 0.1$	-	100%	100%	100%	100%	100%	100%	100%	100%	71.0%
$d_1^* - 0.2$	-	-	100%	100%	100%	100%	100%	100%	100%	77.5%
$d_1^* - 0.3$	-	-	-	100%	100%	100%	100%	100%	100%	81%
$d_1^* + 0.1$	100%	100%	100%	100%	100%	100%	100%	100%	99.0%	-
$d_1^* + 0.2$	100%	100%	100%	100%	100%	100%	100%	100%	-	-
$d_1^* + 0.3$	100%	100%	100%	100%	100%	100%	100%	-	-	-

3. ESTIMATION OF THE MEMORY PARAMETER

To implement the FD-F test, we need a value of the memory parameter d under the alternative hypothesis. Sometimes, we can have an *a priori* knowledge of this value but, in general, we will need to estimate it. There is substantial literature on the estimation of $FI(d)$ models, both in the frequency and the time domain. In view of the results in section 2, the good power properties of the test do not rely much on the accuracy of the estimated value of d neither in finite samples nor in the asymptotics (see Table 3 and Theorem 3 respectively) and therefore any of the existing methods in the literature could be used for these purposes. Nevertheless, to preserve simplicity in the implementation of the test, this section introduces a minimum distance estimator (MDE) of the memory parameter in the time domain for the i.i.d.*i.i.d.* case discussed in Section 2. This type of minimum distance estimator provides a conceptually very attractive alternative to ML estimators, since it does not require any distributional assumption. The method that we propose presents important advantages within this framework. It is based on the properties under the alternative of the same test-statistics (6) or (7) defined in Section 2, and therefore its computation does not require much additional effort. Further, it can be applied both in stationary and non-stationary settings and the estimators are $\sqrt{T}-$consistent and asymptotically normal in all their domain. Finally, it has a good performance in finite samples.

Along these lines, several approaches can be found in the literature. For example, Robinson (1994*b*) has proposed a semiparametric MDE in the time domain which was applied by Delgado and Robinson (1994) to model Spanish inflation rates. Hall et al. (1997) provide the rates of convergence but their distributional properties are not known yet. Tieslau el al. (1996) have introduced a parametric MDE based on the minimization of a quadratic distance between the estimated and the theoretical autocorrelations of the ARFIMA (p, d, q) process. This estimator, however, is only defined in the stationary range of d and is $\sqrt{T}-$ consistent and asymptotically normal only when $d \in (-0.5, 0.25)$. Chung and Schmidt (1995) introduce a modification to the previous estimator and show, by applying the results on autocorrelations of Hosking (1996), that it is possible to obtain a $\sqrt{T}$-consistent and asymptotically normally distributed estimator of d in the whole invertible and stationary range, provided some functions of the autocorrelations are employed in the criterion function. Lastly, Galbraith and Zinde-Walsh (1997), present a parametric estimator in the time domain based on autoregressive approximation. Since it does not require the existence of autocorrelations, it can be applied to nonstationary settings but its consistency has not been proven yet in this framework.

The basis of our estimation method is the following: let us again consider the OLS estimator of ϕ, its associated t-statistic defined in (6) and (7) and their asymptotic behavior when the true DGP is a simple fractional white noise process. Note that, conditioning on the data, these statistics are just a function of the value d_1 used to run the regression. On the other hand, it is straightforward to check that if the DGP is given by (14), the limit of these statistics evaluated at the true value of $d_1 = d_1^*$, are the following:

$$\phi\left(d_1^*\right) = \frac{\sum \Delta y_t \Delta^{d_1^*} y_{t-1}}{\sum \left(\Delta^{d_1} y_{t-1}\right)^2} \xrightarrow{p} \pi_1 = -\left(1 - d_1^*\right),$$

and,

$$\frac{t_\phi\left(d_1^*\right)}{T^{1/2}} = \frac{\sum \Delta y_t \Delta^{d_1^*} y_{t-1}}{T^{1/2} S_T \left(\sum \left(\Delta^{d_1^*} y_{t-1}\right)^2\right)^{1/2}} \xrightarrow{p} \Pi\left(d_1^*\right),$$

where:

$$\Pi\left(d_1^*\right) = \frac{\pi_1}{\left(\sum_{i=0}^{\infty}\pi_i^2 - \pi_1^2\right)^{1/2}} = \frac{\left(d_1^* - 1\right)\Gamma\left(2 - d_1^*\right)}{\left(\Gamma\left(3 - 2d_1^*\right) - \left(\left(d_1^* - 1\right)\Gamma\left(2 - d_1^*\right)\right)^2\right)^{1/2}}.$$

Note that for other alternative values of d_1, the statistics converge to different expressions from the ones derived above. Next, consider the distance between the statistics and the previous limits evaluated in any value of d_1. It is clear that for large T, d_1^* will make these distances almost negligible. Therefore, we define the following estimators of the memory parameter d_1 :

$$\hat{d}_{T_\phi} = \arg\min_{d \in D} V_{T_\phi}\left(d\right), \tag{15}$$

or alternatively,

$$\hat{d}_{Tt_\phi} = \arg\min_{d \in D} V_{Tt_\phi}\left(d\right); \tag{16}$$

where:

$$V_{T_\phi}\left(d\right) = \left(\phi\left(d\right) - \pi_1\left(1 - d\right)\right)^2, \tag{17}$$

$$V_{Tt_\phi}\left(d\right) = \left(\frac{t\left(d\right)}{T^{1/2}} - \Pi\left(d\right)\right)^2. \tag{18}$$

Thus, the idea behind these estimators is that they minimize a quadratic distance between the OLS estimator of ϕ or its associated t-statistic conveniently standardized and the value to which they converge when the correct value d_1 is inserted in the criterion function.

Let us now state formally the asymptotic properties of the above defined estimators.

Theorem 4 *If y_t is a fractional white noise as defined in (14), D is a subset of $\Re$ given by the interval* $(-1/2, 1)$ *and $d_1^* \in D$, then $\hat{d}_{T_\phi}$ and $\hat{d}_{Tt_\phi}$, as defined in* (15) *and* (16), *are consistent estimators of d_1^*.*

Moreover the asymptotic distributions of $\hat{d}_{T_\phi}$ and $\hat{d}_{Tt_\phi}$ are given in the following theorem.

Theorem 5 *Under the conditions of Theorem 4, and assuming that* $\hat{d}_{T_\phi}$ *and* $\hat{d}_{Tt_\phi}$ $\in D$*, then as* $T \rightarrow \infty$*, it holds that:*

$$T^{1/2}\left(\hat{d}_{T_\phi} - d_1^*\right) \xrightarrow{w} N\left(0, w_1\right), \tag{19}$$

and

$$T^{1/2}\left(\hat{d}_{Tt_\phi} - d_1^*\right) \xrightarrow{w} N\left(0, w_2\right), \tag{20}$$

where

$$w_1 = \frac{\left(1 + \sum_{i=2}^{\infty} \pi_i^2 + 2\pi_2\right)}{\left(1 - \sum_{i=2}^{\infty} \pi_i / \left(i-1\right)\right)^2},$$

and

$$w_2 = \frac{\left(1 + \sum_{i=2}^{\infty} \pi_i^2 + 2\pi_2\right)\left(1 + \sum_{i=2}^{\infty} \pi_i^2\right)}{\left(\pi_1 \sum_{i=0}^{\infty} \pi_i^2 \left(\sum_{j=2}^{i+1} 1/(d_1^* + i - j)\right) - \sum_{i=0}^{\infty} \pi_i^2 \left(1 - \sum_{i=2}^{\infty} \frac{\pi_i}{i-1}\right)^2\right)^2}.$$

It is worth noticing that the previous result do not allow to test for a unit root, since the previous distributions do not hold when $d_1^* = 1$.

3.1. Behavior of the estimator of d in finite samples

In this section we evaluate the accuracy of the previous estimators of d by means of Monte Carlo experimentation. We have generated processes of the form (14) with $\varepsilon_t \sim NID\left(0,1\right)$, where $\hat{d}_{T_\phi}$ and $\hat{d}_{Tt_\phi}$ have been computed as the values which minimize (17) and (18) respectively. The initial values needed to start the minimization algorithm where set equal to zero in all cases.

The following table presents the mean of the estimated d and its standard deviation, that is: $\hat{d} = \sum_1^m \hat{d}_i/m$ and $\hat{\sigma} = \left(\sum_1^m (\hat{d}_i - \hat{d})^2/m\right)^{1/2}$, where m is the number of replications and is equal to 500, for two different sample sizes T=100 and T=400, corresponding to the estimator based on the t-statistic as defined in (16). Very similar figures were obtained for the estimator based on the OLS coefficient and therefore they are not reported here, although they are available upon request. Note that the Monte-Carlo standard deviation of $\hat{d}$ is equal to $\hat{\sigma}/\sqrt{m}$.

Table 4. **Estimation of d in the i.i.d. case based on the** $t(d)$ **−statistic**.

$T = 100$										
$\hat{d}/d_1^*$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{d}_{T_{t\phi}}$	0.001	0.115	0.205	0.311	0.407	0.501	0.602	0.704	0.801	0.901
$\hat{\sigma}$	0.109	0.110	0.106	0.108	0.110	0.110	0.109	0.109	0.107	0.104

$T = 400$										
$\hat{d}/d_1^*$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{d}_{T_{t\phi}}$	0.00	0.102	0.207	0.296	0.405	0.501	0.59	0.701	0.800	0.901
$\hat{\sigma}$	0.046	0.050	0.049	0.048	0.049	0.050	0.050	0.048	0.042	0.047

In general, we it is found that the proposed estimator provides very accurate estimates for both sample sizes and therefore it turns out to be very useful in implementing a feasible FD-F test.

3.2. Power of the test with estimated d_1

In this section we examine the power of the FD-F test when the value of the memory parameter is unknown but is estimated according to the MDE criterium defined in (16). The same experiment as in section 2.2. was conducted but in this case the t-value was computed with the estimated value of d_1. The following table reports the power of the test in this case.

Table 5. **Power of the FD − F test when** $\mathbf{d}_1$ **is estimated**.

T\\d_1^*	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
T=100	100%	100%	100%	100%	100%	99.4%	97.7%	86.4%	65.4%	25.1%
T=400	100%	100%	100%	100%	100%	100%	100%	100%	98.9%	63.1%
T=1000	100%	100%	100%	100%	100%	100%	100%	100%	100%	93.7%

As expected, the basic observation that we draw from this simulation study when compared to the results in Table 1, is that replacing the true d_1 by its estimated value hardly affects the power properties of the FD-F test.

In what follows we will examine whether these promising results extend beyond the framework of a simple fractional white noise model examined so far.

4. THE AUGMENTED FRACTIONAL DICKEY-FULLER (AFD-F)

For some applications, such as financial time series, the fractional white noise hypothesis could be adequate but, for many others, serially correlated errors should be allowed for. In this section, following the augmented Dickey-Fuller (AD-F) spirit, we show that the asymptotic distribution stated in Theorem 2 for the test based on the t-statistic still holds provided some lags of $\Delta^{d_0} y_t$ are appropriately introduced in the regression proposed in Section 2.

4.1. Definitions and asymptotic properties

Consider the model in (2) and assume now that u_t follows an autoregressive process[4] of order p, namely, $\alpha(L) u_t = \varepsilon_t$ and $\alpha(L) = 1 - \alpha_1 L - ... - \alpha_p L^p$, with all its roots outside the unit circle.

Multiplying equation (4) by $\alpha(L)$ and setting $d_0 = 1$, we obtain:

$$\begin{aligned} \alpha(L)\Delta y_t &= -\alpha(1)(1-\rho)^{1-d_1}\Delta^{d_1} y_{t-1} + \\ &\quad +\alpha(L)\left(1 - (\Phi^*(L) + \Delta^{d_1}(1-\rho)^{1-d_1}\left(1 + \frac{\tilde{\alpha}(L)}{\alpha(L)}L\right))\right)\Delta y_t + \varepsilon_t \end{aligned} \tag{21}$$

where $\tilde{\alpha}(L) = \Delta^{-1}(\alpha(L) - \alpha(1))$ results from the B-N decomposition of the polynomial $\alpha(L)$. As before, we can rewrite expression (21) and obtain:

$$\Delta y_t = \phi \Delta^{d_1} y_{t-1} + \alpha_1 \Delta y_{t-1} + ... + \alpha_p \Delta y_{t-p} + \xi_t \tag{22}$$

where $\xi_t = \alpha(L)\left(1 - (\Phi^*(L) + \Delta^{d_1}(1-\rho)^{1-d_1}\left(1 + \frac{\tilde{\alpha}(L)}{\alpha(L)}L\right))\right)\Delta y_t + \varepsilon_t$ and

[4] As in the standard Augmented D-F, an extension to ARMA(p, q) processes is possible following the same arguments as in Said and Dickey (1984) and Ng and Perron (1995).

$\phi = -\alpha(1)(1-\rho)^{1-d_1}$. Thus, as in (5), the null hypothesis of unit root and the alternative of $FI(d_1)$ can be both stated in terms of the parameter ϕ, corresponding the former to the case where $\phi = 0$, and the latter to the case where $\phi < 0$. Again, this motivates a test based on the t-statistic[5] of the OLS estimator of ϕ in equation (22). It is noteworthy that, as in the FD-F case, the process ξ_t presents serial correlation but under the null hypothesis of $d_0 = 1$, becomes an i.i.d. process. Therefore, as in (5), this serial correlation has no influence on the asymptotic behavior of the statistic under the null and no correction is needed in this case either. We find that the asymptotic behavior of $t_{\hat{\phi}_{ols}}$ calculated in (22), does not change with respect to the uncorrelated case under the null hypothesis. This result encompasses the one obtained by Dickey and Fuller in the standard $I(1)$ versus $I(0)$ case in an augmented autoregression, as it is stated in the following theorem.

Theorem 6 *Under the null hypotheses of $I(1)$,*

- *The distribution of $t_{\hat{\phi}_{ols}}$ is the same as in the i.i.d. case given in Theorem 2.*
- *$(\hat{\alpha}_1, ...\hat{\alpha}_p)'$ are asymptotically normally distributed for any value $d_1 \in [0,1)$ used in the regression and standard inference applies.*

Notice that, as in the AD-F framework, a test based on the estimated coefficient $\hat{\phi}_{ols}$ will depend on nuisance parameters, therefore the results in Theorem 1 do not hold when short term structure is allowed in the errors.

Let us now consider the behavior of the test under the alternative hypothesis of fractional integration. Kramer (1998) proves the consistency for the augmented D-F test (AD-F) when the DGP is a fractional white noise, provided that the number of lags included in the regression does not tend to infinity too fast. To our knowledge, however, no results available on the consistency of the AD-F when the DGP is a

[5] A test on the autoregressive coefficient ϕ could also be possible but, as in the standard ADF approach, it will depend on nuisance parameters. See Xiao and Phillips (1998) for a semiparametric treatment of this problem in the ADF framework.

more general ARFIMA process. The following theorem states the consistency of the augmented FD-F (AFD-F) test when the DGP is an ARFIMA(p, d_1, q) and the number of lags included in the regression tends to infinity with T.

Theorem 7 *If the DGP is given by $\Delta^{d_1^*} y_t = u_t$, with $d_1^* \in [0, 1)$ and u_t is a stationary and invertible ARMA(p, q) process satisfying $a_p(L) u_t = b_q(L) \varepsilon_t$, $\varepsilon_t \sim i.i.d\,(0, \sigma^2)$, $a_0 = b_0 = 1$ with p and q possibly unknown, the t-statistic associated to the OLS estimator of ϕ in the regression $\Delta y_t = \phi \Delta^{d_1^*} y_{t-1} + \delta_1 \Delta y_{t-1} + ... + \delta_k \Delta y_{t-k} + \eta_{tk}$ diverges to $-\infty$ at a rate of $\sqrt{T}$ as $k, T \to \infty$ and $k/T \to 0$, implying the consistency of the test.*

4.2. Power of the AFD-F test in finite samples

A new Monte Carlo study has been carried out to evaluate the performance of the test in this more general framework. As in Tanaka (1999), we have considered $FI\,(d_1)$ processes with an AR(1) structure in the errors, that is:

$$\Delta^{d_1^*} y_t = u_t, \tag{23}$$

$$u_t = \alpha u_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim n.i.d.\,(0, 1) \tag{24}$$

for different values of d_1 and α, having computed the t-statistic associated to $\hat{\phi}$ in the model:

$$\Delta y_t = \phi \Delta^{d_1^*} y_{t-1} + \zeta \Delta y_{t-1} + a_t. \tag{25}$$

We have computed the size of the test using the same critical values as in the $i.i.d.$ case and we have obtained similar figures than in Table 1. This implies that the test is very well behaved in terms of size even when short term structure is allowed in the errors since it is able to capture this structure in u_t. Figures are not reported although they are available upon request.

Table 6 provides the rejection frequencies at the 5% significance level. The number of replications is 1000. In all cases, the true d_1 has been used to run the regression.

Recall that, by expression (21), $\phi = -\alpha(1)(1-\rho)^{1-d_1^*}$. Therefore the closer the polynomial $\alpha(L)$ is to have a unit root, the closer to zero ϕ is and, consequently, the smaller the power of the test will be.

In the simulations below we observe that when the value of α is small in absolute value or negative, the test remains quite powerful. Nevertheless, as expected, for values of α close to 1, the power decreases as a consequence of the difficulty of identifying the source of the persistence.

Table 6. Power of the test with stationary AR(1) **errors.**

α=-0.8										
T/d_1^*	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
T=100	100%	100%	100%	100%	100%	100%	99.1%	97.2%	77.2%	31.0%
T=400	100%	100%	100%	100%	100%	100%	100%	100%	100%	78.6%
α=-0.5										
T=100	100%	100%	100%	100%	100%	100%	99.0%	94.0%	69.0%	27.4%
T=400	100%	100%	100%	100%	100%	100%	100%	100%	99.9%	71.5%
α=0.2										
T=100	100%	100%	100%	100%	100%	98.8%	90.0%	65.6%	39.2%	19%
T=400	100%	100%	100%	100%	100%	100%	100%	99.6%	85.6%	38.6%
α=0.6										
T=100	100%	100%	99.7%	95.5%	84.3%	63.5%	42.5%	36.0%	17.2%	9.9%
T=400	100%	100%	100%	100%	100%	100%	95.4%	68.2%	32.3%	14.1%

We have also used simulations to study the robustness of the test when d_1 is misspecified. Table 7 reports the power of the test for $\alpha = 0.2$ when an incorrect value of d_1 is used to run the regression. As in the *i.i.d.* case, we have considered deviations of $\pm 0.1, \pm 0.2$ and ± 0.3 from the true d_1. Alternative values of α have been considered and similar results hold. We can observe that the power remains very close to the one obtained in Table 6 when the true value of d_1 was used.

Table 7. Power of the FAD-F test under misspecification of d_1.

$\alpha = 0.2$										
$T = 100$										
$d_1 \backslash d_1^*$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$d_1^* - 0.1$	-	100%	100%	100%	100%	96.2%	86.4%	64.9%	40.1%	17.0%
$d_1^* - 0.2$	-	-	100%	100%	100%	97.2%	83.3%	53.1%	43.2%	17.0%
$d_1^* - 0.3$	-	-	-	100%	100%	94.1%	85.2%	58.2%	35.4%	15.2%
$d_1^* + 0.1$	100%	100%	100%	100%	100%	97.2%	89.3%	58.7%	38.2%	-
$d_1^* + 0.2$	100%	100%	100%	100%	100%	96.2%	88.5%	55.3%	-	-
$d_1^* + 0.3$	100%	100%	100%	100%	100%	94.6%	79.4%	-	-	-
$T = 400$										
$d_1 \backslash d_1^*$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$d_1^* - 0.1$	-	100%	100%	100%	100%	100%	100%	100%	83.1%	35.0%
$d_1^* - 0.2$	-	-	100%	100%	100%	100%	100%	100%	85.3%	35.2%
$d_1^* - 0.3$	-	-	-	100%	100%	100%	100%	100%	80.4%	33.5%
$d_1^* + 0.1$	100%	100%	100%	100%	100%	100%	100%	100%	79.2%	-
$d_1^* + 0.2$	100%	100%	100%	100%	100%	100%	100%	100%	-	-
$d_1^* + 0.3$	100%	100%	100%	100%	100%	100%	100%	-	-	-

4.3. Power of the AFD-F test when d_1 is estimated

Finally, we analyze the power of the test when d_1 is estimated according to the MD criterion. For illustrative purposes, we have considered the ARFIMA$(1, d, 0)$ process. Since in this case the process presents short term structure, we consider the OLS estimator of the autoregressive parameter α for a given value of d, namely,

$\hat{\alpha}(d) = \frac{\sum \Delta^d y_{t-1} \Delta^d y_t}{\sum (\Delta^d y_{t-1})^2}$, and define the residuals:

$$u_t(d) = y_t - \hat{\alpha}(d) y_{t-1}.$$

If we consider the normalized coefficient or the t statistics defined in (6) and (7) associated to the process $u_t(d)$ it is clear that, evaluated in d_1^*, they converge to the same limits as in the simple $i.i.d.$ case, as a result of the consistency of $\hat{\alpha}(d_1^*)$. Therefore it is possible to estimate d_1 using the criteria defined in Section 3, albeit this time on the series $u_t(d)$.

Table 8 reports the rejection frequencies at the 5% signification level of an ARFIMA$(1, d_1, 0)$ process with autoregressive parameter $\alpha = 0.2$ and 0.6 based on the t-statistic associated to ϕ in the regression model: $\Delta y_t = \phi \Delta^{\hat{d}} y_{t-1} + \zeta \Delta y_{t-1}$. The value of d used in the regression was estimated according to the procedure described above.

Table 8. Power of the FAD-F with estimated **d** and AR(1) errors.

$\alpha = 0.2$										
T\\d_1^*	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
T=100	100%	99.3%	99.2%	98.4%	98.0%	93.5%	78.5%	49.3%	38%	18.2%
T=400	100%	100%	100%	100%	100%	100%	99.3%	97%	83.5%	30%

$\alpha = 0.6$										
T\\d_1^*	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
T=100	99.1%	95.4%	89.7%	88.2%	82.9%	48.3%	39.3%	34.2%	15%	8.4%
T=400	100%	100%	100%	100%	100%	96.5%	91.5%	58%	31%	12.4%

Again we observe that the decrease in power when the true value of d is substituted by its corresponding estimated value is small, and similar to the $i.i.d.$ case. Summarizing, in Section 4 we have shown that serially correlated errors can be treated in the same way as in the AD-F framework.

5. APPLICATIONS.

In order to provide some empirical applications of the testing and estimation methods proposed in this paper, we have examined several political, financial and macroeconomic series for which evidence of fractional integration has been found before in the literature. More specifically, in the political context we have analyzed U.K. Gallup opinion poll series (cf. Byers et al., 1997). In the financial area, stock returns data and some transformations of them (cf. Lo, 1991, Ding et al. 1993, Lobato and Savin, 1997). Finally, in the macroeconomic field, the Nelson and Plosser's (1982) data set has also been examined. The following subsections present our main conclusions on any of the analyzed series.

5.1. Political data

First, we consider the Gallup opinion poll series of support for the Conservative and the Labour parties in the UK, analyzed by Byers et al. (1997). The basic finding of their work is that the logistic transformation of these series (after removal of the deterministic components due to the 'election cycle' effect[6]) is well characterized by a fractional white noise wih $0.5 < d < 1$, so it is nonstationary albeit mean reverting. The results obtained using our method are very similar. The data is monthly and its length is 434 observations. We have considered the logistic transformation of the series centered by their sample mean (which corresponds to the case of treating the effect of the election cycle as a constant). Table 9a presents the results of applying the FD-F test on both series, i.e., regressing Δy_t on $\Delta^{\hat{d}_1} y_{t-1}$ and computing the t-statistic associated to the estimated coefficient . Asterisks denote rejection of the null hypothesis at the nominal 5% level, using the N(0, 1) critical values since $\hat{d}_1$ is the nonstationary range.

[6] This effect refers to the tendency of a party's support to depend on the proximity of an election in a deterministic way.

Table 9a. FD-F test on opinion polls.

	$t_{H_0:d_0=1}$	$\hat{d}_1$
Conservatives	-4.60*	0.772
Labour	-7.09*	0.707

The test strongly rejects the null of unit root, showing evidence of fractional integration. The estimated values of d_1 are close to the ones obtained by Byers et al. using Sowell's maximum likelihood approach: 0.779 and 0.726 for the Conservative and Labour cases, respectively.

In view of the results, the conclusion that can be drawn is that the shocks (news) have a transitory effect on voting intentions and that voters forget eventually but not rapidly, i.e., voters have a long memory of events. Similar results are obtained for the Spanish opinion polls (see Dolado et al.,1999).

5.2. Financial data

To illustrate the use of the proposed test with a null hypothesis different from the unit root, we consider stock returns data. There have been several papers analyzing the long term properties of stock returns (cf. Lo, 1991, Ding et al. 1993, Lobato and Savin, 1997) and their central conclusion is that no evidence of long memory is found in the original series. Yet it does appear that when some transformations of them are used, i.e., when the absolute value and the squared returns are considered, there is evidence of long-memory behaviour. We analyze monthly data of the S&P 500 index from February 1973 to December 1997 (299 observations). The data source is Datastream.

In this case we have considered four different null hypotheses: $d_0 = 0.1$, 0.2, 0.3 and 0.4 against smaller integration orders. As it was mentioned in section 2.1, the statistics are normally distributed, since the process is stationary under the null hypothesis. For both the original series and the squared transformation a fractional white noise was fitted. The estimated values of d_1 together with the results of the FD-F test are provided in Table 9b. The Box-Ljunj Q-statistics calculated on the residuals after differencing the series for all orders of lag up to $\sqrt{T}$ yield no significant values, which implies that the null hypothesis of white noise residuals cannot be rejected. For instance, the $Q(18)$-statistic, asymptotically distributed as a χ^2_{17} (see Li and McLeod, 1986), yields values of 16.53 and 20.69 respectively, which are non rejection values at the nominal 5% level. On the contrary, for the absolute value transformation, the fractional white noise model does not seem to be appropriate since evidence of correlation was found in the errors. Table 9b also reports the estimation of d_1 in an ARFIMA$(1, d, 0)$ together with the corresponding AFD-F test results for the absolute value transformation. In this case, the Box-Ljunj test cannot reject the null hypothesis of white noise residuals since no significant values were found.

Our results are in line with those found by Lobato and Savin (1997). Long memory is rejected for the original stock returns but it cannot be rejected for the above mentioned transformations. As in previous studies a greater degree of memory is found in the absolute value rather than in the squared transformation. The original stock returns series and the squared transformation are well characterized by white noise and fractional white noise processes respectively. For the later, the estimation of d_1 produces a value of 0.133 and the FD-F cannot reject the null hypothesis of $d_0 = 0.2$ but it does reject $d_0 = 0.3$ ($d_0 = 0.1$ is not considered since in this case the alternative would be greater than the null hypothesis). The absolute value transformation, in turn, presents a bigger degree of fractional integration with an estimated value of d_1 equal to 0.199. Further, the AFD-F cannot reject either the null of $d_0 = 0.2$ or $d_0 = 0.3$ but rejects $d_0 = 0.4$. This last result is also in line with previous studies which find stronger evidence of long memory in the absolute value rather than in the squared transformation.

Table 9b. FD-F and AFD-F test on stock returns data.

S&P 500	$t_{H_0:d_0=0.1}$	$t_{H_0:d_0=0.2}$	$t_{H_0:d_0=0.3}$	$t_{H_0:d_0=0.4}$	$\hat{d}_1$
Original	-1.77*	-3.57*	-5.17*	-6.87*	-0.004
Abs. value	–	0.180	-1.32	-2.79*	0.199
Squared	–	-1.48	-3.21*	-4.86*	0.133

5.3. Macroeconomic data

Finally, we consider the extended version of Nelson and Plosser's data set. These series have been analyzed, among others, by Gil-Alaña and Robinson (1997) who by applying Robinson's LM test, find evidence of fractional integration in many of them.

In particular, we examine the unemployment rate, which is the series where there is much controversy about its degree of integration. The data is annual, starting in 1891 and ending in 1988 and, as in Nelson and Plosser (1982) it is been transformed to natural logarithms. Unlike most of the above considered examples, the fractional white noise hypothesis does not seem to be appropriate in this case. We consider different specifications for a_t in equation (5), namely that a_t is white noise with[7] and without constant in the model and that it is an AR(p) process. For the first two specifications, the null hypothesis of unit root cannot be rejected. Nevertheless, this result is hardly reliable since the residuals after differencing show significant autocorrelations, presenting evidence of misspecification in the model (the values for the $Q\,(10)$-statistic are 19.58 and 18.13 respectively, with are rejection values for a χ^2_9 distribution at the nominal 5% level). Table 9c also reports the values of the test when an AR(1) is fitted to a_t. None of the autocorrelations of the residuals

[7] As in the D-F framework, the asymptotic behavior changes when deterministic components are introduced in the regression equation. It is straightforward to check that when a constant is introduced in the model and the DGP is a random walk without drift, the limiting distributions correspond to the ones derived in Theorem 3, with the Brownian motions (BM) replaced by demeaned BM. Appendix 2 incorporates the critical values for the nonstandard range of values of d.

were significant in this case, making this specification preferable to the previous ones (the $Q(10)$ statistic is distributed in this case as a $\chi^2_{8,}$ and yields a value of 13.47) and therefore the null of white noise residuals cannot be rejected. Further, the test rejects the unit root null hypothesis, a result which coincides with that obtained by Gil-Alaña and Robinson (1997) and by Nelson and Plosser (1982) in the context of ARIMA models. The estimated value of d is in this case 0.412 and since it lies in the stationary range of values of d although close to the non-stationary frontier, there is interest in testing for the stationarity of the series. The second column in Table 9c reports the values of the test for the null hypothesis of $d_0 = 0.5$ (nonstationary) against smaller values of d (stationarity). In this case, the null of nonstationarity cannot be rejected.

Table 9c. FD-F and AFD-F test on unemployment rate.

	$t_{H_0:d_0=1}$	$t_{H_0:d_0=0.5}$	$\hat{d}_1$
a_t white noise	0.57	-	0.852
a_t white noise and constant	0.36	-	0.863
a_t AR(1)	-2.5*	0.63	0.412

6. CONCLUDING REMARKS

In this paper we have proposed Wald type tests of $H_0 : FI(d_0)$ versus $H_1 : FI(d_1)$, with $d_1 < d_0$, based upon the same principles as the Dickey-Fuller test for $I(1)$ versus $I(0)$ processes. The test statistics are based on the standardized OLS and the t-ratio in a very simple regression model. In particular when $d_0 = 1$ and $0 \leq d_1 < 1$, the regression model consists of regressing Δy_t on $\Delta^d y_{t-1}$ and, possibly, lags of Δy_t to correct for serial correlation. To implement the tests, a $\sqrt{T}$-consistent and asymptotically normal minimum distance estimator of d is proposed which turns out to work well in practice. Several empirical illustrations of how to use and interpret these tests are provided. Monte Carlo experiments support the analytical results and show the good finite sample properties of the proposed methods.

Further research is currently going towards extending the new methods in the same direction as the D-F test has been extended in the unit root literature: exogenous and endogenous structural breaks, heterokedastic error terms, multivariate time series, etc.

REFERENCES

Baillie, R.T., 1996, "Long memory processes and fractional integration in economics and finance", *Journal of Econometrics*, 73.

Beran, J., 1994, *Statistics for long memory processes*, New York: Chapman & Hall.

Brocwell, P. J. and R.A. Davies, 1993, *Time series: Theory and methods*, 2nd ed. (Springer-Verlag, New York ,NY).

Byers, D., J. Davidson and D.A. Peel, 1998, "The dynamics of aggregate political popularity: Evidence from eight countries", Unpublished manuscript.

Byers, D., J. Davidson and D.A. Peel, 1997, "Modelling political popularity: An analysis of long-range dependence in opinion polls series", *Journal of the Royal Statistical Society Series* A, 471-490.

Chung C.F. and Schmidt P., " The minimum distance estimator for fractionally integrated ARMA processes", Econometrics and Economic Theory Paper, No. 9408, Michigan State University.

Delgado M. A. and P. M. Robinson, 1994, "New Methods for the analysis of long memory time series: Application to Spanish inflation", *Journal of Forecasting*, 13, 97-107.

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

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

Diebold, F.X. and G.D. Rudebusch, 1991, "On the power of the Dickey-Fuller tests against fractional alternatives", *Economic Letters*, 35, 155-160.

Ding, Z., C.W.J. Granger and R.F. Engle, 1993, "A long memory property of stock returns and a new model", *Journal of Empirical Finance*, 1, 83-106.

Dolado J.J., J. Gonzalo and L. Mayoral, 1999, "Long Memory in Spanish political opinion polls series", Unpublished Manuscript.

Dolado J.J., and F. Mármol, 1999, "Asymptotic inference for nonstationary fractionally integrated processes", Unpublished Manuscript..

Fox, R. and M.S. Taqqu, 1986, "Large sample properties os parameter estimates for strongly dependent stationary Gaussian tiem series", *The Annals of Statistics*, 14, 517-532

Galbraith, J. W. and V. Zinde-Walsh, 1997, "Time domain methods for the estimation of fractionally-integrated time series models", Forthcoming in Biometrica..

Geweke, J. and S. Porter-Hudak, 1983, "The estimation and application of long memory time series models", *Journal of Time Series Analysis* 4, 221-238.

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

Hall, P., H. L. Koul and B.A. Turlauch, 1997, "Note on convergence rates of semiparametric estimators of dependence index", *The Annals of Statistics* Vol. 25, N^{o} 4., 1725-1739.

Helland, I. S., 1982, "Central limit theorems for martingales with discrete or continuous time", *Scandinavian Journal of Statistics* 9, 79-94.

Hosking, J.R.M., 1981, "Fractional differencing", *Biometrica* 68, 165-176.

Hosking, J.R.M., 1996, "Asymptotic distribution of the sample mean, autocovariances and autocorrelations of long memory time series", *Journal of Econometrics*, 73, 261-284.

Kramer, W., 1998, "Fractional Integration and the Augmented Dickey-Fuller test", *Economic Letters*, 61, 269-272.

Kwiatkowski, D., P.C.P. Phillips, P. Schmitd and Y. Shin, 1992, "Testing the null hypothesis of stationarity against the alternative of a unit root", *Journal of Econometrics* 54, 159-178.

Lee D. and P. Schmidt, 1996, "On the power of the KPSS test of stationarity against fractionally integrated alternatives", *Journal of Ecometrics* 73, 285-302.

Li, W.K and A.I. McLeod, 1986, "Fractional time series modelling", *Biometrica,* 73, 217-221.

Lo, A. W. ,1991, "Long-term memory in stock market prices", *Econometrica* 59, 1279-1313.

Lobato I. and N. E. Savin, 1998, "Real and spurious long memory properties of stock-market data", *Journal of Business &Economic Statistics*, Vol 186, 261-283.

Mármol, F, 1998, "Searching for fractional evidence using combined unit root tests", Working Paper 98-39, Universidad Carlos III de Madrid.

Nelson, C.R., and C.I. Plosser, 1982, "Trends and random walks in macroeconomic time series", *Journal of Monetary Economics* 10, 139-162.

Ng, S. and P. Perron, 1995, "Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag", *Journal of the American Statistical Association*, 90, 268-281.

Phillips P.C.B., 1987, "Time series regression with a unit root", *Econometrica* 55, 277-302.

Phillips P.C.B. and Z. Xiao ,1998, "A primer on unit root testing", *Journal of Economic Surveys*, Vol 12, N^o 5.

Phillips, P.C.B. and P. Perron, 1988, "Testing for unit roots in time series regression", *Biometrica*, 64, 381-412.

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

Robinson, P.M., 1994b, "Time series with strong dependence", in C. Sims (ed.), Advances in Econometrics, Sixth World Congress. Cambridge: Cambridge University Press.

Robinson, P.M., 1992, "Semiparametric analysis of long memory time series", *Annals of Statistics* 22, 515-539.

Said, S.E and D. A. Dickey , 1984, "Testing for unit roots in autoregressive moving average models of unknown order", *Biometrica*, 71, 599-608.

Sargan, J. D. and A. Barghava, 1983, "Maximum likelihood estimation of regression models with first order moving average errors when the root lies on the unit circle", *Econometrica* 51. 799-820.

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

Sowell, F.B., 1990, "The fractional unit root distribution", *Econometrica* 58, No 2, 495-505.

Tanaka, K., 1999, "The nonstationary fractional unit root". Forthcoming in *Econometric Theory*, No. 4, Vol. 15, 1999.

Tieslau, M., Schmidt P. and R. Baillie, 1996, "A minimum distance estimator for long-memory errors", *Journal of Econometrics* 71, 249-264.

Xiao, Z. and P.C.B. Phillips, 1998, "An ADF coefficient test for a unit root in ARMA models of unknown order with empirical applications to the US economy", *Econometrics Journal*, Vol. 1, 27-43.

APPENDIX 1

Proof of lemma 1.

The proof of the first and second part of this lemma is based on checking that $\{x_{t-1}\varepsilon_t\}$ verifies the conditions of the Central Limit Theorem (CLT) for martingale differences (MD) (see Helland, Theorem 2.5 page 82, 1982). These conditions are: i) the process should be a MD sequence, ii) the sum of the conditional variances should be equal to unity and iii) the Lindeberg condition (LC) should hold. Let us first consider the stationary case:

A. $0 < \delta < 0.5$.

Define:

$$\tilde{\varepsilon}_t = \sigma^{-1}\varepsilon_t,$$
$$\tilde{x}_t = \left(\sigma^2 \frac{\Gamma(1-2\delta)}{\Gamma^2(1-\delta)}\right)^{-1/2} x_t,$$
$$X_{T.t} = T^{-1/2}\tilde{x}_{t-1}\tilde{\varepsilon}_t.$$

Let $F_{T.t}$ be an array $\sigma - fields$ such that $F_{T.t-1} \subset F_{T.t}$. i) holds since:

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

With respect to the sum of the conditional variances:

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

$$T^{-1}\sum\left(E\left(\tilde{x}_{t-1}^2|F_{T.t-1}\right)E\left(\tilde{\varepsilon}_t^2|F_{T.t-1}\right)\right) = T^{-1}\sum\tilde{x}_{t-1}^2 \xrightarrow{p} 1. \qquad (A.3)$$

Finally, a sufficient condition for the LC to hold is:

$$\sum E\left(\left|X_{T.t}\right|^{2+\psi}|F_{T.t-1}\right) \xrightarrow{p} 0, \qquad (A.4)$$

where $\psi > 0$. Set $\psi = 2$. We obtain:

$$\sum E\left(\left|X_{T.t}\right|^4|F_{T.t-1}\right) = \frac{1}{T^2}\sum E\left(\left|\tilde{x}_{t-1}\tilde{\varepsilon}_t\right|^4|F_{T.t-1}\right) = \frac{1}{T^2}\sum\left(\tilde{x}_{t-1}\right)^4 E\left(\left|\tilde{\varepsilon}_t\right|^4|F_{T.t-1}\right) =$$

$$= \frac{1}{T}\frac{\sum\left(\tilde{x}_{t-1}\right)^4}{T}E\left(\tilde{\varepsilon}_t^4\right). \qquad (A.5)$$

Applying the law of large numbers (LLN) is easy to check that $\frac{\sum(\tilde{x}_{t-1})^4}{T} = O_p(1)$ provided $E\left(\tilde{\varepsilon}_t^4\right) < \infty$, which clearly implies that $(A.5)$ tends to zero and then $(A.4)$ holds.

Conditions $(A.1)$, $(A.3)$ and $(A.4)$ jointly imply the desired result.

B. $\delta = 0.5$.

Now, let us define:

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

$$\tilde{x}_t = \left((\log T)\sigma^2/\pi\right)^{-1/2} x_t$$

$$X_{T.t} = T^{-1/2}\tilde{x}_{t-1}\varepsilon_t.$$

Condition i) is clearly verified also in this case. With respect to condition ii):

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

Taking into account the results in Tanaka (1999), it is known that $T^{-1}\sum^T \tilde{x}_{t-1}^2 \xrightarrow{p} 1$.

Finally, in order to check that the iii) is verified, we prove that $(A.4)$ holds. For that obtain:

$$\frac{\sum\left(\tilde{x}_{t-1}\right)^4}{T} = \left(\pi/(\log T)\sigma^2\right)^2 \frac{\sum^T \pi_i^4(-\delta)\,\varepsilon_{t-i}^4 + \sum_i^T\sum_j^T \pi_i^2(-\delta)\,\pi_j^2(-\delta)\,\varepsilon_{t-i}^2\varepsilon_{t-j}^2}{T} + o_p(1),$$

and since $\frac{\sum \pi_i^2(-\delta)}{\log T} \to 1/\pi$ and $\sum \pi_i^4(-\delta) = O(1)$, it follows that $\sum\left(\tilde{x}_{t-1}\right)^4/T = O_p(1)$. Therefore,

$$E\left(\tilde{\varepsilon}_t^4\right)\frac{\sum\left(\tilde{x}_{t-1}\right)^4}{T} = O_p(1),$$

which clearly implies that $(A.5)$ tends to zero and then $(A.4)$ holds.

For the proof of the last part of the lemma, see Dolado and Mármol (1999).□

Proof of Theorem 1.

Under the null hypothesis, $\Delta^{d_1}y_t \sim FI(1-d_1)$. Therefore, for values of $d_1 \in [0, 0.5]$ will be nonstationary and will be stationary when $d_1 \in (0.5, 1]$. Applying the LLN, Lemma 1 and results in Dolado and Mármol (1999), we obtain:

$$T^{1-d_1}\hat{\phi}_{ols} = \frac{T^{-(1-d_1)}}{T^{-2(1-d_1)}}\frac{\sum \varepsilon_t\Delta^{d_1}y_{t-1}}{\sum\left(\Delta^{d_1}y_{t-1}\right)^2} \xrightarrow{w} \frac{\int_0^1 B_{1-d_1}(r)\,dB}{\int_0^1 B_{1-d_1}^2(r)\,dr} \qquad \text{if } 0 \le d_1 < 0.5,$$

$$T^{1/2}(\log T)^{1/2}\hat{\phi}_{ols} = \frac{T^{-(1/2)}\sum \varepsilon_t \Delta^{d_1} y_{t-1}}{T^{-1}\sum \Delta^{d_1} y_{t-1}^2} \xrightarrow{w} N(0,\pi) \qquad \text{if } d_1 = 0.5,$$

$$T^{1/2}\hat{\phi}_{ols} = \frac{T^{-(1/2)}\sum \varepsilon_t \Delta^{d_1} y_{t-1}}{T^{-1}\sum \Delta^{d_1} y_{t-1}^2} \xrightarrow{w}$$

$$\frac{N\left(0, \sigma^4 \frac{\Gamma(2d_1-1)}{\Gamma^2(d_1)}\right)}{\sigma^2 \Gamma(2d_1 - 1)/\Gamma^2(d_1)} = N\left(0, \frac{\Gamma^2(d_1)}{\Gamma(2d_1-1)}\right) \qquad \text{if } 0.5 < d_1 < 1. \square$$

Proof of Theorem 2.

The proof of this theorem is analogous to the previous one and therefore is omitted.

Proof of Theorem 3.

Under the DGP (14) in the main text, the process $\Delta y_t \sim FI(d_1^* - 1)$ and $\Delta^{d_1} y_{t-1} \sim I(d_1^* - d_1)$. According to the sign of the difference of $d_1^* - d_1$, the process $\Delta^{d_1} y_{t-1}$ can be either stationary with intermediate memory ($d_1^* - d_1 < 0$), stationary with long memory ($0 < d_1^* - d_1 < 0.5$) or non stationary $(d_1^* - d_1 \geq 0.5)$. Therefore, the asymptotic behavior will depend upon the value of d_1. Let us consider separately the three following cases:

1. $d_1^* - d_1 < 0$.

   In this case both Δy_t and $\Delta^{d_1} y_{t-1}$ are stationary processes. The denominator of the t-ratio is $O_p\left(T^{1/2}\right)$ since, by the LLN, $S_T = O_p(1)$ and $\left(\sum \left(\Delta^{d_1} y_{t-1}\right)^2\right)^{1/2} = O_p\left(T^{1/2}\right)$. On the other hand, the numerator:

$$\frac{\sum \Delta y_t \Delta^{d_1} y_{t-1}}{T} \xrightarrow{p} cov\left(\Delta y_t, \Delta^{d_1} y_{t-1}\right). \qquad (A.6)$$

   Then, it is an $O_p(T)$ and therefore, the t-statistic diverges. It remains to be shown that the limit is -∞. The denominator is always positive and therefore the sign of the ratio is determined by the numerator.

$$cov\left(\Delta y_t, \Delta^{d_1} y_{t-1}\right) = cov\left(\Delta^{1-d_1^*}\varepsilon_t, \Delta^{d_1-d_1^*}\varepsilon_{t-1}\right) = \qquad (A.7)$$

$$= cov(\varepsilon_t + \pi_1\varepsilon_{t-1} + ..., \varepsilon_{t-1} + \pi_1(d_1 - d_1^*)\varepsilon_{t-2} + (.A.8)$$

$$= \sigma^2\left(\pi_1 + \pi_2\pi_1(d_1 - d_1^*) + \pi_3\pi_2(d_1 - d_1^*) + ...\right). \qquad (A.9)$$

All of these coefficients are negative, and:

$$\begin{aligned}\pi_2\pi_1(d_1-d_1^*)+\pi_3\pi_2(d_1-d_1^*)+... &\leq \\ \sup_{j\in[2,\infty)}(-\pi_j)\sum_{i=1}^{\infty}-\pi_i(d_1-d_1^*) &= -\pi_2*(1)<-\pi_1,\end{aligned}$$

then $(A.6)<0$ and the test diverges to $-\infty$.

2. $0<d_1^*-d_1<0.5$. In this case, $\Delta^{d_1}y_{t-1}$ is a stationary long memory process. The denominator of the t-statistic exhibits the same behavior as in the previous case. As $d_1-d_1^*<0$, the coefficients $\pi_i(d_1-d_1^*)$ in $(A.9)$ are positive, so it is immediate to see that the covariance in $(A.6)$ is negative.

3. $d_1^*-d_1>0.5$.

   We can write the t-ratio as:

$$t_{\hat{\phi}_{ols}}=\frac{\sum\Delta y_t\Delta^{d_1}y_{t-1}}{\left(\left(\sum(\Delta y_{t-1})^2\sum(\Delta^{d_1}y_{t-1})^2-\left(\sum\Delta y_t\Delta^{d_1}y_{t-1}\right)^2\right)/T\right)^{1/2}}.$$

   Then:

$$\sum(\Delta y_{t-1})^2=O_p(T); \qquad (A.10)$$

$$\sum\left(\Delta^d y_{t-1}\right)^2=O_p\left(T^{2(d_1-d)}\right); \qquad (A.11)$$

$$\sum\Delta y_t\Delta^d y_{t-1}=O_p(T); \qquad (A.12)$$

While $(A.10)$ and $(A.11)$ are immediate. $(A.12)$ follows from:

$$\sum\Delta y_t\Delta^{d_1}y_{t-1}=\sum\Delta y_t\left(y_{t-1}+\pi_1(d_1)y_{t-2+}\pi_2(d_1)y_{t-3}+...\right),$$

and substituting recursively we obtain:

$$\sum\Delta y_t\Delta^{d_1}y_{t-1}=\left(\sum_{i=0}^{T}\pi_i(d_1)\right)\left(\sum_{t=2}^{T}\Delta y_t y_{t-1}\right) \qquad (A.13)$$

$$-\left(\sum_{i=1}^{T}\pi_i(d_1)\right)\left(\sum_{t=2}^{T}\Delta y_t\Delta y_{t-1}\right)-\left(\sum_{i=2}^{T}\pi_i(d)\right)\left(\sum_{t=3}^{T}\Delta y_t\Delta y_{t-2}\right)+... \qquad (A.14)$$

The term $\sum_{t=2}^{T} \Delta y_t y_{t-1}$ is $O_p(T)$ (see Sowell, 1990) and since $\sum_{i=0}^{T} \pi_i(d_1) \rightarrow 0$, for any $d_1 > 0$, the first term in $(A.13)$ is $o_p(T)$. The remaining terms are sums that verify the LLN so that, divided by T, converge to the autocovariance function of Δy_t; since Δy_t is a stationary short memory process, the sum of its autocovariances is finite. Then, since the denominator is a $O_p\left(T^{d_1^* - d_1}\right)$, (with $d_1^* - d_1 < 1$), and the numerator is an $O_p(T)$, the statistic diverges To see that it diverges to $-\infty$, consider:

$$\frac{\sum \Delta y_t \Delta^{d_1} y_{t-1}}{T} \xrightarrow{p} -cov(\Delta y_t, \Delta y_{t-1}) \sum_{i=1}^{\infty} \pi_i(d_1) - cov(\Delta y_t, \Delta y_{t-2}) \sum_{i=2}^{\infty} \pi_i(d_1) + ...$$

This is a negative quantity because all the covariances $cov(\Delta y_t, \Delta y_{t-i})$ are negative (see, for instance, Hoskings, 1981) and all the sums $\sum_{i=j}^{\infty} \pi_i(d_1)$, $j \geq 1$, are also negative for $d_1 > 0$. Then, the statistic diverges to $-\infty$ and the test is consistent.□

We present the proofs of Theorems 4 and 5 corresponding to the estimator based on $\phi(d)$ defined in (15). The same proofs are valid with slight modifications for the estimator defined in (16). Before presenting them, we need some previous lemmas.

Lemma 2 *Let $V_{T_\phi}(d)$ be defined as in* (17)*, then:*

1. *$V_{T_\phi}(d)$ is a twice continuously differentiable function.*

2. *Further, it verifies:*

$$\sqrt{T} \frac{\partial V_{T_\phi}\left(d_1^* + \delta/\sqrt{T}\right)}{\partial d} \xrightarrow{w} Z^*$$

where:

$$Z^* = \left(\left(1 + 2\pi_2 + \sum_{i=2}^{\infty} \pi_i^2\right)^{1/2} Z - \delta\left(1 - \sum_{i=2}^{\infty} \frac{\pi_i}{i-1}\right)\right)\left(1 - \sum_{i=2}^{\infty} \frac{\pi_i}{i-1}\right)$$

and Z is a $N(0,1)$.

3. *And:*

$$\sqrt{T} \frac{\partial V_{T_\phi}(d_1^*)}{\partial d} \xrightarrow{w} Z^{**}$$

where:

$$Z^{**} = \left(1 + 2\pi_2 + \sum_{i=2}^{\infty} \pi_i^2\right)^{1/2} \left(1 - \sum_{i=2}^{\infty} \frac{\pi_i}{i-1}\right) Z.$$

Proof of Lemma 2.

1. As the coefficients $\pi_i(.)$ are polynomials on d, (recall expression (1)), and $V_{T_\phi}(d)$ is a continuous composition of these coefficients, $V_{T_\phi}(d)$ is a C(∞) function.

2. Consider the first derivative of $V_{T_\phi}(d)$ evaluated in $d = d_1^* + \delta/\sqrt{T}$:

$$\sqrt{T}\left.\frac{\partial V_{T_\phi}(d)}{\partial d}\right|_{d=d_1^*+\delta/\sqrt{T}} = 2\sqrt{T}\left(\phi(d) - \pi_1(1-d)\right)H(d)\Big|_{d=d_1^*+\delta/\sqrt{T}}$$

where $H(d)|_{d=d_1^*+\delta/\sqrt{T}} = \partial\left(\phi(d) - \pi_1(1-d)\right)/\partial d|_{d=d_1^*+\delta/\sqrt{T}}$. The first term of the above expression is given by:

$$\sqrt{T}\left(\left(\phi(d) - \pi_1(1-d)\right)|_{d=d_1^*+\delta/\sqrt{T}}\right) \quad (A.15)$$

$$= \sqrt{T}\left(\frac{\sum \Delta y_t \Delta^{\delta/\sqrt{T}}\varepsilon_{t-1}}{\sum\left(\Delta^{\delta/\sqrt{T}}\varepsilon_{t-1}\right)^2} - \left(1 - d_1^* - \delta/\sqrt{T}\right)\right) \quad (A.16)$$

Since $\Delta^{\delta/\sqrt{T}}\varepsilon_{t-1}$ can be written as (see Tanaka, 1999):

$$\Delta^{\delta/\sqrt{T}}\varepsilon_{t-1} = \varepsilon_{t-1} - \frac{\delta}{\sqrt{T}}\sum_{k=1}^{t-2}\frac{1}{k}\varepsilon_{t-1-k} + O_p\left(\frac{1}{T}\right), \quad (A.17)$$

it follows that:

$$\sum\left(\Delta^{\delta/\sqrt{T}}\varepsilon_{t-1}\right)^2 \xrightarrow{p} \sigma^2, \quad (A.18)$$

and consequently, $(A.16)$ can be rewritten as:

$$\sqrt{T}\left(\sum_{i=0}^{T}\pi_i\hat{\rho}_\varepsilon(i-1) - \pi_1\right) + \delta\left(1 - \frac{\sum_{k=1}^{T} 1/k \sum_{t=k+2}^{T-1}\Delta y_t\varepsilon_{t-1-k}}{\sum\left(\varepsilon_{t-1}\right)^2}\right) + o_p(1).$$

where $\hat{\rho}_\varepsilon(.)$ is the estimated autocorrelation function of the $\varepsilon_t' s$. It is known (Brockwell and Davies, 1993, p.232) that the joint distribution of $\sqrt{T}\hat{\rho}_\varepsilon(1)$, ..., $\sqrt{T}\hat{\rho}_\varepsilon(m)$ with m fixed tends to $\mathrm{N}_m(0, I_m)$. Then, it follows:

$$\sqrt{T}\left(\sum_{i=0}^{T}\pi_i\hat{\rho}_\varepsilon(i-1) - \pi_1\right) \xrightarrow{w} N\left(0, 1 + \sum_{i=2}^{\infty}\pi_i^2 + 2\pi_2\right)$$

and,

$$\delta\left(1 - \frac{\sum_{k=1}^{T} 1/k \sum_{t=k+2}^{T-1}\Delta y_t\varepsilon_{t-1-k}}{\sum\left(\varepsilon_{t-1}\right)^2}\right) \xrightarrow{p} \delta\left(1 - \sum_{i=2}^{\infty}\frac{\pi_i}{i-1}\right).$$

With respect to the second term in expression $(A.16)$:

$$H(d)|_{d=d_1^*+\delta/\sqrt{T}} = \frac{\sum \Delta y_t \left(\Delta^d y_{t-1} \log(1-L)\right) \sum \left(\Delta^d y_{t-1}\right)^2}{\left(\sum \left(\Delta^d y_{t-1}\right)^2\right)^2} - \frac{\left(\sum \Delta y_t \Delta^d y_{t-1}\right) \sum \Delta^d y_{t-1} (\log(1-L) \Delta^d y_{t-1})}{\left(\sum \left(\Delta^d y_{t-1}\right)^2\right)^2} + 1 \Bigg|_{d=d_1+\delta/\sqrt{T}}$$

and, applying $(A.17)$ and $(A.18)$, and taking into account the expansion $\log(1-L) = -(L + L^2/2 + L^3/3...)$ it is easy to check that

$$H(d)|_{d=d_1+\delta/\sqrt{T}} \xrightarrow{p} 1 - \sum_{i=2}^{\infty} \frac{\pi_i}{i-1}. \qquad (A.19)$$

3. This last statement can be proved similarly.

Lemma 3 $V_{T_\phi}\left(d_1^* + \delta/\sqrt{T}\right)$ *is asymptotically strictly convex and its second derivative with respect to* d *verifies:*

$$\frac{\partial^2 V_{T_\phi}\left(d_1^* + \delta/\sqrt{T}\right)}{\partial d^2} = 2H\left(d_1^* + \delta/\sqrt{T}\right)^2 + o_p(1),$$

and:

$$\frac{\partial^2 V_{T_\phi}\left(d_1^* + \delta/\sqrt{T}\right)}{\partial d^2} \xrightarrow{p} 2\left(1 - \sum_{i=2}^{\infty} \frac{\pi_i}{i-1}\right)^2.$$

Proof of Lemma 3.

1. Consider:

$$\frac{\partial^2 V_{T_\phi}\left(d_1^* + \delta/\sqrt{T}\right)}{\partial d^2} = 2H(d)^2 + 2\frac{\partial H}{\partial d}\left(\phi(d) - \pi_1(1-d)\right)\Bigg|_{d=d_1^*+\delta/\sqrt{T}}$$

Lemma 2 implies that $\left(\phi(d) - \pi_1(1-d)\right)|_{d=d_1^*+\delta/\sqrt{T}}$ is $o_p(1)$, then we only need to show that $\frac{\partial H}{\partial d}\big|_{d=d_1^*+\delta/\sqrt{T}} = O_p(1)$. By proceeding in the same way as in Lemma 2, it is easy to check that both numerator and denominator are $O_p(T^4)$, which implies the desired result.

2. It is immediate to check that:

$$H\left(d_1^* + \delta/\sqrt{T}\right) - H\left(d_1^*\right) \xrightarrow{p} 0$$

and therefore, expression $(A.19)$ ensures the result.

Proof of Theorem 4.

We now prove that there exist a MDE $\hat{d}_{T_\phi}$ of d_1 such that, $\hat{\delta} = \sqrt{T}\left(\hat{d}_{T_\phi} - d_1\right) = O_p(1)$. Along the arguments in Tanaka (1999), we first show the existence of a local MDE $\tilde{d}$ such that $\sqrt{T}\left(\tilde{d}_{T_\phi} - d_1^*\right) = \tilde{\delta} = O_p(1)$. For this it is sufficient to show that, for any $\epsilon > 0$ and for all $T > T_0$ with T_0 fixed, there exist a positive constant δ_0 such that:

$$P\left(\left|\tilde{d}_{T_\phi} - d_1^*\right| \leq \delta_0/\sqrt{T}\right) \geq 1 - \epsilon. \qquad (A.20)$$

From Sargan and Bhargava (1983), $(A.20)$ can be proved by showing:

$$P\left(\partial V_{T_\phi}\left(d_1^* + \delta_1/\sqrt{T}\right)/\partial d \leq 0\right) \leq \epsilon$$

and:

$$P\left(\partial V_{T_\phi}\left(d_1^* + \delta_2/\sqrt{T}\right)/\partial d \geq 0\right) \leq \epsilon$$

for all $T \geq T_0$, by taking suitably $\delta_1 (> 0), \delta_2 (< 0)$ and a corresponding $T_{0.}$ Therefore, as in Sargan and Bhargava (1983), we consider a function positively related to $\partial V_{T_\phi}\left(d_1^* + \delta_1/\sqrt{T}\right)/\partial d$, which is $\sqrt{T}\partial V_{T_\phi}\left(d_1 + \delta_1/\sqrt{T}\right)/\partial d$. It follows from Lemma 2 and Chevichev's inequality that:

$$\begin{aligned} P\left(T^{1/2}\partial V_{T_\phi}\left(d_1^* + \delta/\sqrt{T}\right)/\partial d \leq 0\right) &\rightarrow P\left(-Z^* + E\left(Z^*\right) > E\left(Z^*\right)\right) \qquad (A.21) \\ &\leq \frac{var\left(Z^*\right)}{E\left(Z^*\right)^2} = \frac{\left(1 + \sum_2^\infty \pi_i^2 + 2\pi_2\right)}{\delta^2} \qquad (A.22) \end{aligned}$$

and:

$$\begin{aligned} P\left(T^{1/2}\partial V_{T_\phi}\left(d_1^* + \delta/\sqrt{T}\right)/\partial d \geq 0\right) &\rightarrow P\left(Z^* - E\left(Z^*\right) > -E\left(Z^*\right)\right) \qquad (A.23) \\ &\leq \frac{var\left(Z^*\right)}{E\left(Z^*\right)^2} = \frac{\left(1 + \sum_2^\infty \pi_i^2 + 2\pi_2\right)}{\delta^2} \qquad (A.24) \end{aligned}$$

Therefore, it is ensured that the probabilities in $(A.21)$ and $(A.23)$ can be made smaller than ϵ by taking $\delta_0 = \max\left(\delta_1, |\delta_2|, \sqrt{\frac{\left(1+\sum_2^\infty \pi_i^2+2\pi_2\right)}{\epsilon}}\right)$. Then, T_0 can be chosen such that $(A.20)$ holds. Following Tanaka (1999), since $V_{T\phi}\left(d_1^* + \delta/\sqrt{T}\right)$ is asymptotically strictly convex (recall Lemma 3) we can assert that a local minimizer $\tilde{d}$ is asymptotically $\hat{d}$, the MDE of d_1.

Proof of Theorem 5.

As $\hat{d}$ is an interior point of D, it solves the first order conditions so that:

$$\partial V_{T_\phi}\left(\hat{d}\right)/\partial d = 0$$

By a Taylor expansion we have:

$$\partial V_{T_\phi}\left(\hat{d}\right)/\partial d = \partial V_{T_\phi}\left(d_1^*\right)/\partial d + \partial^2 V_{T_\phi}\left(\breve{d}\right)/\partial d^2 (\hat{d} - d_1^*) = 0$$

where $\breve{d}$ lies between $\hat{d}$ and d_1^*. Solving for $(\hat{d} - d_1^*)$ gives:

$$(\hat{d} - d_1^*) = -\left(\frac{\partial^2 V_{T_\phi}\left(\breve{d}\right)}{\partial d^2}\right)^{-1} \frac{\partial V_{T_\phi}\left(d_1^*\right)}{\partial d} \qquad (A.25)$$

As $\hat{d}$ is consistent, $p\lim \breve{d} = d_1^*$, then:

$$\frac{\partial^2 V_{T_\phi}\left(\breve{d}\right)}{\partial d^2} - \frac{\partial^2 V_{T_\phi}\left(d_1^*\right)}{\partial d^2} \xrightarrow{p} 0$$

and therefore,

$$(\hat{d} - d_1^*) = -\left(2H\left(d_1^*\right)^2\right)^{-1} 2H\left(d_1^*\right)\left(\phi\left(d_1^*\right) - \pi_1\right) + o_p(1)$$

Therefore, using Lemma 2:

$$\sqrt{T}(\hat{d} - d_1^*) = -\left(H\left(d_1^*\right)\right)^{-1}\sqrt{T}\left(\phi\left(d_1^*\right) - \pi_1\right) \xrightarrow{w} N\left(0, \frac{\left(1+\sum_{i=2}^\infty \pi_i^2 + 2\pi_2\right)}{\left(1-\sum_{i=2}^\infty \pi_i/(i-1)\right)^2}\right).$$

Proof of Theorem 6.

Let $\hat{\Phi}_{ols}$ be the OLS estimator of the parameters $\Phi = (\alpha_1 ... \alpha_p\ \phi)'$ in equation (22) :

$$\Upsilon_T\left(\hat{\Phi}_{ols} - \Phi\right) = \left(\Upsilon_T^{-1} Y' Y \Upsilon_T^{-1}\right)^{-1} Y' \xi_t \qquad (A.26)$$

where Y is the matrix of regressors:

$$Y = \begin{pmatrix} \Delta y_{t-1} & ... & \Delta y_{t-p} & \Delta^{d_1} y_{t-1} \end{pmatrix}$$

and Υ_T is a weighting matrix as defined below. In order to study its asymptotic behavior, we have to consider three different situations according to the value of d_1 used in the regression, which determines the stationarity or non-stationarity nature of the process $\{\Delta^{d_1} y_t\}$.

1. $0 < d_1 < 0.5$.

   Define the weighting matrix:

$$\Upsilon_T = \begin{pmatrix} T^{1/2} I_p & 0 \\ 0 & T^{1-d_1} \end{pmatrix}$$

   where I_p is the $p \times p$ identity matrix. Applying the LLN and results in Dolado and Mármol. (1999) is easy to check that the first factor $\left(\Upsilon_T^{-1} Y'Y \Upsilon_T^{-1}\right)$ converges to:

$$\begin{pmatrix} \gamma_0 & \gamma_1 & ... & 0 \\ \gamma_1 & ... & & ... \\ ... & & \gamma_0 & 0 \\ 0 & ... & 0 & \int_0^1 B_{1-d_1}^{*2}(r)\, dr \end{pmatrix} = \begin{pmatrix} V & 0 \\ 0 & Q \end{pmatrix}$$

   where γ_j is the autocovariance function of Δy_t. With respect to the second term, $\Upsilon_T^{-1} Y' \xi_t$, recall that under the null hypotheses $\xi_t = \varepsilon_t$. Therefore, the first p elements satisfy the usual CLT for martingale difference sequences with a variance-covariance matrix given by:

$$E \begin{pmatrix} \Delta y_{t-1}\varepsilon_t \\ ... \\ \Delta y_{t-p}\varepsilon_t \end{pmatrix} \begin{pmatrix} \Delta y_{t-1}\varepsilon_t & ... & \Delta y_{t-p}\varepsilon_t \end{pmatrix} = \sigma^2 \begin{pmatrix} \gamma_0 & \gamma_1 & ... & \gamma_{p-1} \\ \gamma_1 & ... & & ... \\ ... & & \gamma_0 & ... \\ \gamma_{p-1} & ... & ... & \gamma_0 \end{pmatrix} = \sigma^2 V$$

Thus:

$$\begin{pmatrix} T^{-1/2}\sum \Delta y_{t-1}\varepsilon_t \\ ... \\ T^{-1/2}\sum \Delta y_{t-p}\varepsilon_t \end{pmatrix} \xrightarrow{w} h_1 = N_p\left(0, \sigma^2 V\right)$$

The last element converges to:

$$T^{-(1-d_1)}\sum \Delta^{d_1} y_{t-1}\varepsilon_t \xrightarrow{w} h_2 = \int_0^1 B^*_{1-d_1}(r)\, dB$$

Therefore:

$$\Upsilon_T\left(\hat{\Phi}_{ols} - \Phi\right) \xrightarrow{w} \begin{pmatrix} V^{-1}h_1 \\ Q^{-1}h_2 \end{pmatrix}$$

The t-statistic associated to $\hat{\phi}_{ols}$ is given by:

$$t_{\hat{\phi}_{ols}} = \frac{\hat{\phi}_{ols}}{\hat{\sigma}_{\hat{\phi}_{ols}}}$$

as $\hat{\Phi}_{ols}$ is consistent, $\hat{\sigma}_{\hat{\phi}_{ols}} \xrightarrow{p} \sigma Q^{-1/2}$, which implies the desired result.

2. $0.5 < d_1 < 1$.

In this case, $\Delta^{d_1} y_t$ is a stationary process. Consider again expression $(A.26)$ where now $\Upsilon_T = T^{1/2} I_{p+1}$ and consider the limit of $\left(\Upsilon_T^{-1} Y' Y \Upsilon_T^{-1}\right)$. In this case the LLN apply to all the elements yielding:

$$\Upsilon_T^{-1} Y' Y \Upsilon_T^{-1} \xrightarrow{p} \begin{pmatrix} \gamma_0 & \gamma_1 & ... & cov(\Delta y_{t-1}, \Delta^{d_1} y_{t-1}) \\ \gamma_1 & ... & & ... \\ ... & & .. & cov(\Delta y_{t-p}, \Delta^{d_1} y_{t-1}) \\ cov(\Delta y_{t-1}, \Delta^{d_1} y_{t-1}) & ... & cov(\Delta y_{t-p}, \Delta^{d_1} y_{t-1}) & var\left(\Delta^{d_1} y_{t-1}\right) \end{pmatrix}.$$

$$= \begin{pmatrix} Q_1 & Q_2 \\ Q_2' & Q_4 \end{pmatrix}.$$

With respect to the second factor, applying the CLT and Lemma 1:

$$\begin{pmatrix} T^{-1/2}\sum \Delta y_{t-1}\varepsilon_t \\ ... \\ T^{-1/2}\sum \Delta^{d} y_{t-1}\varepsilon_t \end{pmatrix} \xrightarrow{w} Z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N_{p+1}\left(0, \begin{matrix} \sigma^2 V & \sigma^2 Q_2 \\ \sigma^2 Q_2' & \sigma^2 var\left(\Delta^{d_1} y_{t-1}\right) \end{matrix}\right).$$

Thus:

$$\Upsilon_T \left(\hat{\Phi}_{ols} - \Phi \right) \xrightarrow{w} \begin{pmatrix} Q_1 & Q_2 \\ Q_2' & Q_4 \end{pmatrix}^{-1} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

Let's call $\begin{pmatrix} a_1 & a_2 \\ a_2 & a_4 \end{pmatrix} = \begin{pmatrix} Q_1 & Q_2 \\ Q_2' & Q_4 \end{pmatrix}^{-1}$; Then,

$$T^{1/2} \left(\hat{\phi}_{ols} - \phi \right) \xrightarrow{w} a_2 z_1 + a_4 z_2.$$

For simplicity of the presentation, we consider the case where $\alpha(L)$ is an AR(1) process. We obtain:

$$a_2 = \frac{-cov\left(\Delta y_{t-1}, \Delta^{d_1} y_{t-1}\right)}{var\left(\Delta y_{t-1}\right) var\left(\Delta^{d_1} y_{t-1}\right) - cov^2\left(\Delta y_{t-1}, \Delta^{d_1} y_{t-1}\right)},$$

$$a_4 = \frac{var\left(\Delta y_{t-1}\right)}{var\left(\Delta y_{t-1}\right) var\left(\Delta^{d_1} y_{t-1}\right) - cov^2\left(\Delta y_{t-1}, \Delta^{d_1} y_{t-1}\right)}.$$

Then, we easily obtain:

$$T^{1/2} \left(\hat{\phi}_{ols} - \phi \right) \xrightarrow{w} N\left(0, \sigma^2 a_4\right)$$

Again, as $\hat{\Phi}_{ols}$ is consistent, so is S_T for σ, therefore the t-statistic converges to:

$$t_{\hat{\phi}_{ols}} \xrightarrow{w} N\left(0, \sigma^2 a_4\right) / \sigma a_4^{1/2} = N\left(0, 1\right).$$

The general case can be proved similarly.

3. $d_1 = 0.5$.

   Define the weighting matrix:

$$\Upsilon_T = \begin{pmatrix} T^{1/2} I_p & 0 \\ 0 & \left(T \log T\right)^{1/2} \end{pmatrix}$$

   The rest of the proof is analogous to the previous case and therefore is omitted.

Proof of Theorem 7.

Since u_t is invertible, the AR representation of u_t is given by:

$$\vartheta(L) u_t = \sum_{i=0}^{\infty} \vartheta_i u_{t-i} = \varepsilon_t \qquad \text{with } \vartheta_0 = 1.$$

Recall now expression (22). y_t can be rewritten as:

$$\Delta y_t = \phi \Delta^{d_1^*} y_{t-1} + \delta_1 \Delta y_{t-1} + ... + \delta_k \Delta y_{t-k} + \eta_{kt},$$

where $\phi = -(1-\rho)^{1-d_1^*} \vartheta(1)$, $\eta_{tk} = \varepsilon_t + \sum_{j=k+1}^{\infty} \delta_j \Delta y_{t-j}$. Let us call: $Y_k = \begin{pmatrix} \Delta y_{t-1} & ... & \Delta y_{t-k} & \Delta^{d_1^*} y_{t-1} \end{pmatrix}$ and $\Phi_k = \begin{pmatrix} \delta_1 & ... & \delta_k & \phi \end{pmatrix}$. It's immediate to check that $\hat{\Phi}_k = \Phi_k + (Y_k' Y_k)^{-1} Y_k \eta_{kt}$ is $O_p(1)$ (remark that all the process involved are stationary). On the other hand, the variance of the OLS estimator is given as usual by $\hat{\sigma}^2 (Y_k' Y_k)^{-1}$ which is a matrix with elements that are $O_p(1/T)$, since $\hat{\sigma}^2$ is $O_p(1)$ and $(Y_k' Y_k)$ is $O_p(T)$. Therefore, the t-statistic diverges at a $\sqrt{T}-$rate. Now, notice that when $k \to \infty$, $\eta_{kt} \xrightarrow{p} \varepsilon_t$, and this implies that $(Y_k' Y_k)^{-1} Y_k \eta_{kt} \xrightarrow{p} 0$, since Y_k and ε_t are uncorrelated, and consequently $\hat{\Phi}_k \xrightarrow{p} \Phi_k$; Note that $\phi < 0$ since $\vartheta(1)$ is positive (with is implied by the fact that ϑ_0 is positive and all the roots of $\vartheta(L)$ lie outside the unit circle) and $|\rho| < 1$, therefore $\phi = -(1-\rho)^{1-d_1^*} \vartheta(1) < 0$. Then, since $\hat{\phi}_{ols}$ is consistent and $\phi < 0$, the t-statistic diverges to $-\infty$ at a $\sqrt{T}$ rate, which ensures the consistency of the test.

APPENDIX 2

In this appendix, we report the critical values corresponding to the range of values of d for which the FD-F does not have a standard distribution under the null. Table 11a presents the critical values for the case where the DGP is a random walk and the t-statistic is computed in equation (5) whereas Table 11b presents the corresponding values for the case where the DGP is the same as in the previous case but a constant is included in regression (5). To compute these critical values, we have generated random walk processes from i.i.d. N(0,1) disturbances and computed the t-statistic associated to ϕ in the corresponding regression. In the following table d is the value used to compute the t-statistic. The number of replications is 10.000.

Table 11a. **Critical Values.**									
DGP: $\Delta y_t = \varepsilon_t$. **Regression:** $\Delta y_t = \phi \Delta^{d_1} y_{t-1}$									
T	$T = 100$			$T = 400$			$T = 1000$		
d_1 / Sig. level	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.0	-1.61	-1.95	-2.6	-1.62	-1.95	-2.6	-1.62	-1.95	-2.6
0.05	-1.59	-1.93	-2.57	-1.61	-1.92	-2.57	-1.59	-1.91	-2.56
0.10	-1.57	-1.87	-2.56	-1.57	-1.9	-2.56	-1.57	-1.90	-2.55
0.15	-1.56	-1.86	-2.55	-1.56	-1.87	-2.52	-1.55	-1.87	-2.50
0.20	-1.51	-1.84	-2.53	-1.53	-1.84	-2.49	-1.53	-1.84	-2.45
0.25	-1.50	-1.83	-2.5	-1.49	-1.83	-2.44	-1.47	-1.82	-2.43
0.30	-1.46	-1.82	-2.49	-1.46	-1.81	-2.43	-1.45	-1.80	-2.43
0.35	-1.45	-1.82	-2.49	-1.44	-1.8	-2.42	-1.44	-1.79	-2.42
0.40	-1.43	-1.81	-2.47	-1.38	-1.79	-2.41	-1.42	-1.75	-2.41
0.45	-1.40	1.80	-2.46	-1.36	-1.75	-2.40	-1.35	-1.71	-2.39
0.50	1.36	-1.75	2.40	-1.33	-1.66	-2.36	-1.33	-1.66	-2.35

Table 11b. **Critical Values. Constant in the model.**									
DGP: $\Delta y_t = \varepsilon_t$; **Regression:** $\Delta y_t = c + \phi\Delta^{d_1} y_{t-1}$									
T	$T = 100$			$T = 400$			$T = 1000$		
d_1 / Sig. level	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.0	-2.58	-3.17	-3.51	-2.57	-2.87	-3.44	-2.57	-2.86	-3.44
0.05	-2.50	-2.88	-3.50	-2.48	-2.77	-3.37	-2.49	-2.82	-3.33
0.10	-2.49	-2.87	-3.49	-2.45	-2.76	-3.30	-2.47	-2.79	-3.31
0.15	-2.44	-2.80	-3.35	-2.41	-2.74	-3.25	-2.39	-2.76	-3.29
0.20	-2.36	-2.67	-3.31	-2.31	-2.57	-3.20	-3.37	-2.70	-3.28
0.25	-2.32	-2.66	-3.30	-2.24	-2.55	-3.15	-2.27	-2.59	-3.09
0.30	-2.27	-2.53	-3.16	-2.14	-2.51	-3.11	-2.15	-2.46	-3.05
0.35	-2.25	-2.51	-3.06	-2.13	-2.36	-3.06	-2.02	-2.45	-3.02
0.40	-2.13	-2.42	-3.01	-1.91	-2.21	-2.86	-1.98	-2.33	–2.98
0.45	-1.96	-2.39	-2.92	-1.87	-2.14	-2.87	-1.85	-2.17	-2.68
0.50	-1.85	-2.20	-2.9	-1.74	-2.10	-2.91	-1.77	-2.09	-2.73

Table 11c. **Critical Values. Constant and trend in the model.**

DGP: $\Delta y_t = c + \varepsilon_{t;}$ **Regression:** $\Delta y_t = c_0 + c_1 t + \phi \Delta^{d_1} y_{t-1}$.									
T	$T = 100$			$T = 400$			$T = 1000$		
d_1 / Sig. level	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.0	-3.20	-3.51	-4.36	-3.13	-3.38	-4.01	-3.14	-3.44	-3.99
0.05	-3.00	-3.35	-3.99	-2.71	-3.11	-3.77	-2.33	-2.69	-3.48
0.10	-2.71	-3.03	-3.67	-2.26	-2.69	-3.49	-1.81	-2.22	-3.01
0.15	-2.54	-2.90	-3.57	-2.01	-2.35	-3.20	-1.77	-2.15	-2.86
0.20	-2.26	-2.62	-3.50	-1.87	-2.27	-3.15	-1.66	-2.06	-2.83
0.25	-2.25	-2.61	-3.47	-1.85	-2.26	-3.02	-1.65	-2.06	-2.80
0.30	-2.24	-2.60	-3.35	-1.77	-2.21	-2.90	-1.62	-2.05	-2.79
0.35	-2.09	-2.45	-3.19	-1.76	-2.13	-2.88	-1.60	-2.03	-2.78
0.40	-2.05	-2.44	-2.88	-1.65	-2.11	-2.86	-1.59	-1.99	-2.74
0.45	-2.04	-2.40	-2.87	-1.63	-2.09	-2.83	-1.58	-1.95	-2.59
0.50	-1.91	-2.37	-2.73	-1.61	-2.08	-2.63	-1.56	-1.85	-2.47