A FRACTIONAL DICKEY-FULLER TEST FOR UNIT ROOTS

By Juan J. Dolado, Jesús Gonzalo, and Laura Mayoral

This paper presents a new test for fractionally integrated (FI) processes. In particular, we propose a testing procedure in the time domain that extends the well-known Dickey-Fuller approach, originally designed for the $I(1)$ versus $I(0)$ case, to the more general setup of $FI(d_0)$ versus $FI(d_1)$, with $d_1 < d_0$. When $d_0 = 1$, the proposed test statistics are based on the OLS estimator, or its t-ratio, of the coefficient on $\Delta^{d_1} y_{t-1}$ in a regression of Δy_t on $\Delta^{d_1} y_{t-1}$ and, possibly, some lags of Δy_t. When d_1 is not taken to be known a priori, a pre-estimation of d_1 is needed to implement the test. We show that the choice of any $T^{1/2}$-consistent estimator of $d_1 \in [0, 1)$ suffices to make the test feasible, while achieving asymptotic normality. Monte-Carlo simulations support the analytical results derived in the paper and show that proposed tests fare very well, both in terms of power and size, when compared with others available in the literature. The paper ends with two empirical applications.

Keywords: ARFIMA, Dickey-Fuller test, fractional processes, long memory, unit roots.

1. INTRODUCTION

It has become quite a standard practice in applied work to perform tests on whether a variable is integrated or stationary using both the null hypotheses of $I(1)$ and $I(0)$; 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 time 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, has proved very useful in capturing the persistence properties of many long-memory processes; cf. Baillie (1996), Beran (1994), and Granger and Joyeux (1980).

In general, unit root tests are consistent when the alternative is a $FI(d)$ process but their power turns out to be quite low (see Diebold and Rudebusch (1991), 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 a growing literature on this subject that can be basically classified into two strands. First, there are Wald-type tests that, 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, inter alia, Geweke and Porter-Hudak (1983), Fox and Taqqu (1986), Sowell (1992), Robinson (1992)). However, most of them lack power when used for testing purposes. On the one hand, the semiparametric techniques tend to yield large confidence intervals that include too often the null hypothesis. On the other hand, although in general the parametric methods present narrower confidence intervals, the precision with which the parameters are estimated hinges on the correct specification of the model (see Hauser, Potscher, and Reschenhofer (1999)).

Within the second group, Robinson (1994) and Tanaka (1999) have proposed useful LM tests in the frequency and the time domain, respectively. A distinctive feature of both approaches is that, in contrast to the classical unit root tests where asymptotic distributions are nonstandard and require case-by-case numerical tabulation, they do have standard asymptotic distributions. In this respect, Robinson (1994) has attributed this different limit behavior to the use of an explicit autoregressive (AR) alternative in the classical unit root-testing approach. Nonetheless, despite the advantage of having a standard limit distribution, a possible shortcoming of the LM approach is that, by working under the null hypothesis, it does not yield any direct information about the correct long-memory parameter, d, when the null is rejected.

In order to overcome that drawback, we propose in this paper a simple Wald-type test in the time domain that has acceptable power properties and, as a by-product of its implementation, provides information about the values of d under the alternative hypothesis. It turns out to be a generalization of the well-known Dickey-Fuller (D-F) test, originally developed for the case of $I(1)$ 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 normalized-OLS coefficient, or on its t-ratio, of the coefficient on $\Delta^{d_1} y_{t-1}$ 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 will concentrate on the case where $d_0 = 1$ and $0 \le d_1 < 1$, since this is the case which has received most attention in the literature. Specifically, we shall show that, in contrast to what happens with the LM test, the standard or nonstandard limiting behavior of the proposed test statistics hinges on 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 make it feasible. When a simple hypothesis is considered, i.e., $H_0 : d = d_0$ against $H_1 : d = d_1$, this last value is used to run the test. However, when a more general alternative hypothesis is envisaged, namely $H_0 : d = d_0$ versus $H_1 : d < d_0$, a pre-estimation of d under the alternative has to be used in

order to implement the test. We show that the choice of a $T^{1/2}$-consistent estimator of d in its relevant range suffices to make the FD-F test feasible while retaining asymptotic normality.

The advantages of the FD-F test arise from several sources. First, by generalizing the simple D-F framework to deal with $FI(d)$ processes, it keeps simplicity as one of its key features. Secondly, in contrast to some of the LM tests, it does not need to assume any known density for the errors and, therefore, offers potential for greater robustness. Thirdly, for the particular case where $d_0 = 1$, the FD-F approach inherits the flexibility of the standard D-F test for unit roots in providing a natural framework to test the null of $I(1)$ against some interesting composite alternatives. For instance, we might be interested in testing against the alternative of $FI(d)$ plus a break in the mean or in any other parameters of the processes, an extension that, to the best of our knowledge, cannot be easily undertaken with the available tests for fractional integration. Lastly, it fares very well in finite samples, in terms of power and size, when compared to other competing tests.

The rest of the paper is structured as follows. In Section 2, the FD-F test is defined and its asymptotic properties are derived. Further, its finite sample behavior is studied via Monte Carlo simulation and a comparison of its properties with those of other leading unit root tests is undertaken. For expository purposes, we just consider fractional white noise processes in this section, whose integration order, d, 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, it is shown that the required property of the estimator is $T^{1/2}$-consistency. Section 4, in turn, extends the results in Section 2 to more general $FI(d)$ processes. Section 5 discusses some empirical applications of the previous tests. Finally, Section 6 draws some concluding remarks.

Proofs of theorems and lemmae are collected in Appendix A, and critical values needed to implement the FD-F test for the particular case where it follows a nonstandard limiting distribution are reported in Appendix B.

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 nonstationary (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 for convenience at the beginning of Appendix A. Moreover, the following conventional notation is adopted throughout the paper: L is the lag operator, $\Delta = (1 - L)$, $\Gamma(\cdot)$ denotes the gamma function, and $\{\pi_i(d)\}$ represents the sequence of coefficients associated to the expansion of Δ^d in powers of L, which are defined as

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

The indicator function is denoted by $1_{(\cdot)}$.; $B(\cdot)$ represents standard Brownian motion (BM), whereas $B_d(\cdot)$ and $W_d(\cdot)$ are standard fractional Brownian motions (FBM) corresponding to the limit distributions of standardized partial sums of stationary and asymptotically stationary (truncated) $FI(d)$ processes,

respectively, as defined in Marinucci and Robinson (1999); see also the definitions of both processes and their limiting distributions at the beginning of Appendix A. Finally, $\overset{w}{\to}$ and $\overset{p}{\to}$ denote weak convergence and convergence in probability, respectively.

2. FRACTIONAL DICKEY-FULLER (FD-F) TEST

In this section we lay out the basis of the FD-F test in the simple case where a variable is considered to be a random walk under the null hypothesis, and a $FI(d)$ process under the alternative. When considering the class of ARIMA process, the Dickey and Fuller (1979, 1981) test statistic turns out to be one of the most popular tests in analyzing whether a process is $I(1)$ or $I(0)$. Its widespread use in applied work is mainly due both to its computational simplicity and adaptability to more general setups, such as serial correlation in the residuals, seasonality, breaking trends, etc. In its simplest formulation, the D-F statistic is based upon testing for the statistical significance of the parameter ϕ in the following regression model:

(1) $\qquad \Delta y_t = \phi y_{t-1} + u_t.$

If $u_t = \varepsilon_t$ is i.i.d., y_t is a random walk when $\phi = 0$ whilst y_t is a stationary $AR(1)$ process if $\phi < 0$. Regression (1) is an *unbalanced* one since the regressand and the regressor have different orders of integration under the null hypothesis: the regressand has been differenced to achieve an $I(0)$ variable under the null, whereas the regressor is in levels since it is $I(0)$ under the alternative. When considering the class of $FI(d)$ processes, the previous setup turns out to be very restrictive since 1 and 0 are only two specific points in the parameter space $d \in \Re$. Despite the fact that the D-F test is consistent against fractional alternatives, its low power makes it convenient to consider other testing procedures in such a case. In what follows, we generalize the regression model in (1) to test the null hypothesis that a series is $FI(d_0)$ against the alternative that it is $FI(d_1)$, where d_0 and $d_1 \in \Re$. Specifically, our proposal is based upon testing for the statistical significance of coefficient ϕ in the following regression:

(2) $\qquad \Delta^{d_0} y_t = \phi \Delta^{d_1} y_{t-1} + u_t,$

where u_t is an $I(0)$ process. Notice that, as in the standard D-F test, (2) is again an unbalanced regression where regressand and regressor have been differenced in agreement with their degree of integration under the null and the alternative hypothesis, respectively. Assuming that $u_t = \varepsilon_t$ in (2) then, when $\phi = 0$, the series follows the process

(3) $\qquad \Delta^{d_0} y_t = \varepsilon_t,$

implying that y_t is $FI(d_0)$. When $\phi < 0$, y_t can be expressed as

(4) $\qquad \left(\Delta^{d_0 - d_1} - \phi L \right) \Delta^{d_1} y_t = \varepsilon_t.$

The polynomial $\Pi(z) = ((1-z)^{d_0-d_1} - \phi z)$ has absolutely summable coefficients and verifies $\Pi(0) = 1$ and $\Pi(1) = -\phi \neq 0$. All the roots of the polynomial are outside the unit circle if $-2^{1-d_1} < \phi < 0$. As in the D-F framework, this condition excludes explosive processes.[2] Under the previous restriction on ϕ, $\Delta^{d_1} y_t$ is $I(0)$, so that y_t is an $FI(d_1)$ process that can be rewritten as

$$(5) \qquad \Delta^{d_1} y_t = C(L)\varepsilon_t,$$

where $C(z) = \Pi(z)^{-1} = ((1-z)^{d_0-d_1} - \phi z)^{-1}$, with $C(0) = 1$ and $0 < C(1) < \infty$.

In this manner, we are able to formulate a test statistic that is based either on the normalized-OLS estimated coefficient or on its t-ratio, as in the standard D-F testing approach, namely

$$(6) \qquad H_0 : \phi = 0, \ y_t \text{ is } FI(d_0),$$

$$(7) \qquad H_1 : \phi < 0, \ y_t \text{ is } FI(d_1),$$

such that when $d_0 = 1$ and $d_1 = 0$ the conventional $I(1)$ vs. $I(0)$ framework is recovered.

For simplicity, we shall restrict our analysis in the sequel to the specific case where $d_0 = 1$, namely, y_t is $I(1)$ under the null, whereas it is $FI(d_1)$, $0 \leq d_1 < 1$, under the alternative hypothesis. The choice for this case is dictated by its empirical relevance since that is the most treated case in the literature. However, in principle, the proposed framework can be potentially extended to deal with more general cases, such as $FI(d_0)$ against $FI(d_1)$ with $d_0 > d_1$, and in several instances throughout the rest of the paper we will make some conjectures about which results hold under this more general setup. Note that the case where $d_1 < 0$ is not considered since standard unit root tests behave well in such a case (see Mármol (1998)).

At this stage, it is convenient to discuss briefly the nature of the $FI(d)$ processes that we will use in the theoretical derivations, as well as the computation of the regressor $\Delta^{d_1} y_{t-1}$ in equation (2).[3] In agreement with the results in Lemma 1 below, we will use for the theoretical derivations the truncated version of an $FI(d)$ process, namely $y_t = \Delta^{-d_1} u_t 1_{(t>0)}$ where u_t is an $I(0)$ disturbance, such that $u_t = \varepsilon_t$ for the fractional white noise process. For convenience, we shall use the notation $\Delta^{d_1} y_t = u_t 1_{(t>0)}$ for that process. This truncation ensures that y_t has finite variance (albeit evolving at rate t^{2d-1}) and it implies that $y_t = 0$, $t \leq 0$. For practical purposes, we shall compute $\Delta^{d_1} y_t = \sum_{i=0}^{t-m-1} \pi_i(d_1) y_{t-i}$ where m is the integer part of $(d_1 + 1/2)$. Notice that a number of observations equal to m is lost when computing the filtered series in order to be consistent with the way in which integer differences would be applied to the data. For instance, if $d_1 = 1$, only the first observation would be lost.

---

[2] This restriction includes $\phi \in (-2, 0]$ as a particular case (when $d_1 = 0$), which is the familiar one in the D-F set-up.

[3] Since only the case $d_0 = 1$ is considered in the paper, we will omit the discussion about how to compute $\Delta^{d_0} y_t$, which, however, can be done along similar lines to the computation of $\Delta^{d_1} y_t$.

As mentioned above, a particularly interesting situation arises when $d_0 = 1$ and ε_t is i.i.d. in (3). In this case, we will be testing whether a series is a *random walk* against the alternative of a mean-reverting $FI(d)$ process. In order to convey the main properties of the FD-F test in a simple way, we will adopt this restrictive setting until Section 4 where the results are extended to the more general case where u_t is considered to be a linear process.

2.1. *The Test and its Asymptotic Properties*

Let us consider $d_0 = 1$ and $u_t = \varepsilon_t$ in (3), where $\{\varepsilon_t\}$ is a sequence of zero-mean i.i.d. random variables with unknown variance σ^2 and the finite fourth-order moment. The OLS estimator of ϕ, $\hat{\phi}_{ols}$, and its t-ratio, $t_{\hat{\phi}_{ols}}$, are given by the usual least-squares expressions

$$(8) \qquad \hat{\phi}_{ols} = \frac{\sum_{t=2}^{T} \Delta y_t \Delta^{d_1} y_{t-1}}{\sum_{t=2}^{T} (\Delta^{d_1} y_{t-1})^2},$$

$$(9) \qquad t_{\hat{\phi}_{ols}} = \frac{\sum_{t=2}^{T} \Delta y_t \Delta^{d_1} y_{t-1}}{S_T \left(\sum_{t=2}^{T} (\Delta^{d_1} y_{t-1})^2 \right)^{1/2}},$$

where the variance of the residuals, S_T^2, is given by

$$(10) \qquad S_T^2 = \frac{\sum (\Delta y_t - \hat{\phi}_{ols} \Delta^{d_1} y_{t-1})^2}{T}.$$

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

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

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

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

and

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

Then the following processes verify:
- *if* $-0.5 < d < 0.5$,

$$(11) \qquad T^{-1} \sum_{t=1}^{T} (x_t - x_t^*) = o_p(1),$$

$$(12) \qquad T^{-1} \sum_{t=1}^{T} (x_t^2 - x_t^{*2}) = o_p(1),$$

and

(13) $$T^{-1} \sum_{t=1}^{T} (x_t x_{t+k} - x_t^* x_{t+k}^*) = o_p(1);$$

- *if $d = -0.5$,*

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

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

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

LEMMA 2: *Let ε_t, x_t and x_t^* and z_t^* be defined as in Lemma 1. Then the following processes are martingale differences and verify:*
- *if $0 < d < 0.5$,*

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

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

- *if $d = -0.5$,*

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

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

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

(20) $$T^{-(1+d)} \sum_{t=2}^{T} z_{t-1}^* \varepsilon_t \xrightarrow{w} \sigma^2 \int_0^1 W_d(r) dB(r).$$

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

THEOREM 1: *Under the null hypothesis that y_t is a random walk, $\hat{\phi}_{ols}$ is a consistent estimator of $\phi = 0$ and converges to its true value at a rate T^{1-d_1} when*

$0 < d_1 < 0.5$, $(T \log T)^{1/2}$ when $d_1 = 0.5$, and at the standard rate $T^{1/2}$ when $0.5 < d_1 < 1$. Its asymptotic distribution is given by

$$(21) \qquad T^{1-d_1} \hat{\phi}_{ols} \xrightarrow{w} \frac{\int_0^1 W_{-d_1}(r) dB(r)}{\int_0^1 W_{-d_1}^2(r) dr} \quad if \quad 0 \le d_1 < 0.5,$$

$$(22) \qquad (T \log T)^{1/2} \hat{\phi}_{ols} \xrightarrow{w} N(0, \pi) \quad if \quad d_1 = 0.5,$$

and

$$(23) \qquad T^{1/2} \hat{\phi}_{ols} \xrightarrow{w} N\left(0, \frac{\Gamma^2(d_1)}{\Gamma(2d_1 - 1)}\right) \quad if \quad 0.5 < d_1 < 1.$$

THEOREM 2: *Under the null hypothesis that y_t is a random walk, the asymptotic distribution of $t_{\hat{\phi}_{ols}}$ is given by*

$$(24) \qquad t_{\hat{\phi}_{ols}} \xrightarrow{w} \frac{\int_0^1 W_{-d_1}(r) dB(r)}{(\int_0^1 W_{-d_1}^2(r) dr)^{1/2}} \quad if \quad 0 \le d_1 < 0.5,$$

and

$$(25) \qquad t_{\hat{\phi}_{ols}} \xrightarrow{w} N(0, 1) \quad if \quad 0.5 \le d_1 < 1.$$

Theorems 1 and 2 state that the standard or nonstandard asymptotic behavior of the previous test statistics depend on the distance between the null and the alternative hypotheses. When the alternative is also a nonstationary process ($0.5 \le d_1 < 1$), the limit distributions are standard (gaussian) and, conversely, they are nonstandard (functionals of FBM) when the alternative is a stationary (or asymptotically stationary) process ($0 \le d_1 < 0.5$). More generally, if the extended $FI(d_0)$ vs $FI(d_1)$ framework is considered, the asymptotic distribution of the FD-F test statistics will be standard when either the process is stationary (or asymptotically stationary) under both hypotheses or when the process is nonstationary under the null hypothesis ($d_0 > 0.5$) and $d_0 - d_1 < 0.5$, and nonstandard otherwise. Thus, in this framework, it is the distance between both the null and the alternative hypotheses, rather than the autoregressive specification under the alternative in the classical unit root approach, as argued by Robinson (1994), that determines the nature of the asymptotic distribution. Finally, note once again that when $d_0 = 1$ and $d_1 = 0$, we recover again the super-consistency of $\hat{\phi}_{ols}$, and the asymptotic distributions correspond to those derived by Dickey and Fuller (1979, 1981).

Next, we consider the behavior of the test under the alternative hypothesis. For simplicity we shall start assuming that the data generating process (DGP) is a simple fractional white noise defined as

$$(26) \qquad \Delta^{d_1} y_t = \varepsilon_t 1_{(t>0)},$$

where ε_t is a zero-mean i.i.d process. Denoting $\theta = d_1 - 1$, then $\Delta^{d_1} y_t = \Delta^{1+\theta} y_t$ and can be rewritten as

$$(27) \qquad \Delta y_t = \Delta^{-\theta} \varepsilon_t 1_{(t>0)} = \varepsilon_t + \pi_1(-\theta)\varepsilon_{t-1} + \sum_{i=2}^{t-1} \pi_i(-\theta)\varepsilon_{t-i},$$

where the coefficients $\pi_i(-\theta)$ are derived from the expansion of $(1-L)^{-\theta}$ in terms of powers of L. Since $\pi_1(-\theta) = \theta = d_1 - 1 (<0)$, under H_1 we can write

$$(28) \qquad \Delta y_t = \theta \Delta^{d_1} y_{t-1} + a_t,$$

with $a_t = \varepsilon_t + \sum_{i=2}^{t-1} \pi_i(-\theta)\varepsilon_{t-i}$. In this case, the coefficient on $\Delta^{d_1} y_{t-1}$ can be interpreted as the distance between the orders of integration of the processes under the alternative and the null hypotheses. Note that, despite the fact that a_t is not a white noise process, θ can be consistently estimated by OLS since a_t and $\Delta^{d_1} y_{t-1}$ are uncorrelated. It is a well-known result that the D-F test is consistent against fractional alternatives (see Sowell (1990) and Mármol (1998), for the stationary and nonstationary cases, respectively). In the following theorem we prove a more general result, which states the consistency of the FD-F test when the DGP is a fractional white noise with a true integration order, d_1^*, which could possibly differ from the order d_1 that is used to compute the proposed test statistics. Therefore, this theorem encompasses the results obtained for the standard D-F test, when $d_1 = 0$, namely, when the regressor in equation (2) is y_{t-1}.

THEOREM 3: *If the DGP is given by*

$$(29) \qquad \Delta^{d_1^*} y_t = \varepsilon_t 1_{(t>0)}, \quad d_1^* \in [0,1),$$

the test statistics based upon $\hat{\phi}_{ols}$ or $t_{\hat{\phi}_{ols}}$ in the regression of Δy_t on $\Delta^{d_1} y_{t-1}$ are consistent for any value of $d_1 \in [0,1)$.

Note that the previous theorem turns out to be extremely helpful since it guarantees the consistency of the proposed tests even when, under the alternative, an incorrect value of d_1 is considered to make the test feasible, insofar as $d_1 \in [0,1)$.

Another type of alternative hypothesis that is often considered in the literature is a local alternative to the null hypothesis. The LM test statistics proposed by Robinson (1994) and Tanaka (1999) are, under the assumption of gaussian errors, asymptotically uniformly most powerful invariant (UMPI) tests, in the sense of achieving asymptotically the largest power among all invariant tests under a sequence of local alternatives (see also Sargan and Bhargava (1983), Bhargava (1986), Tanaka (1996) for a more detailed analysis of the optimality properties of this type of test). The FD-F test also has power against that type of alternative but, in contrast to the LM tests, is not UMPI. The following theorem derives the asymptotic distribution of the test statistic based on $t_{\hat{\phi}_{ols}}$, under the assumption that the DGP is given by $\Delta^{1+\delta/\sqrt{T}} y_t = \varepsilon_t 1_{(t>0)}$, with $\delta < 0$.

THEOREM 4: *Under the assumption that the DGP is a fractional white noise defined as*

$$\Delta^{1+\delta/\sqrt{T}} y_t = \varepsilon_t 1_{(t>0)}, \qquad with \qquad \delta < 0,$$

the asymptotic distribution of the statistic $t_{\hat{\phi}_{ols}}$ computed from the regression of Δy_t on $\Delta^{1+\delta/\sqrt{T}} y_{t-1}$ is given by

$$(30) \qquad t_{\hat{\phi}_{ols}} \xrightarrow{w} N(\delta, 1).$$

This result implies that the probability of rejecting H_0: $\Delta y_t = \varepsilon_t$, under the DGP above, is

$$(31) \qquad P\left(t_{\hat{\phi}_{ols}} < -z_\alpha\right) \longrightarrow \Phi(-z_\alpha - \delta),$$

where z_α and $\Phi(z)$ are the quantile $100\alpha\%$ and the distribution function of a $N(0, 1)$ variate, respectively. However, under the assumption of normality, the maximum power of the LM tests against those alternatives is given by

$$(32) \qquad \Phi\left(-z_\alpha - \delta\sqrt{\frac{\pi^2}{6}}\right).$$

Since $\sqrt{\pi^2/6} = 1.28 > 1$, it follows that the FD-F test is not a UMPI test.

However, in order to check how important is that loss of power, Table I reports the power of both the FD-F and Tanaka's test (denoted as TAN) for $T = 25, 50, 100, 400$, and 1000 when the null hypothesis is a random walk against the local alternative of $1 - \delta/T^{1/2}$ with $\delta = (1, 2, 5, 10)$. The Monte Carlo experiment uses 1000 replications in MATLAB 5.3 for UNIX and the significance level is 5%. As it can be observed, the power of the TAN test is higher only for very local alternatives ($\delta = 1, 2$) and for very large sample sizes. Otherwise, the FD-F test outperforms the TAN test in terms of power. This result also holds for the standard unit root test and it is precisely for that reason that the D-F test is more commonly used in practice than the corresponding LM test. Further, we also found evidence that when the errors are nongaussian (for instance, when

TABLE I
POWER OF FD-F AND TAN TESTS

	FD-F						TAN					
δ/T	25	50	100	400	1000	∞	25	50	100	400	1000	∞
0	5.15	6.2	6.1	5.3	5.2	5	0.3	2	3.3	3.3	3.6	5
1	23.3	22.2	25.9	24.9	24.1	26	4.4	10.5	18.4	27.1	26.8	36
2	57.8	63.8	61.6	63.8	63.5	64	13.9	40.9	55.2	74.6	74.4	82
5	99.9	100	99.6	100	100	100	76.1	98.4	99.8	100	100	100
10	100	100	100	100	100	100	99.5	100	100	100	100	100

TABLE II
SIZE OF UNIT ROOT TESTS AGAINST FRACTIONAL 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	5.6%	5.9%	5.5%	4.9%	5.3%				
$T = 400$	d_1	0.5	0.6	0.7	0.8	0.9	5%	4.6%	3.7%	7.9%
	size	5.6%	5.2%	5.1%	5.0%	5.4%				

they follow a zero-mean standardized $\chi^2(1)$ distribution), the power of the FD-F test is higher (68% and 66%) than that of the TAN test (62% and 41%) for two of the smaller sample sizes ($T = 50$ and 100) and $\delta = 2$.

The results on power of the FD-F test that are collected in Theorems 3 and 4 only have asymptotic validity and, in the case of Theorem 4, they are derived against local alternatives. Thus, they turn out not to be too informative about how the test behaves in finite samples and when the alternative hypothesis is not a local one. In the following subsection we shall examine, through Monte Carlo simulations, how the FD-F test performs in those cases vis-á-vis other competing tests in the literature.

2.2. *Power of the FD-F Test in Finite Samples*

In order to analyze the size and power of the proposed test, we have generated a random walk and a number of FI processes with order of integration $d_1^* \in [0, 0.9)$, defined as $\Delta^{d_1^*} y_t = \varepsilon_t 1_{(t>0)}$, where ε_t are NID$(0, 1)$ innovations, for two sample sizes: $T = 100$ and $T = 400$. The number of replications is 1000 and the significance level is 5%.

Table II 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, $t_{\hat{\phi}_{ols}}$ was computed using different values of d_1 and size is reported for each of those values.[4] Table II also offers the size of the standard D-F test, the Geweke and Porter-Hudak (1983) test, and the LM tests proposed by Tanaka (1999) and Robinson (1994), denoted as GPH, TAN, and ROB, respectively. The number of periodogram ordinates included in the log-periodogram regression for the computation of the GPH test is set equal to $T^{1/2}$, as originally suggested by those authors. Figure 1, in turn, depicts the power of all those tests for $T = 100$, with the order of integration of the process, d_1, represented in the horizontal axis and the rejection frequencies in the vertical axis. For the FD-F test, we have computed the t-ratio $t_{\hat{\phi}_{ols}}$, under the assumption that the true order of integration,

---

[4] The empirical size is only reported for $d_1 \in [0.5, 0.9]$, which is the range whose critical values are those of a $N(0, 1)$ distribution. For $d_1 \in [0, 0.5)$, the critical values were derived numerically and therefore size is always the nominal one.
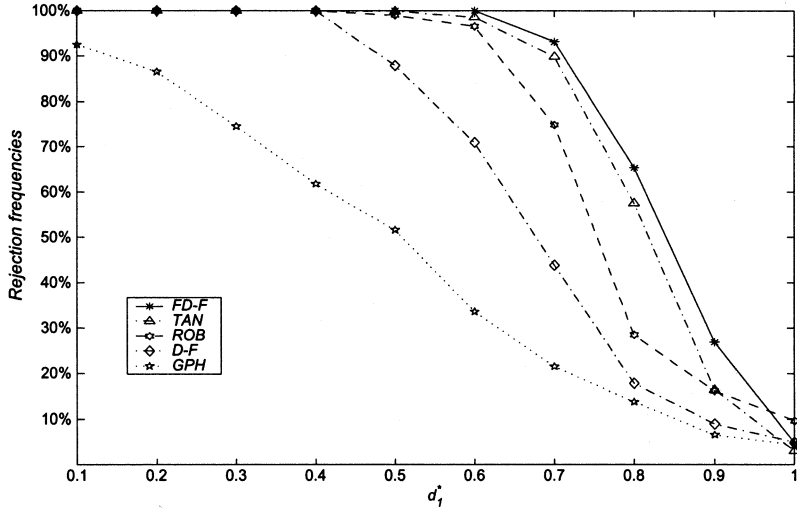
FIGURE 1.—Power of unit root tests.

d_1^*, is known.[5] When $d_1^* \in [0, 0.5)$, the critical values reported in Appendix B, for $T = 100$ and $T = 400$, are used to compute the power whereas, for the remaining values of d_1, the nominal critical values from a $N(0, 1)$ variate are considered.

From Table II and Figure 1, it follows that the FD-F test performs very well both in terms of size and power. As regards size, both LM tests present distortions in opposite directions while the other tests perform correctly. As for power, the FD-F test performs best in samples of moderate size ($T = 100$) and almost as well as TAN for larger sample sizes ($T = 400$), where power is not reported for the sake of brevity. Note that the low power of the GPH test is to be expected since it is designed to cater for general fractional models without any assumption on their spectral density, whereas the FD-F procedure is fully parametric and uses all available information, including the a priori known value of d under the null hypothesis. Finally, for all tests, given their consistency against fractional alternatives, power improves as the sample size grows.

As discussed in the Introduction, two definitions of a $FI(d)$ process (see Appendix A) have been adopted in the literature, depending on how the process is initialized: either it is assumed that the process (under stationarity) starts at $-\infty$ or, by contrast, it is assumed that all pre-sample values are equal to zero. As mentioned above, the latter is the assumption that we adopt in this paper. Nonetheless, it is interesting to examine whether the alternative definition can affect the properties of the test. Note that size will not be affected since, under the null hypothesis, the process is $I(1)$ and the initial condition is assumed to be

---

[5] Below, Table V reports the rejection frequencies of the FD-F test when d_1 is unknown and, thus, needs to be estimated according to the estimation procedure described in Section 3. Notice that in both cases the power is very similar.

TABLE III
Power of the FD-F Tests against FI Alternatives

d_1^*	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%	91.2%	65.9%	25.9%
$T = 400$	100%	100%	100%	100%	100%	100%	100%	99.8%	62.3%

equal to zero, as in the case of our truncated $FI(d)$ process. However, that is not the case when power is considered. This is so since we are dealing with a $FI(d)$ process and, unless we assume that all pre-sample values are set equal to zero, they will appear in the definition of the process. In order to evaluate the effect on the power of the test of using the alternative initializations of the process, we have carried out a small-scale Monte Carlo study, generating $FI(d)$ processes initialized in the distant past to compute the test statistics. More specifically, when $d_1 \leq 0.5$, we have generated $FI(d)$ processes with sample size $T' = K + T$, having discarded the first $K = 1000$ observations. When $d_1 > 0.5$, we have generated the partial sums of the corresponding stationary $FI(d)$ processes with order of integration given by $(d_1 - 1)$. Nevertheless, to compute the test statistics we need to filter the originally generated series and for that only observations $t = 1, \ldots, T$ have been used since, in practice, the values when $t \leq 0$ are unknown. Table III displays the power of the FD-F test for that last case, which turns out to be very similar to that shown in Figure 1. Thus, the effect of the initialization of the process at $t = 0$ seems to have minor effects on the power of the proposed test.

Under the alternative hypothesis, the true value of d_1, d_1^*, is unknown in practice. Thus, it is convenient to analyze, again by means of a Monte Carlo study, how robust is the FD-F test against misspecification in the value of d_1 used to run regression (2). Figure 2 displays the power of such a test for $T = 100$ when incorrect values of d_1 are used. More specifically, we have considered deviations of size ± 0.1, ± 0.2, and ± 0.3 from the true value, d_1^*. As it can be seen, we obtain very similar rejection rates to those reported in Figure 2, when d_1^* was used to compute the power, in accord with the result in Theorem 3. It is worth noticing that the power tends to decrease when values of d_1 larger than d_1^* are employed, especially when $d_1^* > 0.7$. This is so because in that case we are considering an alternative that is close to the null hypothesis and, therefore, the test turns out to be less powerful. Overall, the previous evidence suggests that the FD-F is quite robust in finite samples to misspecification in the value of d_1 and that by and large the good properties of the test do not depend too much on the accuracy of the estimation of d_1 under the alternative.

3. THE FD-F WITH UNKNOWN d_1

To implement the FD-F test, we need a value of the memory parameter d_1 under the alternative hypothesis. Sometimes, we can have an a priori knowledge
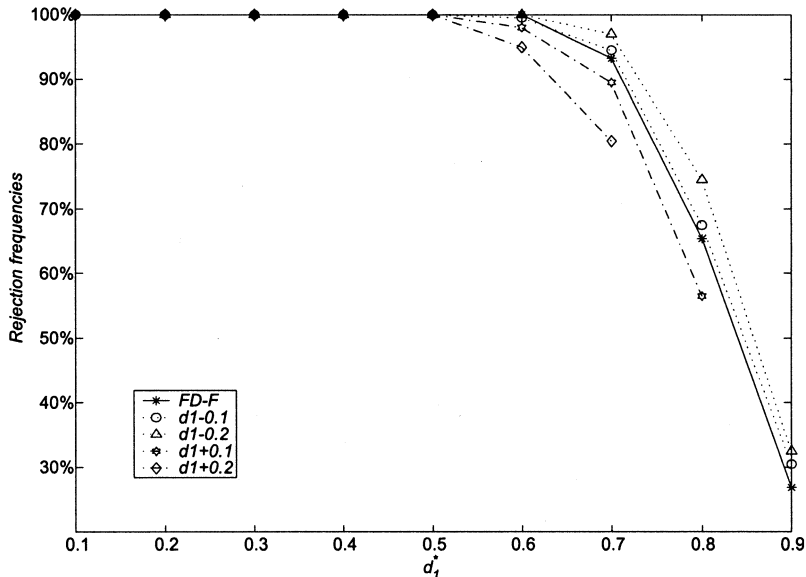
FIGURE 2.—Robustness against a misspecified d_1.

of this value but,[6] more realistically, we will need to estimate it, particularly when a composite alternative is being posed. There is a substantial literature on the estimation of the degree of integration, d, in $FI(d)$ models, both in the frequency and the time domain that can be helpful in order to find an appropriate pre-estimate. Indeed, the results in Section 2 about the relative robustness of the good power properties of the FD-F test to the degree of accuracy in the estimation of d_1, will help us in restricting the search to those estimators that satisfy a certain consistency property within the range $0 \leq d_1 < 1$. In particular, as shown in Theorems 5 and 6 below, the choice of any of the available $T^{1/2}$-consistent estimators of d_1 in the time domain will suffice to make the FD-F feasible. The basic result that we obtain is that if a $T^{1/2}$-consistent estimator of d_1 is used, the asymptotic distribution under H_0 of $t_{\hat{\phi}_{ols}}$ remains a $N(0,1)$. Next, we analyze the behavior of the FD-F test in finite samples through a Monte Carlo study to find that the results of these simulations strongly support the theoretical results.

3.1. *Asymptotic Distribution of the FD-F Test with an Estimated d_1*

To prove the previous result, let us assume that the DGP is a random walk such that $\Delta y_t = \varepsilon_t$. Let $\hat{d}_T$ be a $T^{1/2}$-consistent estimator of $d \leq 1$, such that

---

[6] For example, there is broad evidence that the long-memory parameter characterizing the time series behavior of political opinion polls in most OECD countries is $d = 0.7$; cf. Byers, Davidson, and Peel (1997).

$T^{1/2}(\hat{d}_T - d) \xrightarrow{w} \xi$, where ξ is a non-degenerate distribution. Since the value of d_1 that is needed to implement the test ought to be strictly smaller than 1, the pre-estimated value of d_1, $\hat{d}_1$, is chosen in accord with the following trimming rule:

$$(33) \qquad \hat{d}_1 = \begin{cases} \hat{d}_T, & \text{if } \hat{d}_T < 1 - c, \\ 1 - c, & \text{if } \hat{d}_T \geq 1 - c, \end{cases}$$

where $c > 0$ is a (fixed) value in a neighborhood of zero,[7] such that $(1 - c)$ is sufficiently close to unity. Then, the following theorem derives the asymptotic distribution of the test statistic $t_{\hat{\phi}_{ols}}$ under the null hypothesis that y_t is a random walk and that the value of d_1 needed to implement the test has been chosen in accord with criterion (33).

THEOREM 5: *Under the null hypothesis that y_t is a random walk, the test statistic $t_{\hat{\phi}_{ols}}(\hat{d}_1)$ associated to parameter ϕ in the regression*

$$\Delta y_t = \phi \Delta^{\hat{d}_1} y_{t-1} + a_t,$$

where $\hat{d}_1$ has been chosen according to the criterion defined in (33), *is asymptotically distributed as*

$$(34) \qquad t_{\hat{\phi}_{ols}}(\hat{d}_1) \xrightarrow{w} N(0, 1).$$

The result in the previous theorem implies that when a pre-estimated value of d_1 is used to implement the FD-F test, the corresponding critical values are those from a $N(0, 1)$ distribution, for any value of $\hat{d}_1$ which verifies criterion (33). It is important to note that this result differs from the one obtained before when the value of d_1 was assumed to be a priori known. In that case, when $d_1 \in [0, 0.5)$, the test had a nonstandard distribution under the null. However, the interpretation behind this seemingly contradictory result is a simple one. In effect, the fact that the limit distribution of the feasible FD-F test is always $N(0, 1)$ just reflects that, under the null and for T sufficiently large, the $T^{1/2}$-consistency of the estimator will lead to a small distance between the unit root and the estimated d_1, which is precisely the case where asymptotic normality is achieved in Theorem 1.

3.1.1. *Some Suitable Estimation Methods*

As proven above, in principle, any $T^{1/2}$-consistent estimator of $d_1 \in D \equiv [0, 1)$ could be used in regression (2) to make the FD-F test feasible. This restricts the choice to the class of parametric estimators of d_1, since semiparametric estimators generally have a slower rate of convergence (see, e.g., the survey in Baillie (1996)). Amongst the parametric class of estimators there are several procedures that achieve $T^{1/2}$-consistency within the specified set of values of d_1, both in the frequency and in the time domain. Within the frequency domain,

---

[7] It has to be verified that $0 < c < 0.5$.

Velasco and Robinson (2000) have extended the Whittle estimator, available for $d_1 \in (-0.5, 0.5)$ to the nonstationary range achieving a $T^{1/2}$-consistent and asymptotically normal estimator based on tapered data for $d_1 > 0.5$. That estimator, however, is not efficient for $d_1 > 0.5$ since the tapering increases its variance. Moreover, the amount of tapering to be included in the data increases with d_1, implying lower efficiency of the estimator the higher is the degree of fractional integration. A further shortcoming of that tapered estimator is that it is computationally complex to apply in practice. For that reason, and in parallel with the choice of the time domain as an appropriate underlying framework for the implementation of the FD-F test, we prefer to advocate the use of $T^{1/2}$-consistent estimators of $d_1 \in [0, 1)$ in the time domain to make the test feasible. Within the time-domain class of $T^{1/2}$-consistent estimators in the permissible range of d_1, the most popular ones are the maximum-likelihood estimators (MLE) due to Beran (1995) and Tanaka (1999) and the minimum distance (MD) estimator due to Galbraith and Zinde-Walsh (1997). Although any of those estimators can be used in principle to estimate d_1 we prefer to use a generalized MD estimator proposed by Mayoral (2000) (labelled as GMD) for the following two reasons. First, in contrast to Beran (1995) and Tanaka (1999), gaussianity of the error term is not assumed in the derivation of GMD. And secondly, in contrast with the other MD estimator by Galbraith and Zinde-Walsh (1997), where $T^{1/2}$-consistency is only proved for the stationary range of values of d_1 and conjectured for the nonstationary range, the GMD estimator can be shown to be $T^{1/2}$-consistent in the whole range of $d_1 \in [0, 1)$. Thus the GMD estimator will be used to illustrate the computation of the feasible FD-F test in the sequel. This estimator can be seen as an extension of the MD estimator proposed by Tieslau, Schmidt, and Baillie (1996) where, instead of minimizing a criterion function based on the distance between the theoretical and the sample autocorrelations of the original series y_t (which only exist for stationary $FI(d)$ processes), it minimizes a similar criterion based this time on the autocorrelations of the residuals obtained after filtering the series with the appropriate ARFIMA parameters (which also exist for nonstationary $FI(d)$ processes).

Specifically, the proposed GMD estimator can be used for the general case where u_t is an $ARMA\,(p, q)$ process, i.e., $u_t = \Phi_p(L)^{-1}\Theta_q(L)\varepsilon_t$, such that $\Phi_p(L)$ and $\Theta_q(L)$ are pth and qth order lag polynomials, respectively, with all their roots outside the unit circle. Hence, y_t is assumed to follow the $ARFIMA(p, d, q)$ process

$$(35) \qquad \Phi_p(L)\Delta^d y_t = \Theta_q(L)\varepsilon_t,$$

such that $d \in (-3/4, \infty)$, i.e., a set that contains D.

Let $\Psi = (\phi_1, \ldots \phi_p, \theta_1, \ldots \theta_q)'$ be the vector containing the AR and MA parameters, $\lambda = (d, \Psi')'$ and $\varphi_j(\lambda)$, $j = 0, \ldots, \infty$, be the coefficients associated to the expansion $\Phi_r(L)\Theta_s(L)^{-1}\Delta^d$. Also let λ^* be the true parameter values of

the model in (35). Then, given the observations $y_1, \ldots, y_T$, we can define the residuals

$$(36) \qquad e_t(\lambda) = \sum_{j=0}^{t-1} \varphi_j(\lambda) y_{t-j}.$$

Next, define $V_{Te}(\lambda) = \sum_{i=1}^{k} \hat{\rho}_{ie(\lambda)}^2$, where $\hat{\rho}_{ie(\lambda)}$ is the sample ith order autocorrelation associated with the residuals $e_t(\lambda)$, where $k = o_p(T)$. In practice, $k = T^{1/4}$ is a good choice. Then, the MDE of λ is defined as

$$(37) \qquad \hat{\lambda} = \arg\min_{\lambda \in \Lambda} V_{Te}(\lambda).$$

Notice that, since $e_t(\lambda^*) = \varepsilon_t$, the population autocorrelations are zero and $V_{Te}(\lambda)$ has a unique minimum at $\lambda = \lambda^*$. Mayoral (2000) has proved that $\hat{\lambda}$ turns out to be a $T^{1/2}$-consistent, asymptotically normally distributed estimator of λ, for $d \in (-3/4, \infty)$, with the same asymptotic variance as the MLE derived by Beran (1995). MATLAB programs to implement this estimator are available upon request.

3.1.2. *Behavior of the Estimator of d_1 in Finite Samples*

In this section, we evaluate the accuracy of the proposed GMD estimator of d_1 in finite samples through Monte Carlo experimentation. We have generated fractional white noise processes like (26) with $\varepsilon_t \sim NID(0, 1)$, with the initial values needed to start the minimization algorithm being set equal to zero in all cases. Table IV lists the mean of the estimated $\hat{d}_1$ across simulations. That is, $\hat{d}_1 = \sum_1^N \hat{d}_{1i}/N$ and $\hat{\sigma} = (\sum_1^N (\hat{d}_{1i} - \hat{d}_1)^2/N)^{1/2}$, where $N = 1000$ is the number of replications, for two sample sizes, $T = 100$ and $T = 400$.

In general, we find that the proposed GMD estimator provides very accurate estimates of d_1 for both sample sizes and, therefore, that turns out to be useful in implementing a feasible FD-F test.

TABLE IV

Estimation of d Based on Criterion (37)

					$T = 100$						
d_1^*	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\hat{d}_T$	0.001	0.105	0.203	0.311	0.407	0.501	0.602	0.704	0.801	0.903	1.01
$\hat{\sigma}_{\hat{d}}$	0.092	0.093	0.092	0.092	0.093	0.093	0.091	0.094	0.093	0.091	0.094
					$T = 400$						
d_1^*	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\hat{d}_T$	0.00	0.102	0.203	0.299	0.403	0.501	0.598	0.701	0.800	0.910	0.998
$\hat{\sigma}_{\hat{d}}$	0.043	0.042	0.044	0.042	0.043	0.040	0.041	0.043	0.042	0.041	0.040

TABLE V

TABLE V

SIZE AND POWER OF FD-F TEST WITH AN ESTIMATED d_1

$T \mid d_1^*$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$T = 100$	100%	100%	100%	100%	100%	99.4%	97.7%	86.4%	65.4%	25.1%	4.3%
$T = 400$	100%	100%	100%	100%	100%	100%	100%	100%	98.9%	63.1%	5.1%
$T = 1000$	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	5%

3.2. *Behavior of the FD-F Test with an Estimated d_1*

In this subsection we present the results of the simulations carried out to evaluate the behavior of the FD-F test when d_1 is estimated using the GMD estimator defined above. We designed the same Monte Carlo experiment as in subsection 2.2, albeit using this time the pre-estimated value of d_1 to compute the *t*-ratio. The chosen value of c was 0.02, so that $\hat{d}_1 \leq 0.98$. Table V lists the power and size of the test in that case.

As can be seen, the results in Table V turn out to be very similar to those reported in Table II and Figure 1, where the true value of d_1 was being used to compute the test. In accord with the results derived in Theorems 3 and 5, the basic observation that we draw from this exercise is that replacing the true value of d_1 by $\hat{d}_1$ hardly affects the size and power properties of the proposed test.

Next, Figure 3 displays the normal probability plot (*NP-plot*) which compares the empirical distribution of the feasible *t*-ratio, $t_{\hat{\phi}_{ols}}(\hat{d}_1)$, with a $N(0, 1)$ distribution, for two sample sizes, $T = 100$ and $T = 400$. We can observe that the approximation to a standardized normal variate works well even for the smaller sample size.

Next, we check if the promising results obtained so far for the null that a series is a random walk against the alternative that is a fractional white noise can be extended to more general set-ups where serial correlation in the disturbances is allowed.

4. THE AUGMENTED FD-F TEST

In several applications, such as in the case of financial or opinion poll data, the characterization of a series as a fractional white noise may well be plausible. However, in many other cases, it seems of paramount importance to allow for more general models where there is serial correlation in the errors. In this section, the results in Section 2 are extended to that more general setup. For that, we follow the well-known Augmented Dickey-Fuller (AD-F) testing approach, and find that the asymptotic distribution of the *t*-ratio (see Theorem 2) remains valid in the presence of serial correlation, as long as a sufficient number of lags of $\Delta^{d_0} y_t$ (Δy_t if $d_0 = 1$) is included in the regression.
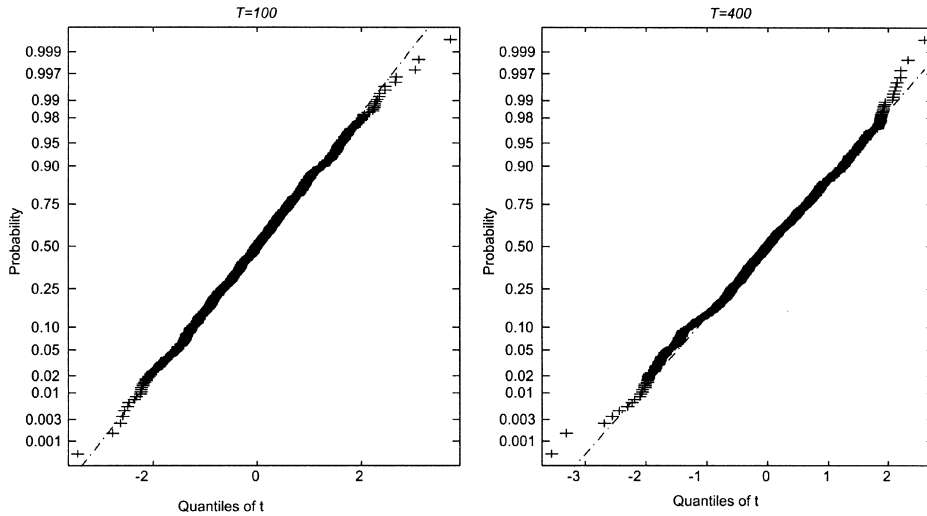
FIGURE 3.—NP-plot of $t_{\hat{\phi}_{ols}}(\hat{d}_1)$.

4.1. *Definitions and Asymptotic Properties*

The proposed testing procedure imitates the traditional DF approach in the context of ARIMA processes. Let us consider again regression (1):

$$\Delta y_t = \phi y_{t-1} + u_t,$$

where now $\alpha(L)u_t = \varepsilon_t$, i.e., an autoregressive process of order p, $AR(p)$, such that $\alpha(L) = 1 - \alpha_1 L - \cdots - \alpha_p L^p$ has all its roots outside the unit circle. Then the AD-F test is based upon the following regression:

$$\Delta y_t = \phi y_{t-1} + \sum_{i=1}^{p} \zeta_i \Delta y_{t-i} + \varepsilon_t.$$

Under the null hypothesis, Dickey and Fuller (1981) showed that the asymptotic distribution of the *t*-ratio $t_{\hat{\phi}_{ols}}$ is identical to the one derived in the absence of serial correlation. Following the same analogy that was used in Section 2, the proposed regression for the case where u_t is an AR (p) process[8] is as follows:

$$(38) \qquad \Delta y_t = \phi \Delta^{d_1} y_{t-1} + \sum_{i=1}^{p} \zeta_i \Delta y_{t-i} + \varepsilon_t.$$

As in regression (2), the null of a unit root and the alternative hypotheses of a $FI(d_1)$ can be expressed in terms of the parameter ϕ when $\phi = 0$ and $\phi < 0$,

[8] As in the case of the AD-F test for unit roots, a generalization to ARMA(p, q) processes can be implemented along the lines of Said and Dickey (1984) and Ng and Perron (1995).

respectively. When $\phi = 0$, it is easy to prove that y_t follows an ARIMA$(p, 1, 0)$ process whereas, when $\phi < 0$, y_t is given by

$$(\alpha(L)\Delta^{1-d_1} - \phi L)\Delta^{d_1} y_t = u_t,$$

where, as in Section 2, the polynomial $\Pi(z) = (\alpha(z)(1-z)^{1-d_1} - \phi z)$ has absolutely summable coefficients which verify that $\Pi(0) = 1$, $\Pi(1) = -\phi$, and $\Delta^{d_1} y_t$ is an invertible process[9] that, under H_1, can be rewritten as

$$\Delta^{d_1} y_t = C(L)\varepsilon_t,$$

where, in this case, $C(z) = (\Pi(z))^{-1}\alpha(z)$ such that $C(0) = 1$ and $0 < C(1) < \infty$.

The previous arguments lead to a test based upon the t-ratio $t_{\hat{\phi}_{ols}}$ computed this time from the extended regression (38). In the following theorem we derive the asymptotic distribution of that test statistic and the remaining coefficients in (38). In parallel with the well-known result for the standard AD-F test, it is found that the asymptotic distribution of the $t_{\hat{\phi}_{ols}}$ is identical to that derived under the assumption of uncorrelated disturbances.

THEOREM 6: *Under the null hypothesis that y_t is an ARIMA$(p, 1, 0)$ process*:
- *The asymptotic distribution of $t_{\hat{\phi}_{ols}}$ in regression* (38) *is the same as in the i.i.d. case, given in Theorem* 2.
- $(\hat{\zeta}_1, \dots, \hat{\zeta}_p)'$ *are asymptotically normally distributed for any value of $d_1 \in [0, 1)$ used in the regression* (38).

Notice that, as is the case in the standard AD-F framework, a test based on the normalized OLS coefficient, $\hat{\phi}_{ols}$, will depend on nuisance parameters. Hence, the results in Theorem 1 do not hold for that test statistic when short-term structure is allowed in the errors. Therefore, in the sequel only the performance of $t_{\hat{\phi}_{ols}}$ will be analyzed.

First, let us now consider the behavior of the test under the alternative hypothesis of fractional integration. Krämer (1998) proves the consistency of the AD-F test 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 are available on the consistency of the AD-F test when the DGP is a more general ARFIMA process. The following theorem states the consistency of the AFD-F test when the DGP is an ARFIMA(p, d_1, q) and the number of lags included in the regression, k, tends to infinity at a slower rate than T, so that $k^3/T \to 0$.

THEOREM 7: *If the DGP is given by $\Delta^{d_1} y_t = u_t 1_{(t>0)}$, with $d_1 \in [0, 1)$ and u_t is a stationary and invertible ARMA (p, q) process satisfying $\alpha_p(L)u_t = \beta_q(L)\varepsilon_t$, $\varepsilon_t \sim i.i.d.(0, \sigma^2)$, $\alpha_0 = \beta_0 = 1$ with p and q possibly unknown, the t-ratio associated to*

---

[9] The condition that $\alpha(-1)2^{1-d_1} < \phi < 0$ is needed to ensure that all the roots of the polynomial $\Pi(z)$ are outside the unit circle.

the OLS estimator of ϕ in the regression $\Delta y_t = \phi \Delta^{d_1} y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_k \Delta y_{t-k} + i_{tk}$ diverges to $-\infty$ when $k \to \infty$ as $T \to \infty$, and $k^3/T \to 0$, implying the consistency of the test.

Since the implementation of the LM test needs precise knowledge of the serial correlation structure of the process, the previous result implies an important advantage of the AFD-F vis-á-vis the use of parametric LM tests. This is so since the semiparametric nature of the AFD-F test does not require such a precise information.

4.2. Behavior of the AFD-F Test in Finite Samples

A 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:

$$(39) \qquad \Delta^{d_1^*} y_t = u_t,$$

$$(40) \qquad u_t = \alpha u_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim n.i.d.(0, 1),$$

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

$$(41) \qquad \Delta y_t = \phi \Delta^{d_1^*} y_{t-1} + \zeta \Delta y_{t-1} + a_t.$$

We have computed the size of the test using the same critical values as in the i.i.d. case and have obtained similar values to those reported in Table II. This implies that the test is very well behaved in terms of size, even when short-term structure is allowed in the errors, due to its ability to capture this structure in u_t. The size values are not reported here although they are available upon request.

Table VI provides the rejection frequencies at the 5% significance level for $T = 100$ and $T = 400$ and $\alpha = -0.8, 0.2$ and 0.6. The two extreme values in that

TABLE VI

POWER OF THE TEST WITH STATIONARY AR(1) ERRORS

$T \mid d_1^*$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					$\alpha = -0.8$					
$T = 100$	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%
					$\alpha = 0.2$					
$T = 100$	100%	100%	100%	100%	100%	98.8%	90.0%	65.6%	39.2%	19.0%
$T = 400$	100%	100%	100%	100%	100%	100%	100%	99.6%	85.6%	38.6%
					$\alpha = 0.6$					
$T = 100$	100%	100%	99.7%	95.5%	84.3%	63.5%	42.5%	36.0%	17.2%	9.90%
$T = 400$	100%	100%	100%	100%	100%	100%	95.4%	68.2%	32.3%	14.1%

range are those analyzed by Tanaka (1999) whereas the middle one has been chosen as an intermediate value. The number of replications is 1000. In all cases, the true d_1 has been used to run the regression. To a large extent, the power of the test depends on how close the polynomial $\alpha(z)$ is to having a unit root. To check that, let us assume that y_t follows an ARFIMA$(1, d_1, 0)$ process defined as $\Delta^{d_1} y_t = \varepsilon_t / (1 - \alpha L)$; then y_t can be rewritten as (see the proof of Theorem 7):

$$\Delta y_t = -\alpha(1)\Delta^{d_1} y_{t-1} + (1 - \Delta^{d_1})\Delta y_t + \varepsilon_t,$$

where the first coefficient in the expansion of $(1 - \Delta^{d_1})$ around $L = 0$ is zero. Note that the closer $\alpha(1)$ is to zero, i.e. the closer the coefficient on $\Delta^{d_1} y_{t-1}$ is to zero, the smaller the power of the test will be. That is to be a priori expected since, in the limit, as α tends to unity, y_t will tend to a $FI(1 + d_1)$ process that contains a unit root within the range of values permitted under the alternative hypothesis, and hence its power will be small. This is confirmed in the simulation when $\alpha = 0.6$. However, the AFD-F test remains quite powerful for other smaller or negative values of α.

We have also carried out simulations to study the robustness of the test when d_1 is misspecified in the regression and $\alpha = 0.2$. As in the i.i.d. case, we have considered deviations of $\pm 0.1, \pm 0.2$, and ± 0.3 from the true value of d_1 and found small changes with respect to the power reported in Table VI. For example, when $d_1^* = 0.6$ and $T = 100$, the power for $d_1^* \pm 0.2$ is 88.5 and 83.3%, respectively, whereas at the true value it is 90.6%. Other alternative values of α have also been considered, yielding similar results.

4.3. Behavior of the AFD-F Test with an Estimated d_1

Finally, we analyze properties of the AFD-F test when d_1 is estimated. As in the i.i.d. case, the asymptotic distribution of the t-ratio with an estimated d_1 is identical to that obtained with known d_1 whenever a $T^{1/2}$-consistent estimator of d_1 is used, as shown in the following Theorem.

THEOREM 8: *Under the null hypothesis that y_t is generated by*

$$(42) \qquad \Delta y_t = u_t, \quad \alpha_p(L)u_t = \varepsilon_t,$$

the asymptotic distribution of the t-ratio $t_{\hat{\phi}_{ols}}(\hat{d}_1)$ associated to coefficient ϕ in regression (38), *where d_1 is an estimator of d that minimizes the criterion function defined in* (33), *is as follows*:

$$(43) \qquad t_{\hat{\phi}_{ols}}(\hat{d}_1) \xrightarrow{w} N(0, 1).$$

Figure 4 portrays the NP-Plot, which compares the distribution of $t_{\hat{\phi}_{ols}}(\hat{d}_1)$ with a $N(0, 1)$ distribution when the DGP is an ARIMA$(1, 1, 0)$ and d_1 has been estimated in two samples of size $T = 100$ and $T = 400$ using the GMD estimator
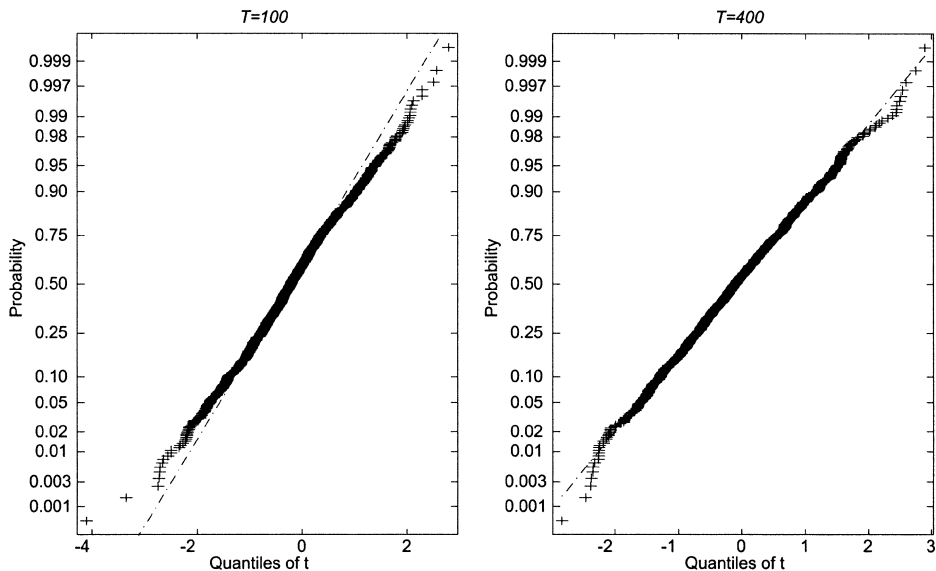
FIGURE 4.—NP-plot of $t_{\hat{\phi}_{ols}}(\hat{d}_1)$.

described in subsection 3.1.1. As in the i.i.d. case, the approximation to a $N(0,1)$ distribution works reasonably well in the finite samples considered.

Finally, to explore the behavior of the test with an estimated d_1, we have run another small Monte Carlo study. Table VII reports the rejection frequencies, at the 5% significance level, of the test based on the t-ratio in the regression: $\Delta y_t = \phi \Delta^{\hat{d}} y_{t-1} + \zeta \Delta y_{t-1} + e_t$. As shown above, the critical values are those of a $N(0,1)$. The DGP is ARFIMA $(1, d_1, 0)$ process with the autoregressive parameter, α, taking the values $\alpha = 0.2$ and 0.6.

As in the i.i.d. case, we can observe that the power loss due to the pre-estimation of d_1 turns out to be small, although it happens to be larger than that obtained under i.i.d. error terms.

TABLE VII
POWER AND SIZE OF THE AFD-F TEST WITH ESTIMATED d_1

	$\alpha = 0.2$										
$T \mid d_1^*$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$T = 100$	100%	100%	100%	100%	99.0%	98.2%	88.5%	63.3%	38%	18.2%	4.5%
$T = 400$	100%	100%	100%	100%	100%	100%	99.9%	98.1%	83.5%	37.1%	6.0%
	$\alpha = 0.6$										
$T \mid \backslash d_1^*$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$T = 100$	100%	100%	99.7%	92.2%	82.9%	59.3%	41.3%	34.2%	15%	9.4%	5.4%
$T = 400$	100%	100%	100%	100%	100%	97.5%	94.5%	67.8%	31.1%	13.4%	5.1%

4.4. *Analogies between the FD-F and the LM Tests*

As mentioned above, Robinson (1994) and Tanaka (1999) have developed LM tests in the frequency and time domain, respectively. Both those tests and the FD-F and AFD-F tests share some features that makes them different from the traditional unit root tests, their most remarkable property being that they have standard limit distributions. To analyze the analogies and differences between those tests and the one we propose, we will focus on Tanaka's test since, as with the FD-F test, it is defined in the time domain.

Tanaka's (1999) LM test statistic is designed to test $H_0 : \theta = 0$ against $H_1 : \theta < 0$ or $H_1 : \theta > 0$ in the model

$$(44) \qquad \Delta^{d_0+\theta} y_t = u_t, \qquad y_t \sim N(0, \Omega),$$

where d_0 is a given constant and $u_t = \beta_q(L)/\alpha_p(L)\varepsilon_t$ is an ARMA(p, q) process. Let $L(\cdot)$ be the log-likelihood and $\psi = (\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)'$. Then, the proposed LM test statistic, LM_{TAN}, is given by

$$(45) \qquad LM_{TAN} = \frac{\partial L(\theta, \psi, \sigma^2)}{\partial \theta}\bigg|_{H_0} = T \sum_{k=1}^{T-1} \frac{1}{k} \hat{\rho}_k,$$

where $\hat{\rho}_k = \sum_{j=k+1}^{T} \hat{\varepsilon}_{j-k} \hat{\varepsilon}_j / \sum_{j=1}^{T} \hat{\varepsilon}_j^2$ is the kth autocorrelation of the residuals $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_T$, calculated under the null and therefore defined as

$$\hat{\varepsilon}_t = \hat{\alpha}(L)\hat{\beta}(L)^{-1}(1-L)^{d_0} y_t.$$

Under H_0, it can be shown that the previous test statistic is asymptotically distributed as a $N(0, W)$, where W is a function of both the elements of the inverse of the Fisher matrix and the coefficients ψ (for a detailed description, see Tanaka (1999)). As pointed out by the author, the computation of W can become cumbersome when u_t follows a general linear process. Note that this is not the case for the AFD-F test whose asymptotic distribution is always a $N(0, 1)$. However, the latter test needs, in most cases, a pre-estimation of d under H_1, in contrast to the LM tests.

In the absence of serial correlation, it is possible to carry out a detailed analytical comparison of both tests. In such a case, the LM_{TAN} test in (45) and the FD-F test, based upon the OLS estimator of ϕ are, respectively

$$LM_{TAN} = T \frac{\sum_{t=1}^{T} \Delta^{d_0} y_t \left(\sum_{k=1}^{t-1} \frac{1}{k} \Delta^{d_0} y_{t-k} \right)}{\sum_{t=1}^{T} (\Delta^{d_0} y_t)^2},$$

$$\hat{\phi}_{ols} = \frac{\sum_{t=1}^{T} \Delta^{d_0} y_t \Delta^{d_1} y_{t-1}}{\sum_{t=1}^{T} (\Delta^{d_1} y_{t-1})^2}.$$

Under the null hypothesis, they can be written as

$$(46) \qquad LM_{TAN}|_{H_0} = T \frac{\sum_{t=2}^{T} \varepsilon_t \left(\sum_{k=1}^{t-1} \frac{1}{k} \varepsilon_{t-k} \right)}{\sum_{t=1}^{T} (\varepsilon_t)^2} = T \sum_{k=1}^{T-1} \frac{1}{k} \hat{\rho}_\varepsilon(k)$$

and

$$(47) \qquad \hat{\phi}_{ols}\big|_{H_0} = \frac{\sum_{t=2}^{T} \varepsilon_t \left(\sum_{k=1}^{t-1} \pi_i (d_1 - d_0) \varepsilon_{t-k} \right)}{\sum_{t=2}^{T} \left(\sum_{k=1}^{t-1} \pi_i (d_1 - d_0) \varepsilon_{t-k} \right)^2}.$$

A comparison of (46) and (47) yields the following interesting observations. On the one hand, the numerators in the expressions above look similar except that the weights on the innovations ε_{t-k} differ. In the case of $\hat{\phi}_{ols}$, the weights are a function of the distance between d_0 and d_1. When that distance is large, the size of the weights increases $(0 \le d_1 < 0.5)$ implying that a CLT does not hold, and therefore that the limit distribution of the test is a nonstandard one. This feature does not arise with the LM_{TAN} test since its weights do not depend on d_1. On the other hand, a similar argument applies to the denominator, which precludes a choice of weights that would make both expressions equivalent. When dealing with serially correlated disturbances, the analytical comparison between both test statistics becomes much more cumbersome since, in that case, they differ not only in the weighting factors but in many other dimensions.

5. EMPIRICAL APPLICATIONS

In order to provide some empirical illustrations of how the testing and estimation methods proposed in this paper can be applied in practice, we have examined two series for which evidence of fractional integration has been found before in the literature. More specifically, the first one is the UK Gallup opinion poll series (cf. Byers, Davidson, and Peel (1997)) whereas the second one is the US unemployment rate series in Nelson and Plosser's (1982) data set.

5.1. *Opinion Poll Data*

We consider the Gallup opinion poll series of support for the Conservative and the Labour parties in the UK, analyzed by Byers, Davidson, and Peel (1997). Their basic finding is that the logistic transformation of these series (after removal of the deterministic components due to the 'election cycle' effect)[10] is well characterized by a fractional white noise with $0.5 < d < 1$, so it is nonstationary, albeit mean reverting. The results obtained through our method are very similar. The data are monthly and their length consists of 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 VIII presents the results of applying the FD-F test to both series, i.e., regressing Δy_t on $\Delta^{\hat{d}_1} y_{t-1}$, where $\hat{d}_1$ is the GMD estimate, and computing the t-ratio 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 d_1 has been pre-estimated.

---

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

TABLE VIII

	$t_{H_0:d_0=1}$	$\hat{d}_1$
Conservative	−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, Davidson, and Peel (1997) using the Sowell (1992) exact maximum likelihood approach, which are 0.779 and 0.726 for the Conservative and Labour cases, respectively.

In view of those results, the conclusion that can be drawn is that the shocks (news) have a transitory effect on voting intentions and that voters forget eventually about them, albeit not rapidly, implying that voting intentions can be characterized as a long-memory process. Similar results are obtained for the Spanish opinion polls (see Dolado, Gonzalo, and Mayoral (2001)).

5.2. *Unemployment Rate 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). By applying Robinson's LM test, they find evidence of fractional integration in many of them. In particular, we examine the unemployment rate, which is the series where there is more controversy about its degree of integration. The data are annual, from 1891 to 1988 and, as in Nelson and Plosser (1982), they have been transformed to natural logarithms. Unlike in most of the above-considered examples, the fractional white noise hypothesis does not seem to be appropriate in this instance. Thus, we consider different specifications of u_t in equation (2), namely that u_t is white noise, with and without a constant in the model,[11] and that it is an AR(p) process. For the first two specifications, the null hypothesis of a unit root cannot be rejected. Nevertheless, this result is hardly reliable since the residuals obtained after differencing the series show significant autocorrelations, presenting evidence of misspecification in the model (the values for the $Q(10)$-statistic are 19.58 and 18.13 respectively, which are rejection values for a χ_9^2 distribution at the nominal 5% level). Table IX also reports the values of the test when an AR(1) is fitted to u_t. In this case, none of the autocorrelations of the residuals were significant, making this specification preferable

[11] As in the D-F framework, the asymptotic behavior of the FD-F test changes when deterministic components are introduced in the regression equation for the case where $d_1 < 0.5$. 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 2, with the FBM replaced by demeaned FBM. Appendix B incorporates the critical values for the nonstandard range of values of d.

TABLE IX

FD-F AND AFD-F TEST ON UNEMPLOYMENT RATE

	$t_{H_0:d_0=1}$	$\hat{d}_1$
u_t white noise	0.57	0.852
Constant term and u_t white noise	0.36	0.863

to the previous ones (the $Q(10)$ statistic is distributed in this case as a χ^2_8, and yields a value of 13.47) and therefore the null of white noise residuals cannot be rejected. The AFD-F test rejects the unit root null hypothesis, a result that agrees with those obtained by Gil-Alaña and Robinson (1997) and Nelson and Plosser (1982) in the context of ARIMA models. The estimated value of d_1 is in this case 0.412, lying in the stationary range of values of d_1, although close to the nonstationary boundary.

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, we focus on the case when $d_0 = 1$ and $0 \le d_1 < 1$, where 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, any $T^{1/2}$-consistent estimator of d_1, within the permissible range can be used. In particular, we advocate the use of a Generalized Minimum-Distance estimator of d_1 proposed by Mayoral (2000) in the time domain which achieves $T^{1/2}$-consistency and asymptotic normality, and which turns out to work well in finite samples. Several empirical illustrations of how to use and interpret these tests are provided. Further, various Monte Carlo experiments throughout the paper support the analytical results and show that the proposed tests behave reasonably well in finite samples.

Further research is currently being undertaken towards generalizing the FD-F testing approach along similar directions as the D-F test has been extended in the unit root literature accounting for exogenous and endogenous structural breaks, heteroscedastic error terms, seasonality, fractional cointegration, etc. The extension of the proposed testing approach to a multivariate framework is particularly interesting since some of the best-known cointegration tests, such as Stock-Watson (1988) and Johansen (1991), can be interpreted as multivariate generalizations of the AD-F tests. This extension is also crucial for cointegration analysis since it is known that a mistake in the determination of the parameter d can produce spurious cointegration (see Gonzalo and Lee (1988)).

Dept. of Economics, Universidad Carlos III de Madrid, calle Madrid 126, 28903 Getafe, Madrid, Spain; dolado@eco.uc3m.es; http://www.eco.uc3m.es/personal/ info_contacto/dolado.html,

Dept. of Statistics and Econometrics, Universidad Carlos III de Madrid, calle Madrid 126, 28903 Getafe, Madrid, Spain; jgonzalo@est-econ.uc3m.es; http:// halweb.uc3m.es/jgonzalo,

and

Dept. of Economics and Economic History, Universitat Autonoma de Barcelona, 08193 Bellaterra (Cerdanyola del Valles), Barcelona, Spain; Laura.Mayoral@uab.es; http://selene.uab.es/_cs_u_fonaments/caste/profespe/mayoral.htm.

APPENDIX A

DEFINITION 1 (*Stationary ARFIMA*(p, d, q) *process*): y_t is an *invertible and stationary ARFIMA*(p, d, q) *process* if it can be written as

(A.1) $\qquad \Delta^d y_t = u_t,$

where $\{u_t\}_{t=-\infty}^{\infty}$ is an ARMA(p, q) process such that $u_t = \Phi_p(L)^{-1}\Theta_q(L)\varepsilon_t$ and both lag polynomials have their roots outside the unit circle, where, ε_t is a zero-mean i.i.d. random variable with variance σ^2 and $d \in (-0.5, 0.5)$.

DEFINITION 2 (*Asymptotically stationary ARFIMA*(p, d, q) *process*): Let u_t^* be generated by

$$u_t^* = u_t, \qquad t > 0,$$
$$u_t^* = 0, \qquad t \leq 0,$$

where $\{u_t\}_{t=-\infty}^{\infty}$ verifies the same properties as in Definition 1. Then, y_t is an *asymptotically stationary ARFIMA*(p, d, q) *process* if it can be written as

(A.2) $\qquad \Delta^d y_t = u_t^*,$

with $d \in (-0.5, 0.5)$.

DEFINITION 3 (*Nonstationary ARFIMA*(p, d, q) *process*): Let y_t be an ARFIMA(p, d^*, q) defined as in (A.1). Then, if x_t is defined as

$$x_t = \sum_{i=1}^{t} y_i = \sum_{i=1}^{t}\left(\sum_{j=0}^{\infty} \pi_j(d^*)u_{i-j}\right),$$

it is a *nonstationary ARFIMA*(p, d, q) *process* with $d = 1 + d^*$, i.e., $d \in (0.5, 1.5)$.

DEFINITION 4 (*Nonstationary (truncated) ARFIMA*(p, d, q) *process*): Let y_t be an ARFIMA(p, d^*, q) as defined in (A.2). Then, if x_t is defined as

$$x_t = \sum_{i=1}^{t} y_i = \sum_{i=1}^{t}\left(\sum_{j=0}^{i-1} \pi_j(d^*)u_{i-j}\right),$$

it is a *nonstationary (truncated) ARFIMA*(p, d, q) *process* with $d = 1 + d^*$, i.e., $d \in (0.5, 1.5)$.

Consider the following sequences of partial sums of stationary and asymptotically stationary ARFIMA processes:

$$S_{[Tr]} = \sum_{i=1}^{[Tr]} y_i, \qquad r \in (0, 1],$$

with y_t defined as in (A.1), and

$$S^*_{[Tr]} = \sum_{i=1}^{[Tr]} y^*_i, \qquad r \in (0, 1],$$

with y^*_t defined as in (A.2). Then, it can be proved (see Marinucci and Robinson (1999)) that

$$c^{-1/2} T^{-d-1/2} S_{[Tr]}(r) \xrightarrow{w} B_d(r), \qquad r \in (0, 1],$$

with c a constant such that $\mathrm{var}(S_T) \sim c T^{2(d+1/2)}$, and

$$\Gamma(d+1) \left\{ \frac{d+1/2}{T^{2d-1}\pi g(0)} \right\}^{1/2} S^*_{[Tr]} \xrightarrow{w} W_d(r), \qquad r \in (0, 1],$$

where $g(0)$ is the spectral density of u_t at frequency zero, and $B_d(r)$ and $W_d(r)$ are Fractional Brownian Motions (FBM) as defined in Samorodnitsky and Taqqu (1994) and Mandelbrot and Van Ness (1968), respectively (see also Marinucci and Robinson (1999)).

PROOF OF LEMMA 1: *First Part of the Lemma*: Let us express x_t as

(A.3) $\qquad x_t = x^*_t + \xi_t,$

where ξ_t is given by

(A.4) $\qquad \xi_t = \sum_{i=t}^{\infty} \pi_i(-d) \varepsilon_{t-1}.$

In order to prove (11), it is enough to show that the second moment of $(1/T)\sum_{t=1}^{T} \xi_t$ converges to zero, namely,

(A.5) $\qquad \frac{1}{T^2} E\left(\left(\sum_{t=1}^{T} \xi_t \right)^2 \right) = \frac{\sigma^2}{T^2} \sum_{t=1}^{T} \left(\sum_{i=t}^{\infty} \pi_i^2(-d) \right).$

Since $\pi_i^2(-d) = i^{2(d-1)}$ for large i, $\sum_{i=1}^{\infty} \pi_i^2(-d)$ is summable. Taking into account that $\sum_{i=t}^{\infty} \pi_i^2(-d) = O(t^{2d-1})$ and that

(A.6) $\qquad \frac{\sigma^2}{T^2} \sum_{t=1}^{T} t^{2d-1} = O\left(\frac{T^{2d}}{T^2} \right),$

it is straightforward to see that the left side of (11) tends to zero. To prove (12), we can rewrite it as

(A.7) $\qquad \frac{1}{T} \sum_{t=1}^{T} (x_t^2 - x_t^{2*}) = \frac{1}{T} \left(\sum_{t=1}^{T} \xi_t^2 + 2 \sum_{t=1}^{T} x_t^* \xi_t \right),$

whose first moment converges to zero since

$$\frac{1}{T} E\left(\sum_{t=1}^{T} \xi_t^2 + 2 \sum_{t=1}^{T} x_t^* \xi_t \right) = \frac{1}{T} \sum_{t=1}^{T} E(\xi_t^2) = O(T^{2d-1}).$$

To show that the second-order moment also converges to zero, we first apply Loève's inequality (see Davidson (1994, p. 140)) to the first term of the right side of (A.7):

(A.8) $\qquad \frac{1}{T^2} E\left(\left(\sum_{t=1}^{T} \xi_t^2 \right)^2 \right) \leq \frac{1}{T} \sum_{t=1}^{T} E(\xi_t^4)$

$$\leq \frac{E(\varepsilon_t^4)}{T} \sum_{t=1}^{T} \left(\sum_{i=t}^{\infty} \pi_i^4(-d) \right) + \frac{6\sigma^4}{T} \sum_{t=1}^{T} \left(\sum_{i=t}^{\infty} \left(\sum_{j=i+1}^{\infty} \pi_i^2(-d)\pi_j^2(-d) \right) \right).$$

Since $\pi_i^4(-d) = i^{4(d-1)}$, then $\sum_{i=t}^{\infty} \pi_i^4(-d) = O(t^{4d-3})$ and therefore $\sum_{t=1}^{T} t^{4d-3} = O(1)$, implying that the first term in (A.8) converges to zero. With regard to the second term in (A.8)

$$(A.9) \qquad \sum_{t=1}^{T}\left(\sum_{i=t}^{\infty}\pi_i^2(-d)\left(\sum_{j=i+1}^{\infty}\pi_j^2(-d)\right)\right) \leq \sum_{t=1}^{T}\left(\sum_{i=t}^{\infty}\pi_i^2(-d)\sum_{j=t}^{\infty}\pi_j^2(-d)\right)$$
$$\leq \sum_{t=1}^{T}\left(\sum_{i=t}^{\infty}\pi_i^2(-d)\right)^2.$$

Since $\sum_{i=t}^{\infty}\pi_i^2(-d)^2 = O(t^{2(2d-1)})$, then $T^{-1}\sum_{t=1}^{T} t^{4d-2} = O(T^{-1})$ if $d < 1/4$ and $O(T^{4d-2})$ if $d \geq 1/4$. From (A.9) it is clear that the second term in (A.8) also tends to zero. Secondly, applying Loève's inequality to the second term of the right side of (A.7), we get

$$(A.10) \qquad T^{-2}E\left(\left(\sum_{t=1}^{T}x_t^*\xi_t\right)^2\right) \leq T^{-1}\left(\sum_{t=1}^{T}E(x_t^{*2}\xi_t^2)\right) = T^{-1}\left(\sum_{t=1}^{T}E(x_t^{*2})E(\xi_t^2)\right),$$

where the last inequality holds because of the independence between x_t^{*2} and ξ_t^2. Since $E(x_t^{*2}) = O(1)$ and $E(\xi_t^2) = O(t^{2d-1})$, it is straightforward to prove that expression (A.10) tends to zero, leading to the final required result. The proof of (13) is analogous and therefore is omitted.

Second Part of the Lemma: In order to prove the convergence in (14), consider

$$(A.11) \qquad (T\log T)^{-1}\sum_{t=1}^{T}E(z_t^{*2}) = (T\log T)^{-1}\frac{\sigma^2}{\pi}\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\frac{\Gamma^2(i+0.5)}{\Gamma^2(i+1)}\right).$$

Since $(T\log T)^{-1}\sum_{t=1}^{T}(\sum_{i=0}^{t-1}\frac{\Gamma^2(i+0.5)}{\Gamma^2(i+1)}) \to 1$, then $(T\log T)^{-1}\sum_{t=1}^{T}E(z_t^{*2}) \to \sigma^2/\pi$. To show that the variance of $(T\log T)^{-1}\sum_{t=1}^{T}z_t^{*2}$ converges to zero, notice that

$$(A.12) \qquad \mathrm{var}\left((T\log T)^{-1}\sum_{t=1}^{T}z_t^{*2}\right) = (T\log T)^{-2}\left(\sum_{t=1}^{T}\mathrm{var}(z_t^{*2}) + 2\sum_{t=2}^{T}\sum_{j=1}^{t-1}\mathrm{cov}(z_t^2, z_j^2)\right),$$

and

$$(A.13) \qquad \sum_{t=1}^{T}\mathrm{var}(z_t^{*2}) = \left(E(\varepsilon_t^4)\sum_{t=1}^{T}\left(\sum_{i=0}^{t}\pi_i^4(-0.5) + 6\sigma^4\sum_{i=1}^{t-1}\sum_{j=0}^{i-1}\pi_i^2(-0.5)\pi_j^2(-0.5)\right) - T\frac{\sigma^4}{\pi^2}\right)$$
$$= O(T) + O(T(\log T)^2) - O(T) = O(T(\log T)^2).$$

This implies that $(T\log T)^{-2}\sum_{t=1}^{T}\mathrm{var}(z_t^{*2}) \to 0$. It is easy to check, by applying the Cauchy-Schwartz inequality to the second term on the right side of (A.12), that

$$(A.14) \qquad 2\sum_{t=2}^{T}\sum_{j=1}^{t-1}\mathrm{cov}(z_t^2, z_j^2) = O(T(\log T)^2).$$

(A.13) and (A.14) imply the desired result.

Third Part of the Lemma: The result in (15) has already been proved in the literature (see Gourieroux, Maurel, and Monfort (1989)). *Q.E.D.*

PROOF OF LEMMA 2: The proof of the first and second part of this lemma is based on checking that the sequence $\{x_{t-1}\varepsilon_t\}$ verifies the conditions of the standard Central Limit Theorem (CLT) for martingale difference sequences (m.d.s.) (see Hall and Heyde (1980, Chapter 3), 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.

A. $0 < d < 0.5$: Define

(A.15) $\qquad \tilde{\varepsilon}_t = \sigma^{-1}\varepsilon_t,$

(A.16) $\qquad \tilde{x}_t = \left(\sigma^2 \dfrac{\Gamma(1-2d)}{\Gamma^2(1-d)}\right)^{-1/2} x_t,$

and

(A.17) $\qquad X_{T\cdot t} = T^{-1/2}\tilde{x}_{t-1}\tilde{\varepsilon}_t.$

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

(A.18) $\qquad T^{-1/2}E\big(\tilde{x}_{t-1}\tilde{\varepsilon}_t|F_{T\cdot t-1}\big) = T^{-1/2}E\big(\tilde{x}_{t-1}|F_{T\cdot t-1}\big)E\big(\tilde{\varepsilon}_t|F_{T\cdot t-1}\big) = 0.$

Regarding condition (ii), we have

(A.19) $\qquad T^{-1}\sum \mathrm{var}\big(\tilde{x}_{t-1}\tilde{\varepsilon}_t|F_{T\cdot t-1}\big) = T^{-1}\sum\left(E\big(\tilde{x}_{t-1}^2\tilde{\varepsilon}_t^2|F_{T\cdot t-1}\big) - E\big(\tilde{x}_{t-1}\tilde{\varepsilon}_t|F_{T\cdot t-1}\big)^2\right)$

$$= T^{-1}\sum\big(E\big(\tilde{x}_{t-1}^2|F_{T\cdot t-1}\big)E\big(\tilde{\varepsilon}_t^2|F_{T\cdot t-1}\big)\big) = T^{-1}\sum \tilde{x}_{t-1}^2 \xrightarrow{p} 1.$$

Finally, condition (iii) holds since

(A.20) $\qquad \displaystyle\sum_{t=1}^{T} E\big(|X_{T\cdot t}|^2 I\{|X_{T\cdot t}| > \varrho|\}\big) = E\big(|\tilde{x}_{t-1}\tilde{\varepsilon}_t|^2 I_{\{|\tilde{x}_{t-1}\tilde{\varepsilon}_t| > T^{1/2}\varrho|\}}\big) \longrightarrow 0, \quad \text{for all} \quad \varrho > 0.$

Conditions (A.18), (A.19), and (A.20) 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 $T^{-1}(\sum(\tilde{x}_{t-1}^*)^2 - \sum(\tilde{x}_{t-1})^2) = o_p(1)$ (see Lemma 1) implies $T^{-1}\sum(\tilde{x}_{t-1}^*)^2 \xrightarrow{p} 1$. Lastly, a sufficient condition for condition (iii) is Liapunov's condition, $1/T^2\sum_{t=1}^{T}E(\tilde{x}_{t-1}^4\tilde{\varepsilon}_t^4) \to 0$. To prove this, consider

(A.21) $\qquad \dfrac{1}{T^2}\displaystyle\sum_{t=1}^{T}E\big(\tilde{x}_{t-1}^4\tilde{\varepsilon}_t^4\big) = \dfrac{1}{T^2}E\big(\tilde{\varepsilon}_t^4\big)\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\pi_i^4(-d) + 6\sum_{i=1}^{t-1}\sum_{j=0}^{i-1}\pi_i^2(-d)\pi_j^2(-d)\right).$

Noticing that $\pi_i(-d) = i^{d-1}$, it is easy to check that

$$\sum_{i=0}^{T-1}\pi_i^4(-d) + 6\sum_{i=1}^{T-1}\sum_{j=0}^{i-1}\pi_i^2(-d)\pi_j^2(-d) = O(1)$$

and therefore $T^{-1}(\sum_{i=0}^{T-1}\pi_i^4(-d) + 6\sum_{i=1}^{T-1}\sum_{j=0}^{i-1}\pi_i^2(-d)\pi_j^2(-d)) \to 0$. This implies that expression (A.21) tends to zero as desired.

B. $d = 0.5$: Now let us define

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

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

and

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

Condition (i) is satisfied by following the same arguments as in the previous case. Condition (ii), follows from (A.19),

$$T^{-1}\sum_{1}^{T}\mathrm{var}(\tilde{x}_{t-1}^*\tilde{\varepsilon}_t|F_{T\cdot t-1}) = T^{-1}\sum^{T}\tilde{x}_{t-1}^{*2},$$

and the results in Tanaka (1999). From them, we know that $((\sum^{T}\tilde{x}_{t-1}^2)/T) \xrightarrow{p} 1$. Condition (iii) can be proved using a similar argument as in the previous part of this lemma.

C. $-0.5 < d < 0$: The convergence for the truncated and nontruncated processes has been proved by Gourieroux, Maurel, and Monfort (1989) and by Davidson and de Jong (2001), respectively.   Q.E.D.

PROOF OF THEOREM 1: Under the null, $\Delta^{d_1} y_t \sim FI(1 - d_1)$. Hence, for values of $d_1 \in [0, 0.5]$ the process will be nonstationary whereas it will be (asymptotically) stationary when $d_1 \in (0.5, 1]$. Applying Lemmae 1 and 2 and results in Tanaka (1999) and in Gourieroux, Maurel, and Monfort (1989), we obtain

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

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

$$T^{1/2}\hat{\phi}_{ols} = \frac{T^{(-1/2)} \sum \varepsilon_t \Delta^{d_1} y_{t-1}}{T^{-1} \sum (\Delta^{d_1} y_{t-1})^2} \xrightarrow{w} N\left(0, \frac{\Gamma^2(d_1)}{\Gamma(2d_1 - 1)}\right) \quad \text{if} \quad 0.5 < d_1 < 1. \qquad \text{Q.E.D.}$$

PROOF OF THEOREM 2: The proof of this theorem is similar to that of Theorem 1, and therefore is omitted.

PROOF OF THEOREM 3: For the sake of brevity, we present the proof of this theorem for the t-ratio defined in (9). A similar proof remains valid for the OLS estimator defined in (8).

Under DGP (29), $\Delta y_t \sim FI(d_1^* - 1)$ and $\Delta^{d_1} y_{t-1} \sim FI(d_1^* - d_1)$. According to the sign of $d_1^* - d_1$, the process $\Delta^{d_1} y_{t-1}$ can be asymptotically stationary with intermediate memory ($d_1^* - d_1 < 0$), short memory ($d_1^* - d_1 = 0$), asymptotically stationary with long memory ($0 < d_1^* - d_1 < 0.5$) or nonstationary ($d_1^* - d_1 \ge 0.5$). Therefore, the asymptotic behavior of the FD-F test will depend upon the value of d_1 that is used in the regression. Let us consider separately each of the following cases:

1. $d_1^* - d_1 < 0$: In this case both Δy_t and $\Delta^{d_1} y_{t-1}$ are asymptotically stationary processes. Proceeding in a similar way to the proof of Lemma 1, it is easy to prove that

$$T^{-1} \sum_{t=2}^{T} (\Delta y_t \Delta^{d_1} y_t - \Delta y_t^s \Delta^{d_1} y_t^s) = o_p(1),$$

where Δy_t^s and $\Delta^{d_1} y_t^s$ are the corresponding nontruncated stationary processes associated to Δy_t and $\Delta^{d_1} y_t$, respectively, which are defined as follows:

(A.22) $$\Delta y_t^s = \Delta y_t + \sum_{i=t}^{\infty} \pi_i (1 - d_1^*) \varepsilon_{t-i}$$

and

$$\Delta^{d_1} y_t^s = \Delta^{d_1} y_t + \sum_{i=t}^{\infty} \pi_i (d_1 - d_1^*) \varepsilon_{t-i}.$$

That allows us to apply the Law of Large Numbers (LLN) for stationary and ergodic processes to both the denominator and numerator of the t-ratio $t_{\hat{\phi}_{ols}}$. The denominator is $O_p(T^{1/2})$ since S_T is strictly positive and $O_p(1)$, and $\left(\sum (\Delta^{d_1} y_{t-1}^s)^2\right)^{1/2} = O_p(T^{1/2})$. The numerator is $O_p(T)$ since

(A.23) $$\frac{\sum \Delta y_t \Delta^{d_1} y_{t-1}}{T} \xrightarrow{p} \text{cov}(\Delta y_t^s, \Delta^{d_1} y_{t-1}^s).$$

Therefore the t-ratio $t_{\hat{\phi}_{ols}}$ diverges. It remains to be proved that it diverges to $-\infty$. Because the denominator is always positive, the sign of the ratio is determined by the sign of the numerator,

(A.24) $$\text{cov}(\Delta y_t^s, \Delta^{d_1} y_{t-1}^s) = \text{cov}(\Delta^{1-d_1^*} \varepsilon_t, \Delta^{d_1 - d_1^*} \varepsilon_{t-1})$$
$$= \sigma^2 (\pi_1 + \pi_2 \pi_1 (d_1 - d_1^*) + \pi_3 \pi_2 (d_1 - d_1^*) + \cdots).$$

All the coefficients π_i and $\pi_i(d_1 - d_1^*)$ are negative for $i \geq 1$; therefore

$$\pi_2 \pi_1 (d_1 - d_1^*) + \pi_3 \pi_2 (d_1 - d_1^*) + \cdots > 0.$$

To check that (A.24) is negative, notice that

(A.25) $$\pi_2 \pi_1 (d_1 - d_1^*) + \pi_3 \pi_2 (d_1 - d_1^*) + \cdots \leq \sup_{j \in [2,\infty)} (-\pi_j) \sum_{i=1}^{\infty} -\pi_i (d_1 - d_1^*) = -\pi_2 < -\pi_1.$$

Therefore, the covariance in (A.24) is negative and the t-ratio $t_{\hat{\phi}_{ols}}$ diverges to $-\infty$.

2. $d_1^* - d_1 = 0$: The covariance in (A.24) is also a negative quantity since

(A.26) $$\frac{\sum \Delta y_t \Delta^{d_1} y_{t-1}}{T} = \frac{\sum \Delta y_t \varepsilon_{t-1}}{T} \xrightarrow{p} \pi_1 (1 - d_1) = -(1 - d_1).$$

Applying the same argument for the denominator as in the previous case, it is easy to check that the t-ratio $t_{\hat{\phi}_{ols}}$ diverges to $-\infty$.

3. $0 < d_1^* - d_1 < 0.5$: Proceeding in an analogous way as above, it is easy to show that the denominator of $t_{\hat{\phi}_{ols}}$ is positive and that the covariance in (A.24) is negative.

4. $d_1^* - d_1 > 0.5$: In this case, $\Delta^{d_1} y_t$ is a nonstationary process. We can write $t_{\hat{\phi}_{ols}}$ as

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

To prove that $t_{\hat{\phi}_{ols}}$ diverges it is sufficient to show that

(A.27) $$\sum_{t=2}^{T} (\Delta y_{t-1})^2 = O_p(T),$$

(A.28) $$\sum_{t=2}^{T} (\Delta^{d_1} y_{t-1})^2 = O_p(T^{2(d_1^* - d_1)}),$$

and

(A.29) $$\sum_{t=2}^{T} \Delta y_t \Delta^{d_1} y_{t-1} = O_p(T).$$

Expressions (A.27) and (A.28) follow immediately. With respect to (A.29), it follows from Theorem 3 in Dolado and Mármol (1999) and some algebra that

(A.30) $$T^{-1} \sum_{t=2}^{T} \Delta y_t \Delta^{d_1} y_{t-1} \xrightarrow{p} \sum_{j=0}^{\infty} \pi_j (1 - d_1) z_j = O(1),$$

where

(A.31) $$z_j = \sigma^2 \sum_{i=0}^{\infty} \sum_{k=1}^{\infty} \pi_i (1 - d_1^*) \pi_{i+k+j} (1 - d_1^*).$$

The previous probabilistic orders of magnitude imply that the t-statistic diverges. It remains to check that it diverges towards $-\infty$. For that, we need to show first that $z_j < 0$ and that $|z_j| > |z_{j+1}|$ for all $j \geq 0$. With respect to the former inequality, notice that from (A.31), it follows that

(A.32) $$z_j = \sigma^2 \left(\sum_{k=1}^{\infty} \pi_{k+j} (1 - d_1^*) + \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} \pi_i (1 - d_1^*) \pi_{i+k+j} (1 - d_1^*) \right).$$

Since $\pi_j(1-d_1^*) < 0$ for all $j \geq 1$, the first term in (A.32) will be negative and the second one is positive. Now, because

(A.33) $\qquad \left| \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} \pi_i(1-d_1^*)\pi_{i+k+j}(1-d_1^*) \right| \leq \left(\sum_{i=1}^{\infty} |\pi_i(1-d_1^*)| \right) \sum_{k=1}^{\infty} |\pi_{1+k+j}(1-d_1^*)|,$

it is clear that $\sum_{k=1}^{\infty} |\pi_{k+j}(1-d_1^*)| > \sum_{k=1}^{\infty} |\pi_{1+k+j}(1-d_1^*)|$, which in turn implies $z_j < 0$. Since both z_j and z_{j+1} are negative, to prove the second inequality, $|z_j| > |z_{j+1}|$ for all $j \geq 0$, it is equivalent to show that $z_j - z_{j+1} < 0$. Consider

(A.34) $\qquad z_j - z_{j+1} = \pi_{j+1}(1-d_1^*) + \sum_{i=1}^{\infty}\sum_{k=1}^{\infty}[\pi_{i+k+j}(1-d_1^*) - \pi_{i+k+j+1}(1-d_1^*)]$

(A.35) $\qquad = \pi_{j+1}(1-d_1^*) + \sum_{i=1}^{\infty} \pi_i(1-d_1^*)\pi_{i+j+1}(1-d_1^*) < 0,$

where the last inequality follows from noticing that

$$\pi_{j+1}(1-d_1^*) < 0,$$

$$\sum_{i=1}^{\infty} \pi_i(1-d_1^*)\pi_{i+j+1}(1-d_1^*) > 0,$$

but

$$\sum_{i=1}^{\infty} \pi_i(1-d_1^*)\pi_{i+j+1}(1-d_1^*) \leq -\pi_{j+2}(1-d_1^*).$$

Finally, to prove that the probability limit in (A.30) is negative, notice that

(A.36) $\qquad \sum_{j=0}^{\infty} \pi_j(1-d_1)z_j \leq z_0 + z_1 \sum_{i=1}^{\infty} \pi_i(1-d_1) = z_0 - z_1 < 0.$

Thereby, the test statistic diverges to $-\infty$ and the test is consistent. $\qquad\qquad$ Q.E.D.

PROOF OF THEOREM 4: The t-ratio $t_{\hat{\phi}_{ols}}$, computed from the regression of Δy_t on $\Delta^{1+\delta/\sqrt{T}} y_{t-1}$ is given by

(A.37) $\qquad t_{\hat{\phi}_{ols}} = T^{1/2} \dfrac{\sum \Delta y_t \Delta^{1+\delta/\sqrt{T}} y_{t-1}}{\left(\sum(\Delta y_t)^2 \sum\left(\Delta^{1+\delta/\sqrt{T}} y_{t-1}\right)^2 - \left(\sum \Delta y_t \Delta^{1+\delta/\sqrt{T}} y_{t-1}\right)^2 \right)^{1/2}}.$

Assuming that $\Delta^{1+\delta/\sqrt{T}} y_t = \varepsilon_t 1_{(t>0)}$, expression (A.37) can be rewritten as

(A.38) $\qquad t_{\hat{\phi}_{ols}} = T^{1/2} \dfrac{\sum \Delta^{-\delta/\sqrt{T}} \varepsilon_t 1_{(t>0)} \varepsilon_{t-1}}{\left(\left(\sum(\Delta^{\delta/\sqrt{T}}\varepsilon_t)^2\right)\left(\sum(\varepsilon_{t-1})^2\right) - \left(\sum \Delta^{-\delta/\sqrt{T}}\varepsilon_t 1_{(t>0)}\varepsilon_{t-1}\right)^2 \right)^{1/2}}.$

The denominator of (A.38) is $O_p(T)$, and divided by T, converges in probability to σ^2 (see Tanaka (1999)). As for the numerator

(A.39) $\qquad T^{-1/2} \sum \Delta^{-\delta/\sqrt{T}} \varepsilon_t 1_{(t>0)} \varepsilon_{t-1}$

$$= T^{-1/2}\left(\sum_{t=2}^{T} \varepsilon_t \varepsilon_{t-1} + \pi_1(-\delta/\sqrt{T}) \sum_{t=3}^{T} \varepsilon_{t-1}^2 + \sum_{j=2}^{T-1} \pi_j(-\delta/\sqrt{T}) \sum_{t=j+1}^{T} \varepsilon_{t-j}\varepsilon_{t-1} \right).$$

The last term in the right side of (A.39) converges to zero in mean square:

$$T^{-1}E\left(\sum_{j=2}^{T-1}\pi_j(-\delta/\sqrt{T})\sum_{t=j+1}^{T}\varepsilon_{t-j}\varepsilon_{t-1}\right)^2 = T^{-1}E\left(\sum_{j=2}^{T-1}\pi_j^2(-\delta/\sqrt{T})\sum_{t=j+1}^{T}\varepsilon_{t-j}^2\varepsilon_{t-1}^2\right)$$

$$= T^{-1}\sigma^4\sum_{j=2}^{T-1}\pi_j^2(-\delta/\sqrt{T})(T-j)$$

$$= \sigma^4 T^{-1}O(1) - T^{-2}O(T^{3/2}) \longrightarrow 0,$$

since $\pi_j(-\delta/\sqrt{T}) = \delta/(j\sqrt{T}) + O(1/T)$. Therefore (A.39) can be rewritten as

$$(A.40)\qquad T^{-1/2}\sum\Delta^{-\delta/\sqrt{T}}\varepsilon_t 1_{(t>0)}\varepsilon_{t-1} = T^{-1/2}\sum\varepsilon_t\varepsilon_{t-1} + \delta T^{-1}\sum\varepsilon_{t-1}^2 + o_p(1).$$

Applying the corresponding CLT and the LLN for m.d.s. to expression (A.40) and given that the denominator of (A.38) converges in probability to σ^2, then

$$t_{\hat{\phi}_{ols}}(1+\delta/\sqrt{T})|_{H_1: d=1+\delta/\sqrt{T}} \xrightarrow{w} N(\delta,1).$$

<div style="text-align:right">Q.E.D.</div>

PROOF OF THEOREM 5: Since d_T is a consistent estimator of d ($=1$), it follows that if $\hat{d}_1$ is chosen according to the criterion function in (33), $\hat{d}_1 \xrightarrow{p} 1-c$. In order to prove that $(t_{\hat{\phi}_{ols}}(\hat{d}_1) - t_{\hat{\phi}_{ols}}(1-c))$ is $o_p(1)$ we apply the mean value theorem (MVT) on $t_{\hat{\phi}_{ols}}$ around the point $(1-c)$:

$$(A.41)\qquad t_{\hat{\phi}_{ols}}(\hat{d}_1) = t_{\hat{\phi}_{ols}}(1-c) + \frac{\partial t_{\hat{\phi}_{ols}}(\check{d})}{\partial d}(\hat{d}_1 - (1-c)),$$

where $\check{d}$ is an intermediate point between $\hat{d}_1$ and $(1-c)$. Thus, $\check{d} \xrightarrow{p} (1-c)$. In order to prove $(t_{\hat{\phi}_{ols}}(\hat{d}_1) - t_{\hat{\phi}_{ols}}(1-c)) = o_p(1)$, from (A.41), it suffices to show $\partial t_{\hat{\phi}_{ols}}(\check{d})/\partial d = o_p(T^{1/2})$, since $(\hat{d}_1 - (1-c)) = O_p(T^{-1/2})$. The t-ratio $t_{\hat{\phi}_{ols}}$ can be rewritten as

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

and under the null hypothesis

$$t_{\hat{\phi}_{ols}}(d) = T^{1/2}\frac{\sum\varepsilon_t\Delta^{d-1}\varepsilon_{t-1}1_{(t-1>0)}}{\left(\sum(\varepsilon_t)^2\sum\left(\Delta^{d-1}\varepsilon_{t-1}1_{(t-1>0)}\right)^2 - \left(\sum\varepsilon_t\Delta^{d-1}\varepsilon_{t-1}1_{(t-1>0)}\right)^2\right)^{1/2}}.$$

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

$$(A.42)\qquad \frac{\partial t_{\hat{\phi}_{ols}}(d)}{\partial d}\bigg|_{d=1-c}$$

$$= T^{1/2}\frac{\left(\sum\varepsilon_t\left(\log(1-L)\Delta^{-c}\varepsilon_{t-1}1_{(t-1>0)}\right)\right)A_1 - \left(\sum\varepsilon_t\Delta^{-c}\varepsilon_{t-1}1_{(t-1>0)}\right)A_2}{A_1^2}\bigg|_{d=1-c},$$

where

$$(A.43)\qquad A_1|_{d=1-c} = \left(\left(\sum\varepsilon_t^2\right)\left(\sum\Delta^{-c}\varepsilon_{t-1}^2 1_{(t-1>0)}\right) - \left(\sum\varepsilon_t\Delta^{-c}\varepsilon_{t-1}1_{(t-1>0)}\right)^2\right)^{1/2},$$

and

$$(A.44)\qquad A_2 = \partial A_1/\partial d|_{d=1-c}.$$

Taking into account that $c < 0.5$, then $\sum \Delta^{-c} \varepsilon_{t-1}^2 1_{(t-1>0)} = O_p(T)$ and $\sum \varepsilon_t \Delta^{-c} \varepsilon_{t-1} = o_p(T)$, and therefore A_1 is $O_p(T)$. As regards A_2, notice that it can be written as

$$(A.45) \qquad A_2 = A_1^{-1}\left(\left(\sum \varepsilon_t^2\right)\sum(\log(1-L)\Delta^{-c}\varepsilon_{t-1}1_{(t-1>0)})\Delta^{-c}\varepsilon_{t-1}\right)$$
$$- A_1^{-1}\left(\left(\sum \varepsilon_t \Delta^{-c}\varepsilon_{t-1}\right)\sum \varepsilon_t \Delta^{-c}\varepsilon_{t-1}1_{(t-1>0)}\log(1-L)\right).$$

Since the function $\log(1-z)$ is analytic in the convergence disc $|z| < 1$, the expansion $\log(1-z) = -(z + 1/2z^2 + 1/3z^3 + \cdots)$ applies and therefore it follows that $A_2 = O_p(T)$. The first element of the numerator in the right side of (A.42), namely $(\sum \varepsilon_t \Delta^{-c}\varepsilon_{t-1}\log(1-L)) = o_p(T)$, since divided by T, converges in mean square to zero. To check this last statement, notice that

$$(A.46) \qquad T^{-1}\sum_{t=1}^{T}\varepsilon_t\Delta^{-c}\varepsilon_{t-1}1_{(t-1>0)}\log(1-L) = T^{-1}\sum_{j=1}^{T-2}\frac{1}{j}\left(\sum_{t=j+2}^{T}\varepsilon_t\Delta^{-c}\varepsilon_{t-1-j}1_{(t-1-j>0)}\right).$$

To simplify notation, let us call $a_j = \sum_{t=j+2}^{T}\varepsilon_t\Delta^{-c}\varepsilon_{t-1-j}1_{(t-1-j>0)}$. Then,

$$(A.47) \qquad T^{-2}E\left(\sum_{j=1}^{T-2}\frac{1}{j}a_j\right)^2 = T^{-2}E\left(\sum_{j=1}^{T-2}\frac{1}{j^2}a_j^2 + 2\sum_{j=1}^{T-2}\sum_{k>j}^{T-2}\frac{a_ja_k}{jk}\right).$$

The first term of the right side of (A.47) converges to zero since

$$E(a_j^2) = E\left(\sum_{t=j+2}^{T}\varepsilon_t\Delta^{-c}\varepsilon_{t-1-j}1_{(t-1-j>0)}\right)^2 = \sum_{t=j+2}^{T}E(\varepsilon_t^2)E\left(\Delta^{-c}\varepsilon_{t-1-j}1_{(t-1-j>0)}\right)^2$$

$$= \sigma^2\sum_{t=0}^{T}\left(\sum_{i=0}^{t}\pi_i(-c)^2\right) = \sigma^2\sum_{t=0}^{T}(T-t+1)\pi_t(-c)^2$$

$$= T\sum_{t=0}^{T}\pi_t(-c)^2 - \sum_{t=0}^{T}t\pi_t(-c)^2 + \sum_{t=0}^{T}\pi_t(-c)^2 = O(T^{3/2})$$

(see Gray and Zhang (1988) to check the intermediate equalities).

The second term of the right side of (A.47) also converges to zero since

$$T^{-2}\sum_{j=1}^{T-2}\sum_{k>j}^{T-2}\frac{E(a_ja_k)}{jk} \leq T^{-2}\sum_{j=1}^{T-2}\sum_{k=1}^{T-2}\frac{(E(a_j^2)E(a_k^2))^{1/2}}{jk} \leq T^{-2}E(a_j^2)T^{-2}\sum_{j=1}^{T-2}\frac{1}{j}\sum_{k=1}^{T-2}\frac{1}{k}$$

$$= O(T^{-2})O(T^{3/2})(\log T)^2 \longrightarrow 0.$$

Assembling these results, we get

$$(A.48) \qquad \left.\frac{\partial t_{\hat{\phi}_{ols}}(d)}{\partial d}\right|_{d=1-c} = T^{1/2}\frac{o_p(T)O_p(T) - O_p(T^{1/2})O_p(T)}{O_p(T^2)} = T^{1/2}o_p(1).$$

Therefore, the asymptotic distributions of $t_{\hat{\phi}_{ols}}(\hat{d}_1)$ and $t_{\hat{\phi}_{ols}}(1-c)$ are identical, namely, $t_{\hat{\phi}_{ols}}(\hat{d}_1) \xrightarrow{w} N(0,1)$.
$Q.E.D.$

PROOF OF THEOREM 6: Let $\widehat{\Phi}_{ols}$ be the OLS estimator of the vector of parameters $\Phi = (\zeta_1, \ldots, \zeta_p, \phi)'$ in regression (38):

$$(A.49) \qquad \Upsilon_T(\widehat{\Phi}_{ols} - \Phi) = \left(\Upsilon_T^{-1}Y'Y\Upsilon_T^{-1}\right)^{-1}\Upsilon_T^{-1}Y'\varepsilon_t,$$

where Y is the matrix of regressors

(A.50) $\qquad Y = (\Delta y_{t-1}, \quad \cdots, \quad \Delta y_{t-p}, \quad \Delta^{d_1} y_{t-1}),$

and Υ_T is a weighting matrix to be defined below. In order to analyze the asymptotic behavior of Φ_{ols}, we need to consider three different cases according to the value of d_1 that is used in the regression. This value determines the (asymptotic) stationary or nonstationary nature of $\{\Delta^{d_1} y_t\}$.

   1. $0 \le d_1 < 0.5$: Define the following weighting matrix:

(A.51) $\qquad \Upsilon_T = \begin{pmatrix} T^{1/2} I_p & 0 \\ 0 & T^{1-d_1} \end{pmatrix},$

where I_p is the $p \times p$ identity matrix. Applying the corresponding LLN for stationary ARMA processes and results in Gourieroux, Maurel, and Monfort (1989), it can be proved that the first term of the right side of (A.49), $(\Upsilon_T^{-1} Y' Y \Upsilon_T^{-1})$, converges in probability to

(A.52) $\qquad \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & & 0 \\ \gamma_1 & \cdots & & & \cdots \\ \cdots & & \gamma_0 & & 0 \\ 0 & \cdots & 0 & \sigma^2 \int_0^1 W_{-d_1}^{*2}(r)\, dr \end{pmatrix} = \begin{pmatrix} V & 0 \\ 0 & Q \end{pmatrix},$

where γ_j is the autocovariance function of u_t. With respect to the second term, $\Upsilon_T^{-1} Y' \varepsilon_t$, the first p elements satisfy the standard CLT for m.d.s. with a variance-covariance matrix given by

(A.53) $\qquad E\left\{ \begin{pmatrix} \Delta y_{t-1} \varepsilon_t \\ \cdots \\ \Delta y_{t-p} \varepsilon_t \end{pmatrix} (\Delta y_{t-1}, \varepsilon_t, \ldots, \Delta y_{t-p}, \varepsilon_t) \right\} = \sigma^2 \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \cdots & & \cdots \\ \cdots & & \gamma_0 & \cdots \\ \gamma_{p-1} & \cdots & \cdots & \gamma_0 \end{pmatrix} = \sigma^2 V.$

Thus

(A.54) $\qquad \begin{pmatrix} T^{-1/2} \sum \Delta y_{t-1} \varepsilon_t \\ \cdots \\ T^{-1/2} \sum \Delta y_{t-p} \varepsilon_t \end{pmatrix} \xrightarrow{w} h_1 = N_p(0, \sigma^2 V).$

The last element in $\Upsilon_T^{-1} Y' \varepsilon_t$ converges to (see Lemma 2)

(A.55) $\qquad T^{-(1-d_1)} \sum \Delta^{d_1} y_{t-1} \varepsilon_t \xrightarrow{w} h_2 = \sigma^2 \int_0^1 W_{-d_1}^*(r)\, dB.$

Putting together (A.52), (A.54), and (A.55), we get

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

Consistency of $\widehat{\Phi}_{ols}$ implies $\hat{\sigma}_{\hat{\phi}_{ols}} \xrightarrow{p} \sigma Q^{-1/2}$, and therefore the desired result for $t_{\hat{\phi}_{ols}} = \hat{\phi}_{ols} / \hat{\sigma}_{\hat{\phi}_{ols}}$ follows.

   2. $0.5 < d_1 < 1$: In this case, $\Delta^{d_1} y_t$ is a (asymptotically) stationary process. Consider again expression (A.49) where now $\Upsilon_T = T^{1/2} I_{p+1}$. In this case the LLN applies to all the elements of the first

term of the right side of (A.49), yielding

$$
\text{(A.56)} \quad \Upsilon_T^{-1} Y' Y \Upsilon_T^{-1} \xrightarrow{p}
\begin{pmatrix}
\gamma_0 & \gamma_1 & \cdots & \mathrm{cov}(\Delta y_{t-1}^s, \Delta^{d_1} y_{t-1}^s) \\
\gamma_1 & \cdots & & \cdots \\
\cdots & & \cdots & \mathrm{cov}(\Delta y_{t-p}^s, \Delta^{d_1} y_{t-1}^s) \\
\mathrm{cov}(\Delta y_{t-1}^s, \Delta^{d_1} y_{t-1}^s) & \cdots & \mathrm{cov}(\Delta y_{t-p}^s, \Delta^{d_1} y_{t-1}^s) & \mathrm{var}(\Delta^{d_1} y_{t-1}^s)
\end{pmatrix}
$$

$$
\text{(A.57)} \quad =
\begin{pmatrix}
Q_1 & Q_2 \\
Q_2' & Q_4
\end{pmatrix}.
$$

Applying both the corresponding CLT for m.d.s. and Lemma 2 to the second term of the right side of (A.49) yields

$$
\text{(A.58)} \quad
\begin{pmatrix}
T^{-1/2} \sum \Delta y_{t-1} \varepsilon_t \\
\cdots \\
T^{-1/2} \sum \Delta^d y_{t-1} \varepsilon_t
\end{pmatrix}
\xrightarrow{w} Z =
\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}
\sim N_{p+1}\left(0,\;
\begin{pmatrix}
\sigma^2 V & \sigma^2 Q_2 \\
\sigma^2 Q_2' & \sigma^2 \mathrm{var}(\Delta^{d_1} y_{t-1})
\end{pmatrix}
\right).
$$

Hence

$$
\text{(A.59)} \quad \Upsilon_T(\widehat{\Phi}_{ols} - \Phi) \xrightarrow{w}
\begin{pmatrix}
Q_1 & Q_2 \\
Q_2' & Q_4
\end{pmatrix}^{-1}
\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.
$$

Using the notation

$$
\begin{pmatrix}
b_1 & b_2 \\
b_2 & b_4
\end{pmatrix}
=
\begin{pmatrix}
Q_1 & Q_2 \\
Q_2' & Q_4
\end{pmatrix}^{-1},
$$

then

$$
\text{(A.60)} \quad T^{1/2}(\hat{\phi}_{ols} - \phi) \xrightarrow{w} b_2 z_1 + b_4 z_2.
$$

To simplify the presentation, we consider the case where $\alpha(L)$ is an AR(1) process. In that case

$$
\text{(A.61)} \quad b_2 = \frac{-\mathrm{cov}(\Delta y_{t-1}^s, \Delta^{d_1} y_{t-1}^s)}{\mathrm{var}(\Delta y_{t-1}^s)\,\mathrm{var}(\Delta^{d_1} y_{t-1}^s) - \mathrm{cov}^2(\Delta y_{t-1}^s, \Delta^{d_1} y_{t-1}^s)},
$$

and

$$
\text{(A.62)} \quad b_4 = \frac{\mathrm{var}(\Delta y_{t-1}^s)}{\mathrm{var}(\Delta y_{t-1}^s)\,\mathrm{var}(\Delta^{d_1} y_{t-1}^s) - \mathrm{cov}^2(\Delta y_{t-1}^s, \Delta^{d_1} y_{t-1}^s)}.
$$

Then, it is easy to show that

$$
\text{(A.63)} \quad T^{1/2}(\hat{\phi}_{ols} - \phi) \xrightarrow{w} N(0, \sigma^2 b_4).
$$

Again, since $\widehat{\Phi}_{ols}$ is consistent, so is S_T for σ, and thus the t-ratio converges to

$$
\text{(A.64)} \quad t_{\hat{\phi}_{ols}} \xrightarrow{w} N(0, \sigma^2 b_4)/\sigma b_4^{1/2} = N(0, 1).
$$

The more general case can be proved along a similar line of reasoning.

3. $d_1 = 0.5$: Define the following weighting matrix:

$$
\text{(A.65)} \quad \Upsilon_T =
\begin{pmatrix}
T^{1/2} I_p & 0 \\
0 & (T \log T)^{1/2}
\end{pmatrix}.
$$

The rest of the proof is analogous to the previous one and therefore is omitted. _Q.E.D._

PROOF OF THEOREM 7: The following proof is similar to the one proposed by Said and Dickey (1984) and Xiao and Phillips (1998) in the context of the AD-F test. Both are based upon Berk (1974).

The u_t process is a stationary and invertible ARMA with the following autoregressive representation:

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

where $\vartheta(L) = \alpha(L)/\beta(L)$. Applying the Beveridge-Nelson decomposition to $\vartheta(L)$ we get $\vartheta(L) = \vartheta(1) + \tilde{\vartheta}(L)(1-L)$, where the coefficients of $\tilde{\vartheta}(L)$ decay at an exponential rate. Thus, we can write y_t as follows:

(A.66) $\qquad \Delta y_t = -\Delta^{d_1} y_{t-1} + (1 - \Delta^{d_1})\Delta y_t + u_t.$

Multiplying (A.66) by $\vartheta(L)$ and rearranging terms,

(A.67) $\qquad \Delta y_t = -\vartheta(1)\Delta^{d_1} y_{t-1} + (1 - (\vartheta(1) + \tilde{\vartheta}(L))\Delta^{d_1})\Delta y_t + \varepsilon_t.$

Define $\Lambda(z) = (1 - (\vartheta(1) + \tilde{\vartheta}(z))(1-z)^{d_1})$. Notice that $\Lambda(0) = 0$, since $\vartheta(1) + \tilde{\vartheta}(0) = 1$. This result implies that the right-hand side of (A.67) does not contain the contemporaneous value of Δy_t, but only lagged values of Δy_t. Let $\{\delta_i\}_{i=1}^{\infty}$ be the coefficients in the expansion of polynomial $\Lambda(z)$ in increasing powers of L. Then, expression (A.67) can be rewritten as

$$\Delta y_t = \phi \Delta^{d_1} y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots + \varepsilon_t.$$

The rate at which the $\{\delta_i\}$ coefficients decay is given by the rate at which the coefficients on $(1-z)^{d_1}$ decay, since the latter are those with a slower rate of decay in the polynomial $\Lambda(z)$. Hence, for large j, $\delta \sim j^{-d_1 - 1}$. This implies that those coefficients are absolutely summable for all $d_1 > 0$.

Next, we show that the OLS estimators of the parameters in the regression model $\Delta y_t = \phi \Delta^{d_1} y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_k \Delta y_{t-k} + \xi_{tk}$ (where only k lags of Δy_t have been included, and the remaining terms appear in the disturbance, ξ_{tk}) are consistent, when k grows to infinity (as $T \to \infty$) at a chosen rate. Define

$$X_t' = (\Delta y_{t-1}, \dots, \Delta y_{t-k}),$$
$$U_t' = (\Delta^{d_1} y_{t-1} : X_t'),$$

and

$$R_0 = \sum_{t=k+1}^{T} U_t U_t' = \begin{pmatrix} \sum (\Delta^{d_1} y_{t-1})^2 & \sum \Delta^{d_1} y_{t-1} X_t' \\ \sum \Delta^{d_1} y_{t-1} X_t & \sum X_t X_t' \end{pmatrix}.$$

Then,

$$(\hat{\phi}, \hat{\delta}_1, \dots, \hat{\delta}_k)' - (\phi, \delta_1, \dots, \delta_k)' = R_0^{-1} \sum_{t=k+1}^{T} U_t \xi_{tk}.$$

To show that $\sum_{t=k+1}^{T} U_t \xi_{tk}/(T-k) = o_p(1)$, consider

(A.68) $\qquad (T-k)^{-2} \left\| \sum_{t=k+1}^{T} U_t \xi_{tk} \right\|^2 = (T-k)^{-2} \left(\left(\sum_t \Delta^d y_{t-1} \xi_{tk} \right)^2 + \left(\sum_{j=1}^{k} \sum_t \Delta y_{t-j} \xi_{tk} \right)^2 \right).$

Applying Lemma 2 in Berk (1974), we obtain

(A.69) $\qquad E\left(\left(\sum_t \Delta^d y_{t-1} \xi_{tk} \right)^2 \right) \leq (T-k)^2 \sum_{j=k+1}^{\infty} \delta_j^2,$

and

$$(\text{A.70}) \qquad E\left(\left(\sum_{j=1}^{k}\sum_{t}\Delta y_{t-j}\xi_{tk}\right)^2\right) \le (T-k)^2 k \sum_{j=k+1}^{\infty}\delta_j^2.$$

As $\delta_j \sim j^{-d-1}$, then $\sum_{j=k+1}^{\infty}\delta_j^2 = O_p(k^{-2d-1})$. Thus, (A.68) tends to zero, implying that $(T-k)^{-1}\sum_{t=k+1}^{T}U_t\xi_{tk}$ does it as well. Further, Berk (1974) has shown that

$$k^{1/2}\left\|(T-k)^{-1}R_0^{-1} - R^{-1}\right\| = o_p(1)$$

if $k^3/T \to 0$, where R is the population variance-covariance matrix associated to sample moments matrix R_0 (recall that all the processes in R_0 are either $I(0)$ or $I(d_1-1)$). This implies that

$$R_0^{-1}\sum_{t=k+1}^{T}U_t\xi_{tk} \xrightarrow{p} 0,$$

and hence $(\hat{\phi},\hat{\delta}_1,\ldots,\hat{\delta}_k)' - (\phi,\delta_1,\ldots,\delta_k)' = o_p(1)$. To check that the t-ratio $t_{\hat{\phi}_{ols}}$ diverges, note that its numerator is $O_p(1)$ while the denominator is $O_p((T-k)^{1/2})$, since $\hat{\sigma}^2$ is $O_p(1)$ and $(X_k'X_k)$ is $O_p(T-k)$. Both properties imply that $t_{\hat{\phi}_{ols}}$ diverges. As $\hat{\phi} \to \phi = -\vartheta(1) < 0$ (since ϑ_0 is positive and all the roots of $\vartheta(L)$ are outside the unit circle), the divergence is towards $-\infty$. \qquad Q.E.D.

PROOF OF THEOREM 8: The proof is similar to the one used in Theorem 5. Applying the MVT to $t_{\hat{\phi}_{ols}}(\hat{d}_1)$ around $(1-c)$ yields

$$(\text{A.71}) \qquad t_{\hat{\phi}_{ols}}(\hat{d}_1) = t_{\hat{\phi}_{ols}}(1-c) + \frac{t_{\hat{\phi}_{ols}}(\breve{d})}{\partial d}(\hat{d}_1 - (1-c)).$$

Since $\breve{d}$ in this expression is an intermediate point between $\hat{d}_1$ and $1-c$ and $\hat{d}_1 \xrightarrow{p} 1-c$, then $\breve{d} \xrightarrow{p} 1-c$. Next, we check that the second term in (A.71) is $o_p(1)$. As before, for the sake of simplifying the presentation, we will focus on the case where u_t is a stationary $AR(1)$ process, namely $\alpha(L) = 1 - \alpha L$. In such a case, the t-ratio $t_{\hat{\phi}_{ols}}(d)$ can be written as

$$(\text{A.72}) \qquad t_{\hat{\phi}_{ols}}(d) = \frac{(\sum \Delta y_{t-1}^2)(\sum \Delta^d y_{t-1}\varepsilon_t) - (\sum \Delta^d y_{t-1}\Delta y_{t-1})(\sum \Delta y_{t-1}\varepsilon_t)}{\hat{\sigma}(\sum \Delta y_{t-1}^2)^{1/2}(\sum \Delta y_{t-1}^2 \sum \Delta^d y_{t-1}^2 - (\sum \Delta^d y_{t-1}\Delta y_{t-1})^2)^{1/2}\hat{\sigma}(\sum \Delta y_{t-1}^2)^{1/2}} = \frac{A_3}{A_4}.$$

The first derivative of $t_{\hat{\phi}_{ols}}(d)$ with respect to d is given by

$$(\text{A.73}) \qquad \frac{\partial t_{\hat{\phi}_{ols}}(d)}{\partial d} = \frac{(\partial A_3(d)/\partial d)A_4 - A_3(\partial A_4(d)/\partial d)}{A_4^2},$$

where

$$(\text{A.74}) \qquad \frac{\partial A_3(d)}{\partial d} = \left(\sum \Delta y_{t-1}^2\right)\sum \varepsilon_t\left(\Delta^d y_{t-1}\log(1-L)\right)$$
$$- \left(\sum \Delta y_{t-1}\varepsilon_t\right)\left(\sum \Delta y_{t-1}\Delta^d y_{t-1}\log(1-L)\right),$$

$$\frac{\partial A_4(d)}{\partial d} = \left(\sum \Delta y_{t-1}^2\right)^{1/2}\left(\left(b_1^{1/2}\frac{\partial \hat{\sigma}}{\partial d}\right) + b_1^{-1/2}\sum \Delta y_{t-1}^2 \sum \Delta^d y_{t-1}\log(1-L)\right)$$
$$- b_1^{-1/2}\left(\sum \Delta y_{t-1}^2\right)^{1/2}\left(\sum \Delta^d y_{t-1}\Delta y_{t-1}\right)\sum \Delta y_{t-1}\Delta^d y_{t-1}\log(1-L),$$

and

$$(\text{A.75}) \qquad b_1 = \left[\sum \Delta y_{t-1}^2 \sum \Delta^d y_{t-1}^2 - \left(\sum \Delta^d y_{t-1}\Delta y_{t-1}\right)^2\right].$$

Using the same arguments as in Theorem 5, it can be checked that

$$\text{(A.76)} \qquad \left.\frac{\partial A_3(d)}{\partial d}\right|_{d=1-c} = o_p(T^2),$$

$$\text{(A.77)} \qquad \left.\frac{\partial A_4(d)}{\partial d}\right|_{d=1-c} = O_p(T^{3/2}),$$

$$\text{(A.78)} \qquad A_3|_{d=1-c} = O_p(T^{3/2}),$$

and

$$\text{(A.79)} \qquad A_4|_{d=1-c} = O_p(T^{3/2}),$$

implying that expression (A.73) evaluated at $(1-c)$ is $O_p(1)$. This result implies again that

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

Moreover, since $(\hat{d}_1 - (1-c)) = O_p(T^{-1/2})$, it follows that

$$\text{(A.80)} \qquad \frac{\partial t_{\phi_{ols}}(\check{d}_1)}{\partial d}(\hat{d}_1 - (1-c)) = o_p(1).$$

Therefore $(t_{\phi_{ols}}(\hat{d}_1) - t_{\phi_{ols}}(1-c)) = o_p(1)$, implying that

$$\text{(A.81)} \qquad t_{\phi_{ols}}(\hat{d}_1) \xrightarrow{w} N(0,1). \qquad\qquad\qquad Q.E.D.$$

APPENDIX B

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 X presents the critical values for the case where the DGP is a random walk and the t-ratio is computed from equation (2) whereas Table XI (Table XII) presents the corresponding values for the case where the DGP is the same as

TABLE X

CRITICAL VALUES

DGP: $\Delta y_t = \varepsilon_t$; REGRESSION: $\Delta y_t = \phi \Delta^{d_1} y_{t-1} + e_t$

T	$T = 100$			$T = 400$			$T = 1000$		
d_1/sig. lev.	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.0	−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 XI
CRITICAL VALUES
DGP: $\Delta y_t = \varepsilon_t$; REGRESSION: $\Delta y_t = c + \phi \Delta^{d_1} y_{t-1} + e_t$

T	$T = 100$			$T = 400$			$T = 1000$		
d_1/sig. lev.	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.0	−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 XII
CRITICAL VALUES
DGP: $\Delta y_t = c + \varepsilon_t$; REGRESSION: $\Delta y_t = c_0 + c_1 t + \phi \Delta^{d_1} y_{t-1} + e_t$

T	$T = 100$			$T = 400$			$T = 1000$		
d_1/sig. lev.	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.0	−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	−28
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

in the previous case but a constant (a constant and a linear trend) is included in regression (2). To compute those critical values, we have generated random walk processes (with a drift in the case of Table XII) from i.i.d. $N(0, 1)$ disturbances and computed the t-ratio associated to $\hat{\phi}_{ols}$ in the corresponding regression. In all the tables above, d is the value used in regression (2) to compute the t-ratio. The number of replications is 10,000.

REFERENCES

BAILLIE, R. T. (1996): "Long Memory Processes and Fractional Integration in Economics and Finance," *Journal of Econometrics*, 73, 15–131.

BHARGAVA, A. (1986): "On the Theory of Testing for Unit Roots in Observed Time Series," *Review of Economic Studies,* 53, 369–384.

——— (1994): *Statistics for Long Memory Processes.* New York: Chapman and Hall.

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.

BERK, K. N. (1974): "Consistency of Spectral Estimates," *The Annals of Statistics,* 2, 489–502.

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,* 160, 471–490.

DAVIDSON, J. (1994): *Stochastic Limit Theory.* New York: Oxford University Press.

DAVIDSON, J., AND R. M. DE JONG (1999): "The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals II: Fractionally Integrated Processes," Unpublished Manuscript, Michigan State University.

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.

——— (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.

DOLADO, J. J., J. GONZALO, AND L. MAYORAL (2001): "Long-range Dependence in Spanish Opinion Poll Data," forthcoming in *Journal of Applied Econometrics.*

DOLADO, J. J., AND F. MARMOL (1999): "Asymptotic Inference for Nonstationary Fractionally Integrated Processes," Working Paper Series No. 99-68, Universidad Carlos III de Madrid.

FOX, R., AND M. S. TAQQU (1986): "Large Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time 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," Mimeo.

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.

GONZALO, J., AND T. LEE (1998): "Pitfalls in Testing for Long Run Relationships," *Journal of Econometrics,* 86, 129–154.

GOURIEROUX, C., F. MAUREL, AND A. MONFORT (1989): "Least Squares and Fractionally Integrated Regressors," Document de Travail No. 8913, INSEE.

GRANGER, C. W. J., AND K. JOYEUX (1980): "An Introduction to Long-memory Time Series and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15–29.

GRAY, H. L., AND N. F. ZHANG (1988): "On a New Definition of the Fractional Difference," *Mathematics of Computation*, 50, 513–529.

JOHANSEN, S. (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–1580.

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and its Applications.* New York: Academic Press.

HAUSER, M. A., B. M. POTSCHER, AND E. RESCHENHOFER (1999): "Measuring Persistence in Aggregate Output: ARMA Models, Fractionally Integrated ARMA Models and Nonparametric Procedures," *Empirical Economics,* 24, 243–269.

HELLAND, I. S. (1982): "Central Limit Theorems for Martingales with Discrete or Continuous Time," *Scandinavian Journal of Statistics,* 9, 79–94.

KRAMER, W. (1998): "Fractional Integration and the Augmented Dickey-Fuller Test," *Economic Letters,* 61, 269–272.

LEE, D., AND P. SCHMIDT (1996): "On the Power of the KPSS Test of Stationarity against Fractionally Integrated Alternatives," *Journal of Econometrics,* 73, 285–302.

MANDELBROT, B. B., AND J. W. VAN NESS (1968): "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, 10, 422–437.

MARINUCCI, D., AND P. M. ROBINSON (1999): "Alternative Forms of Brownian Motion," *Journal of Statistical Planning and Inference,* 80, 11–122.

MARMOL, F. (1998): "Searching for Fractional Evidence Using Combined Unit Root Tests," Working Paper Series No. 98-39, Universidad Carlos III de Madrid.

MAYORAL, L. (2000): "A New Minimum Distance Estimation Procedure of ARFIMA Processes," Working Paper Series No. 00-17, Universidad Carlos III de Madrid. Revised version (2002) available upon request.

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., AND Z. XIAO (1998): "A Primer on Unit Root Testing," *Journal of Economic Surveys,* 12, 423–470.

ROBINSON, P. M. (1992): "Semiparametric Analysis of Long Memory Time Series," *Annals of Statistics,* 22, 515–539.

——— (1994): "Efficient Tests of Nonstationary Hypotheses," *Journal of the American Statistical Association,* 89, 1420–1437.

SAID, S. E., AND D. A. DICKEY (1984): "Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order," *Biometrika,* 71, 599–608.

SAMARODNITSKI, G., AND M. S. TAQQU (1994): *Stable Non-Gaussian Random Processes.* New York: Chapman and Hall.

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. (1990): "The Fractional Unit Root Distribution," *Econometrica,* 58, 495–505.

——— (1992): "Maximum Likelihood Estimation of Stationary Univariate Fractionally-integrated Time-series Models," *Journal of Econometrics,* 53, 165–188.

STOCK, J. H., AND M. W. WATSON (1988): "Testing for Common Trends," *Journal of the American Statistical Association,* 83, 1097–1107.

TANAKA, K. (1996): *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory.* New York: Wiley.

——— (1999): "The Nonstationary Fractional Unit Root," *Econometric Theory,* 15, 549–582.

TIESLAU, M., P. SCHMIDT, AND R. BAILLIE (1996): "A Minimum Distance Estimator for Long-Memory Errors," *Journal of Econometrics,* 71, 249–264.

VELASCO, C., AND P. M. ROBINSON (2000): "Whittle Pseudo-Maximum Likelihood Estimation for Nonstationary Time Series," *Journal of the American Association,* 95, 1229–1243.

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,* 1, 27–43.