INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO

CENTRO DE INVESTIGACIÓN ECONÓMICA

Discussion Paper Series

Optimal Fractional Dickey-Fuller Tests for Unit Roots

Ignacio N. Lobato
Instituto Tecnológico Autónomo de México

and

Carlos Velasco
Universidad Carlos III de Madrid

February 2004
Discussion Paper 04-01

Av. Camino a Santa Teresa # 930
Col. Héroes de Padierna
México, D.F. 10700
M E X I C O

OPTIMAL FRACTIONAL DICKEY-FULLER TESTS FOR UNIT ROOTS

Ignacio N. Lobato and Carlos Velasco
Centro de Investigación Económica
Instituto Tecnológico Autónomo de México (ITAM)
Av. Camino Sta. Teresa 930, México D.F. 10700, Mexico
e-mail:ilobato@itam.mx
Departamento de Econometría y Estadística
Universidad Carlos III de Madrid
Avda. de la Universidad 30, 28911 Leganés (Madrid), Spain.
e-mail:cavelas@est-econ.uc3m.es

Abstract

This article studies the fractional Dickey-Fuller (FDF) test for unit roots recently introduced by Dolado, Gonzalo and Mayoral (2002). Apart from the analogy with the Dickey-Fuller test, the main motivation for their method relies on simulations since these authors do not provide any justification for their particular implementation of the FDF test. In order to give additional rationale to the test, we frame the FDF test in a model where a nuisance or auxiliary parameter is not identified under the null hypothesis. Within this framework we investigate optimality aspects of the class of tests indexed by this auxiliary parameter and show that the test proposed by these authors is not optimal. In addition, we propose feasible FDF tests with good asymptotic and finite sample properties.

1 Introduction

In a recent paper, Dolado, Gonzalo and Mayoral (2002, hereinafter DGM) have introduced a fractional Dickey-Fuller (hereinafter, FDF) test for testing the null of unit root against the alternative of fractional integration. This problem had already been posed and optimally solved by Robinson (1994) and Tanaka (1999) for the Gaussian case. DGM showed that their FDF test can equal or outperform the finite sample power of these optimal tests for some particular implementations, even for Gaussian time series. However, in contrast with the optimal tests, DGM's testing procedure is not supported by any statistical principle or optimality criterion and its practical relevance appears to rely exclusively on the results of Monte Carlo experiments. This superiority in the finite sample behavior motivates to investigate whether it is possible to provide some rationality and to modify the FDF test proposed by DGM to improve its asymptotic and finite sample properties. By considering the FDF test as a class of tests indexed by an auxiliary or nuisance parameter that is not identified under the null hypothesis, in this article we derive optimal FDF tests that improve the *ad hoc* implementation of the FDF test proposed in DGM.

In DGM's simplest framework y_t denotes a fractionally integrated process whose true order of integration is d, $\Delta^d y_t = (1-L)^d y_t = \varepsilon_t$, where ε_t are i.i.d. random variables with zero mean and finite variance and L is the lag operator. The fractional difference operator is defined by

$$\Delta^\alpha y_t := \sum_{i=0}^{t-1} \pi_i(\alpha) y_{t-i}, \quad t = 1, 2, \ldots,$$

for any real α, where $\pi_i(\alpha) = (i-\alpha-1)!/(i!(-\alpha-1)!)$ are the coefficients of the binomial expansion of $(1-L)^\alpha$.

DGM consider testing the null hypothesis $d = 1$ versus either a simple alternative ($d = d_A$) or a composite alternative ($d < 1$), by means of the t-statistic of the coefficient of $\Delta^{d_1} y_{t-1}$ of the regression of

$$\Delta^1 y_t \qquad \text{on} \qquad \Delta^{d_1} y_{t-1},$$

where according to DGM, d_1 is "the true value of d under the alternative hypothesis". That is, the y_t variable in the left hand side (LHS) has been differenced once (the value of d under the null hypothesis) whereas the y_{t-1} variable in the right hand side (RHS) has been differenced d_1 times ("the value of d under the alternative hypothesis"). Despite DGM realized that their test was consistent against local alternatives, DGM did not provide any further rationale for their test, and just relied on the analogy with the Dickey-Fuller (hereinafter, DF) test. In the DF framework the null hypothesis is $d = 1$ and the alternative is $d = 0$. The DF test is based on the t-statistic of the coefficient of y_{t-1} of the regression of Δy_t on y_{t-1}, i.e.

$$\Delta^1 y_t \quad \text{on} \quad \Delta^0 y_{t-1},$$

where the y_t variable in the LHS has been differenced once whereas the y_{t-1} variable in the RHS has not been differenced.

However, the analogy with the DF test used by DGM to motivate their FDF test is tenuous. An alternative FDF test would use as test statistic the t-statistic of the coefficient of y_{t-1} of the regression of

$$\Delta^{1-d_1} y_t \quad \text{on} \quad \Delta^{1-1} y_{t-1}.$$

The "rationale" of this type of FDF test is that the y_t variable in the LHS has been differenced "one minus the value of d under the alternative" times whereas the y_{t-1} variable in the RHS has been differenced "one minus the value of d under the null" times, as in the DF regression. Note that, instead of having the same *regressand* as DF, as DGM propose, the previous FDF test would have the same *regressor* as DF. In fact, this alternative version of the FDF is closer to the original DF in the sense that the asymptotic null distribution of the associated t-ratio is nonstandard for any value of d_1, similar to the original DF, and contrary to DGM's FDF test where the asymptotic null distribution of their t-ratio is the standard normal when $d_1 \geq 0.5$. Obviously, these two proposals are not the only tests that can originate from the DF's set up, showing that invoking an analogy with the DF test is not sufficient for deriving sound statistical tests.

DGM considered the OLS estimation of the model

$$\Delta y_t = \phi \Delta^{d_1} y_{t-1} + u_t, \qquad t = 1, \ldots, T, \tag{1}$$

and proposed the FDF test statistic, which is the t-ratio associated with the OLS estimate $\widehat{\phi}$ of ϕ,

$$t(d_1) = \frac{\sqrt{T} \sum_{t=2}^{T} \Delta y_t \Delta^{d_1} y_{t-1}}{\sqrt{\sum_{t=2}^{T} (\Delta y_t - \widehat{\phi} \Delta^{d_1} y_{t-1})^2 \sum_{t=2}^{T} \left(\Delta^{d_1} y_{t-1}\right)^2}}, \tag{2}$$

where T denotes the sample size. Based on their particular analogy with the DF test, DGM identified d_1 with "the true value of d under the alternative hypothesis", and hence they inappropriately stated that for implementing the FDF test "a value of d is needed under the alternative hypothesis to make [the test] feasible" (p. 1964). Hence, they went on to arbitrarily recommend the use of $d_1 = d_A$ when the alternative hypothesis is simple, and the use of $d_1 = \widehat{d}$ when the alternative is composite, where $\widehat{d}$ is a trimmed version of a $\sqrt{T}$-consistent estimator for d, such that $\widehat{d}$ is strictly smaller than 1 with probability one. We will show that this interpretation is wrong and that there is no need to estimate d even for the composite alternative case. Consequently, DGM's qualification of the FDF test as a Wald test on d is inadequate. DGM identified d_1 with the "true value of d under the alternative hypothesis" and this interpretation of d_1 led them to propose an inefficient, arbitrary and complicated implementation of the FDF test, as we will show. By contrast, we argue that the parameter d_1 has a very concrete statistical meaning, since it defines a *class of tests* indexed by d_1, as it is emphasized in expression (2) by writing explicitly the input value d_1 as an argument of the test statistic. This interpretation of d_1 will allow us to derive simple (since there is no need of trimming or even of estimating d) and efficient implementations of the FDF test.

In order to gain a new perspective on the DGM framework, notice that in model (1), ϕ represents the slope of the best proportional predictor of Δy_t given $\Delta^{d_1} y_{t-1}$ when $d_1 > 0.5$, that is,

$$\phi = \phi(d_1) = \text{plim}_{T\rightarrow\infty} \frac{T^{-1}\sum_{t=1}^{T} \Delta y_t \Delta^{d_1} y_{t-1}}{T^{-1}\sum_{t=1}^{T}(\Delta^{d_1} y_{t-1})^2}.$$

The previous probability limits will be evaluated in Section 2. The key insight is that the parameter ϕ can be different for each d_1, so its dependence on d_1 has been stressed by writing $\phi(d_1)$. Notice that under the null hypothesis, Δy_t and $\Delta^{d_1} y_{t-1}$ are uncorrelated, implying that $\phi(d_1) = 0$ for all d_1. In Sections 2 and 3, we will see that maximizing the power of the FDF test is achieved by maximizing the correlation between Δy_t and $\Delta^{d_1} y_{t-1}$. Then, rather than "the true value of d", d_1 is "a tuning parameter that determines the power of the FDF test". The important difference with DGM is that d_1 is not some arbitrary value derived from DGM's particular analogy with the DF test, but a parameter that the researcher chooses to maximize the correlation between Δy_t and $\Delta^{d_1} y_{t-1}$, and hence to maximize the power of the FDF test.

As we have seen, under the null hypothesis, Δy_t and $\Delta^{d_1} y_{t-1}$ are uncorrelated, that is, $\phi(d_1) = 0$ for all d_1. In addition, under the alternative, the true d_1 can be defined as the degree of differencing of y_{t-1} that maximizes the correlation between Δy_t and $\Delta^{d_1} y_{t-1}$. Then, under the alternative d_1 is properly defined, but under the null d_1 is not identified, since $\phi(d_1) = 0$ for all d_1. Therefore, this testing framework is similar, for instance, to the nonlinear regression set up considered in Section 2 in Hansen (1996). This situation is nonstandard since d_1 can be regarded as a nuisance or auxiliary parameter that is not identified under the null hypothesis that states that the parameter of interest d equals 1. This statistical problem has been addressed, among others, by Davies (1977, 1987), Andrews and Ploberger (1994) and Hansen (1996). These references propose tests statistics that consider all the possible values that the auxiliary parameter can take. However, in this article we will show that, in the pure fractional case, the true optimal d_1 just depends on the true value of d on a simple (linear) way; and in particular, the optimal FDF test against local alternatives is obtained by setting d_1 equal to the true value of d_1 when d tends to 1, approximately 0.69. Hence, for this pure fractional case, there is no need of employing tests procedures that consider a range of values of the nuisance parameter d_1. However, for more complicated cases where the augmented FDF case is required, the true d_1 can be difficult to define and derive since it depends on the short run properties of y_t, so tests that employ all the possible values of the nuisance parameter can be more useful.

The plan of the article is the following. In the next section we derive optimal FDF tests for pure $I(d)$ processes using an asymptotic local power criterion in terms of d_1 under a sequence of local alternatives that converge to the null at the parametric rate. Also, within this framework, we analyze testing procedures, previously employed in econometrics and statistics, which consider all the values of the auxiliary parameter in a given interval. Section 3 studies optimal tests for fixed alternatives introducing the maximal squared correlation as a criterion to define d_1. This criterion function allows the derivation of a feasible and optimal implementation of the FDF test when a consistent estimator of d is available. A brief Monte Carlo exercise compares the finite sample performance of the considered tests. Finally, Section 4 concludes. For simplicity, we have followed the notation in DGM as close as possible.

2 An optimal FDF test for local alternatives

In the previous section we have seen that the actual value of d_1 used to implement the FDF test may have very important implications on the properties of the test. In order to motivate this point further, in Table 1 we report the results of a small Monte Carlo exercise. The data is fractionally integrated with $d = \{0, 0.1, ..., 0.9, 1\}$ with Gaussian errors and the selected $d_1 = \{0, 0.1, ..., 0.9\}$. We have chosen this grid of values for d_1 because they are the values considered by DGM and because we will show later that for the pure fractional case the optimal d_1's are always below 1. The considered sample size is 100 and the number of replications is 30,000. Simulations have been carried out in double precision Fortran 90.

The set up is similar to that employed by DGM in their Figures 1 and 2. Table 1 reports rejection percentages for the $t(d_1)$ tests based on 5% asymptotic critical values. Note that for the case $d_1 < 0.5$, we employ the estimated critical values reported by DGM in their Table X in p. 2003. Notice that when $d_1 = d$, $t(d_1)$ represents an unfeasible FDF test, as proposed by DGM, which does not take into account the sampling variation associated with the estimation of d. There are two main lessons from our Table 1. First, for any value of d_1 the empirical power is essentially 1 for the case when $d < 0.5$. Hence, for this sample size the most interesting case is when $d \geq 0.5$. Second, for the case when $d \geq 0.5$ the optimal selection of d_1 is not d, as DGM propose, but a lower value. Inspection of Table 1 reveals that with respect to DGM's selection of d_1, the empirical power can increase up to 35% by optimally choosing d_1.

In fact, the same two conclusions also appear implicitly in Figure 2 in DGM where they show that the empirical power increases by choosing d_1 lower than d (notice that in Figure 2 in DGM the true value d is denoted by d_1^*). DGM were aware of this fact, and in p. 1975 they provided an explanation: "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*" (italics are ours). This sentence illustrates that DGM continuously confused the role of d_1 with that of d. In this article, it will become clear that the true (optimal) d_1, in a sense to be established, and the value of d_1 that maximizes the test statistic $t(d_1)$ for each sample are closely related with d, the true value of the fractional parameter, but they are not the same.

It could be argued that the previous Monte Carlo findings are questionable because in Table 1 there is a considerable size distortion, especially for the case $d_1 = 0.5$ (notice that the first column of Table 2 in DGM also provides similar evidence). In order to confirm that the previous findings are robust, we also calculated size-adjusted power. In Table 2 we report rejection percentages for the $t(d_1)$ tests based on 5% empirical critical values for the same set up as Table 1. Table 2 offers similar messages to Table 1. Mainly, compared with DGM's selection of d_1, the empirical power can increase substantially by optimally choosing d_1.

From the previous simulation results it is clear that some criterion to optimally select d_1 is desirable. Robinson (1994) and Tanaka (1999) consider a sequence of local alternatives to the null hypothesis and derive asymptotically uniformly locally most powerful tests under the assumption of Gaussian errors. In the DGM framework a similar analysis is limited since no distributional assumptions are imposed and the class of test statistics is given by (2). However, we can still use the same principle and define the true value of d_1 as that which maximizes the correlation between Δy_t and $\Delta^{d_1} y_{t-1}$, and hence maximizes the power of $t(d_1)$ against local alternatives.

The following theorem establishes the asymptotic distribution of the class of test statistics $t(d_1)$ under the sequence of local alternatives $d = 1 - \delta/\sqrt{T}$ for all values of $d_1 \geq 0.5$. We consider this range because the asymptotic distribution of $t(d_1)$ is the standard normal only for these values of d_1, and therefore power comparisons are analytically tractable. Notice that this analysis includes the case $d_1 = 1$ that would not make any sense under DGM's rationality since $d_1 = 1$ is the value of d under the null hypothesis. DGM's Theorem 4 also analyzes local alternatives but, following their particular reasoning, they just consider the case $d_1 = d = 1 - \delta/\sqrt{T}$. On the contrary, we study any $d_1 \geq 0.5$.

Theorem 1. *Under the assumption that the DGP is a fractional white noise defined as*

$$DGP : \Delta^{1-\delta/\sqrt{T}} y_t = \varepsilon_t 1_{t>0} \quad \text{with} \quad \delta \geq 0,$$

where ε_t *is i.i.d. with finite second moment, for* $d_1 \geq 0.5$, *the asymptotic distribution of the test statistic* $t(d_1)$ *is given by:*

$$t(d_1) \xrightarrow{w} N(-\delta h(d_1), 1),$$

where

$$h(d_1) = \frac{\Gamma(d_1)}{d_1 \sqrt{\Gamma(2d_1 - 1)}},$$

and Γ *represents the gamma function.*

The proof of the theorem is in the Appendix. Note that the noncentrality parameter of the Gaussian asymptotic distribution of $t(d_1)$ is a positive function $h(d_1)$, $d_1 > 0.5$. It achieves a maximum at $d_1 = d^* \simeq 0.69145$, $h(d^*) \simeq 1.2456$, and satisfies that $h(0.5) = 0$ and $h(1) = 1$, in agreement with Theorem 4 of DGM where the drift of the distribution of $t(d_1)$ for $d_1 = 1 - \delta/\sqrt{T} \to 1$ is obtained. In addition, as d_1 tends to infinity, $h(d_1)$ tends to zero, see the plot of the function $h(d_1)$ in Figure 1. Figure 2 plots $h(d_1)$ for the more interesting range $0.5 \leq d_1 \leq 1$. Theorem 1 is remarkable because it shows that, contrary to the previous econometrics literature (see Andrews and Ploberger (1994) or Hansen (1996)), there exists a unique true optimal d_1 independent of δ.

In Figures 1 and 2 we have added horizontal lines at $\sqrt{\pi^2/6} \simeq 1.2825$, which represents the noncentrality parameter achieved by the optimal Robinson-Tanaka test, and at $1 = h(1)$, which is the noncentrality parameter for DGM's original proposal, $d_1 = d = 1 - \delta/\sqrt{T} \to 1$. Notice that $h(d_1)$ and $\sqrt{\pi^2/6}$ are quite close at the true optimal $d_1 = d^*$ (in fact, for the Gaussian case the relative efficiency loss of the optimal FDF test with respect to Robinson-Tanaka's test is smaller than 4%) and that as d_1 approaches 0.5, $h(d_1)$ tends to zero and has a vertical asymptote, reflecting the infinite efficiency loss incurred by choosing $d_1 = 0.5$. In particular, since $h(0.5) = 0$, the test cannot detect root-T alternatives when $d_1 = 0.5$. However, it is simple to check that for the $d_1 = 0.5$ case the test can detect local alternatives converging to the null at the rate $T^{-1/2} \log T$.

In summary, Theorem 1 indicates that any value of $d_1 \geq 0.5$ implies a consistent FDF test, hence there is no need to estimate consistently d and the FDF test should not be qualified as a Wald test as DGM do. Furthermore, with respect to an asymptotic local power criterion, an optimal implementation of the FDF test requires employing the true optimal $d_1 = d^*$.

Next, we give further insight on the FDF test by comparing the optimal FDF test with Robinson-Tanaka's optimal test. Following the proof of Theorem 1 (see (13) in the appendix), it is easy to show that under local alternatives, the FDF test is asymptotically equivalent to

$$\mathcal{T}(d_1) := T^{1/2} \sum_{j=1}^{T-1} \omega_j(d_1) \hat{\gamma}_j$$

where the weights $\omega_j(d_1)$ are defined by

$$\omega_j(d_1) := \frac{\pi_{j-1}(d_1 - 1)}{\left(\sum_{i=0}^{\infty} \pi_i^2(d_1 - 1)\right)^{1/2}}, \quad j = 1, 2, \ldots,$$

and $\hat{\gamma}_j$ is the sample autocovariance of order j of Δy_t,

$$\hat{\gamma}_j = \frac{1}{T} \sum_{t=j+1}^{T} \Delta y_t \Delta y_{t-j}.$$

Under the null hypothesis, and for $d_1 \geq 0.5$, it is obvious that $\mathcal{T}(d_1)$ is distributed asymptotically as a $N(0,1)$, since $\Delta y_t = \varepsilon_t$ which is i.i.d., so the $T^{1/2}\hat{\gamma}_j$ are asymptotically distributed as i.i.d. standard normal variates.

The Robinson-Tanaka's test, see Robinson (1991) and Tanaka (1999), belongs to the same class as $\mathcal{T}(d_1)$, but it employs a different weighting scheme that does not depend on d_1 or in any other parameter,

$$\mathcal{T}^{RT} = T^{1/2} \sum_{j=1}^{T-1} \omega_j^{RT} \hat{\gamma}_j$$

where Robinson-Tanaka's weights are given by

$$\omega_j^{RT} := \frac{j^{-1}}{\left(\sum_{i=0}^{\infty} i^{-2}\right)^{1/2}} = \frac{j^{-1}}{(\pi^2/6)^{1/2}}, \quad j = 1, 2, \ldots.$$

Robinson (1991, p.83) proposed these weights and Tanaka showed that they are optimal in an asymptotic local power context. It is easy to show that $\mathcal{T}(d^*)$ is closest to $\mathcal{T}^{RT}$ in the sense that d^* is the value of d_1 that minimizes

$$\sum_{j=1}^{\infty} \left(\omega_j(d_1) - \omega_j^{RT}\right)^2.$$

This can be easily checked because it is straightforward to show that minimizing the previous criterion is equivalent to maximizing $h(d_1)$. In Figure 3, we have plotted the weights ω_j^{RT} and the weights $\omega_j(d_1)$ for three values of d_1: d^*, 0.8 and 0.6. Note that the weights associated to d^* are slightly lower than Robinson-Tanaka's weights for moderate autocorrelations. The weights associated with $d_1 = 0.8$ (0.6) are above (below) ω_j^{Tan} for long lags, but $\omega_j(d^*)$ are very similar to ω_j^{RT}.

For the cases where d_1 is below 0.5, the asymptotic null distribution is no longer the standard normal, hence power considerations become rather intricate. In particular, for the case $0 < d_1 < 0.5$ the asymptotic distribution of the FDF test is not pivotal, as stated in Theorem 2 in DGM, and for this case it is possible to show that under local alternatives the FDF statistic converges to a well defined zero mean distribution depending on δ and d_1. Monte Carlo experiments not reported here show that for this region the decrease in power is monotonic in d_1 from the peak at the true optimal $d_1 = d^*$.

For models where a nuisance parameter is not identified under the null hypothesis, the statistics and econometrics literatures have proposed test procedures that take into account simultaneously many values of the nuisance parameter instead of a single value. We have shown that in the framework of this section, there is no need of employing this approach since the true parameter d_1 (which is not identified under the null) is simply defined by $d_1 = d^*$ under the local alternatives. However, as we will comment later, for a general implementation of the FDF test, where additional lags of Δy_t or deterministic trends may be included, the true d_1 may have a complicated definition that depends on all these features. Hence, for the general case, a sensible practical strategy might be to employ the tests statistics that we consider next. In the rest of this section we briefly study this alternative approach that has been developed by Davies (1977, 1987), Andrews and Ploberger (1994) or Hansen (1996) in different contexts for a sequence of local alternatives similar to the one considered above.

The main idea is to consider the statistic $t(d_1)$ as a stochastic process indexed by the nuisance parameter d_1. Under the DGP of Theorem 1, the asymptotic distribution of $t(d_1)$ is pivotal for $d_1 \in D$ where $D = [0.5, \infty)$, hence we restrict our analysis to any closed interval $D_1 = [\underline{d}, \overline{d}]$ that belongs to the interior of D. Notice that the case $d_1 = 0.5$ has to be excluded because of the discontinuity of the asymptotic theory. In order to derive the test statistics and their asymptotic distribution theory, Theorem 2 below establishes the weak convergence of the process $t(d_1)$ in the metric space of the continuous functions over the set D_1, $\mathbb{C}(D_1)$, endowed with the uniform metric. Based on this theorem, test statistics are constructed by selecting continuous functionals φ of $t(d_1)$. For instance, the two most common are the Kolmogorov-Smirnov (KS), $\sup_{D_1}|t(d_1)|$ and the Cramer-von Mises (CvM), $\int_{D_1} t^2(d_1)\mathrm{d}d_1$. The test based on $\sup_{D_1}|t(d_1)|$ parallels the sup Wald test of Andrews and Ploberger (1994), and also similar analysis can be applied to the sup LM and sup LR tests. The basic result that justifies these tests is the following theorem.

Theorem 2. *Under the assumption that the DGP is a fractional white noise defined as*

$$DGP : \Delta^{1-\delta/\sqrt{T}} y_t = \varepsilon_t 1_{t>0} \quad \text{with} \quad \delta \geq 0,$$

where ε_t is i.i.d and has finite fourth moment, for $d_1 \in D_1$,

$$t(d_1) \Rightarrow W(d_1) - \delta h(d_1),$$

where $\Longrightarrow$ denotes weak convergence in the metric space $\mathbb{C}(D_1)$ endowed with the uniform metric, $W(d_1)$ is a zero mean Gaussian process with covariance kernel given by

$$C^W(d_1^a, d_1^b) := \left(\sum_{i=0}^{\infty} \pi_i(d_1^a - 1)\pi_i(d_1^b - 1) \right) \left(V(d_1^a)V(d_1^b) \right)^{-1/2},$$

for $d_1^a, d_1^b \in D_1$, $V(d_1) = \sum_{i=0}^{\infty} \pi_i^2(d_1 - 1) = \Gamma(2d_1 - 1)/\Gamma(-d_1)^2$, and $h(d_1)$ is defined in the statement of Theorem 1.

In particular, notice that under the null hypothesis ($\delta = 0$)

$$t(d_1) \Rightarrow W(d_1), \tag{3}$$

so that for each $d_1 \in D_1$ the asymptotic null distribution of $t(d_1)$ is the standard normal, whereas for local alternatives the asymptotic distribution of $t(d_1)$ is the normal with mean $-\delta h(d_1)$ and unit variance, agreeing with Theorem 1.

Given Theorem 2, it is immediate to derive the asymptotic distributions of the KS and the CvM tests under the null and under local alternatives. Furthermore, since the asymptotic distribution in (3) only depends on d_1, and not on any feature of the data such as any conditional moment, the asymptotic null distributions of these test statistics are pivotal, and critical values can be easily estimated by Monte Carlo simulation. Alternatively, wild bootstrap procedures as those described in Hansen (1996) are valid and simple to implement in this context. Then, consistency of the tests follows by standard arguments. As commented above, contrary to the framework in Hansen (1996), for the pure fractional case there is no need of employing a grid of d_1's since the true optimal $d_1 = d^*$ is known for local alternatives and furthermore, the FDF test which employs $d_1 = d^*$ is consistent against fixed alternatives. However, as we will argue in the Conclusions, for more complicated structures, defining the true (optimal) value for d_1 can be extremely complicated since this value depends on the autocorrelation structure of the data. In these situations, employing KS or CvM with bootstrap based critical values might be the best practical option. In the next section we will briefly investigate the finite sample behavior of these tests.

3 An optimal FDF test for fixed alternatives

In the previous section we have derived a consistent and optimal implementation of the FDF test in a local alternative framework. In this section we will consider a complementary criterion to select d_1 in a fixed alternative framework. We will see that if a semiparametric estimator of d is available, there are alternative versions of the FDF test that are more powerful in finite samples.

Since the asymptotic null distribution of the $t(d_1)$ statistic is the standard normal for any $d_1 \geq 0.5$, for this range of values for d_1 maximizing the power is equivalent to finding the value of d_1 that maximizes the probability limit of $t(d_1)^2$, properly standardized. Equivalently, denoting by $R^2(d_1)$ the squared sample correlation between Δy_t and $\Delta^{d_1} y_{t-1}$, that is,

$$R^2(d_1) = \frac{\left(\sum_{t=2}^{T} \Delta y_t \Delta^{d_1} y_{t-1}\right)^2}{\sum_{t=2}^{T} (\Delta y_t)^2 \sum_{t=2}^{T} (\Delta^{d_1} y_{t-1})^2},$$

the basic relation of simple regression theory,

$$t(d_1)^2 = T \frac{R^2(d_1)}{1 - R^2(d_1)}, \tag{4}$$

implies that maximizing the probability limit of $T^{-1} t(d_1)^2$ is equivalent to maximizing the probability limit of $R(d_1)^2$, defined as $\rho^2(d_1)$.

Therefore under the alternative hypothesis the natural definition of d_1 is the argument that maximizes $\rho^2(d_1)$, that is, the squared population correlation between Δy_t and $\Delta^{d_1} y_{t-1}$. Denote this true optimal d_1 by

$$d_1 = d_1^* := \arg\max_{d_1} \rho^2(d_1).$$

Since d_1 does not appear on the variance of Δy_t,

$$\begin{aligned} d_1^* &= \arg\max_{d_1} \ \mathrm{plim}_{T\to\infty} \frac{\left(T^{-1}\sum_{t=2}^{T} \Delta y_t \Delta^{d_1} y_{t-1}\right)^2}{T^{-1}\sum_{t=2}^{T} (\Delta^{d_1} y_{t-1})^2} \\ &= \arg\max_{d_1} \frac{\left(\lim_{T\to\infty} T^{-1}\sum_{t=2}^{T} Cov(\Delta y_t, \Delta^{d_1} y_{t-1})\right)^2}{\lim_{T\to\infty} T^{-1}\sum_{t=2}^{T} Var(\Delta^{d_1} y_{t-1})}. \end{aligned}$$

Then, using that $\Delta^d y_t = \varepsilon_t$, the objective function can be written as

$$\frac{\left(\lim_{T\to\infty} T^{-1}\sum_{t=2}^{T} Cov\left(\Delta^{1-d}\varepsilon_t, \Delta^{d_1-d}\varepsilon_{t-1}\right)\right)^2}{\lim_{T\to\infty} T^{-1}\sum_{t=2}^{T} Var\left(\Delta^{d_1-d}\varepsilon_{t-1}\right)}.$$

Next, we calculate these expressions starting by the denominator. Using that $\Delta^{d_1-d}\varepsilon_{t-1} = \sum_{i=0}^{t} \pi_i(d_1-d)\varepsilon_{t-1-i}$,

$$\lim_{T\to\infty} T^{-1}\sum_{t=2}^{T} Var\left(\Delta^{d_1-d}\varepsilon_{t-1}\right) = \lim_{T\to\infty} T^{-1}\sum_{t=2}^{T}\sum_{i=0}^{t-2} \pi_i(d_1-d)^2 = \sum_{i=0}^{\infty} \pi_i(d_1-d)^2 < \infty$$

if

$$d_1 - d > -0.5, \tag{5}$$

and in this case, $\sum_{i=0}^{\infty} \pi_i(d_1-d)^2 = \Gamma(2d_1-2d+1)/\Gamma(d-d_1-1)^2$. Note that the previous condition (5) is satisfied for any $d_1 \geq 0.5$. Regarding the numerator, using that $\Delta^{1-d}\varepsilon_t = \sum_{i=1}^{t} \pi_i(1-d)\varepsilon_{t-i}$,

$$\lim_{T\to\infty} T^{-1}\sum_{t=2}^{T} Cov\left(\Delta^{1-d}\varepsilon_t, \Delta^{d_1-d}\varepsilon_{t-1}\right) = \sum_{i=1}^{\infty} \pi_i(1-d)\pi_{i-1}(d_1-d).$$

Hence,

$$d_1^*(d) = \arg\max_{d_1} L(d, d_1)$$

where

$$L(d, d_1) = \frac{\left(\sum_{i=1}^{\infty} \pi_i(1-d)\pi_{i-1}(d_1-d)\right)^2}{\Gamma(2d_1-2d+1)/\Gamma(d-d_1-1)^2}. \tag{6}$$

Ideally, we would like to express analytically the objective function $L(d, d_1)$, and then derive the function $d_1^* = d_1^*(d)$ that provides the true optimal value of d_1 for each value of d.

Unfortunately, this is quite complicated for a general d. In agreement with the results in Section 2, when $d = 1 - \delta/\sqrt{T}$ the optimal selection of d_1 is $d_1^* = d^* \simeq 0.69$. For a general d, we have not been able to find an explicit expression for the numerator of equation (6). However, we can approximate $d_1^* = d_1^*(d)$ numerically with any level of precision, and in Figure 4 we report the $d_1^*(d)$ for some values of d and a truncation at $i = 10^5$ in the infinite sum in (6).

Confirming the results of the previous Monte Carlo experiments, Figure 4 shows that d_1^* is always below the true d. Figure 4 also indicates that the relation between d_1^* and d is essentially linear for the range $d_1^* \geq 0.5$. In Figure 4 we have added the regression line of $d_1^*(d)$ on d. This fit is given by $\widehat{d_1^*}(d) = -0.030 + 0.717d$ using a truncation at $i = 10^5$ in the infinite sum in (6). The standard error of this estimation is 0.0004. Notice that $\widehat{d_1^*}(d) - d > -0.5$, so that the condition (5) is always satisfied. In addition, in agreement with the results of the previous section, $\widehat{d_1^*}(1)$ is very close to d^*, the discrepancy can be attributed to the numerical error in the approximation.

Notice that for values of d below 0.74 (approximately), $\widehat{d_1^*}(d)$ would be below 0.5, so that for these values it may happen that $\widehat{d_1^*}(d)$ is not the optimal value for d_1. However, as previously stated, power comparisons for the case $d_1 < 0.5$ are analytically intractable, and furthermore, as the previous Monte Carlo shows, even for moderate sample sizes, detecting a violation of the null hypothesis is relatively simple when d is not close to 1. Hence, for simplicity, we propose to employ as d_1, a feasible version of $\widehat{d_1^*}(d)$. In particular, for the simple alternative case, we propose to select d_1 as $-0.030 + 0.717d_A$, which is different from the value of d under the alternative hypothesis, as DGM propose. In addition, for the more interesting composite alternative case, we propose to employ as d_1

$$\widehat{d_1^*}(\widetilde{d}) := -0.030 + 0.717\widetilde{d}, \tag{7}$$

where $\widetilde{d}$ is a consistent estimator of d that satisfies

$$T^{1/4} \log T \left(\widetilde{d} - d\right) = o_p(1). \tag{8}$$

Notice that DGM propose to estimate d with a $\sqrt{T}$-consistent estimator, whereas condition (8) also holds for many semiparametric estimators. The reason is that a careful inspection of the proof of Theorem 5 in DGM shows that $\partial t_{\widehat{\phi}_{ols}}(\breve{d})/\partial d$ is $O_p(T^{1/4} \log T)$, which is more accurate than $o_p(T^{1/2})$ as DGM state in their expression (A.48). The previous condition (8) holds for many semiparametric estimators for an appropriate choice of the bandwidth parameter, see Velasco (1999a, b). Of course, for the considered pure fractional case, semiparametric estimators are not necessary since the true model is known. However, for more complicated cases and in empirical applications, where an augmented FDF test is employed instead of the FDF test, semiparametric estimators may be specially relevant. Although DGM did not consider the use of semiparametric estimators, they can be very useful because they allow the derivation of consistent and efficient FDF tests without specifying a correct parametric model.

In addition, note that contrary to DGM, trimming is not necessary. Following their peculiar reasoning, in p. 1977 DGM wrongly stated that the trimming was necessary "since the value of d_1 that is needed to implement the test ought to be strictly smaller than 1". Our analysis shows that this trimming is not necessary since d_1 can be equal or bigger than 1. The following lemma justifies this implementation of the FDF test.

Lemma. *Under the null hypothesis ($d = 1$), the t-ratio statistic associated to the parameter ϕ in the regression*

$$\Delta y_t = \phi \Delta^{\widehat{d}_1^*(\widetilde{d})} y_{t-1} + u_t, \tag{9}$$

where $\widehat{d}_1^*(\widetilde{d})$ is given by (7) and $\widetilde{d}$ satisfies (8), is asymptotically distributed as $N(0,1)$.

The proof of this lemma is omitted since it is similar to the proof of Theorem 5 in DGM, with the commented modification to justify the use of a semiparametric estimator. The intuition of the lemma is also similar to DGM: under the null $\widehat{d}_1^*(\widetilde{d})$ is close to d^*, and hence the asymptotic standard distribution is the standard normal.

Our choice of d_1 is asymptotically optimal when $d_1 \geq 0.5$, but it is natural to wonder about the finite sample behavior of the considered tests. In the rest of the section we comment on the results of a small Monte Carlo study. The framework is similar to that considered in the simulations exercise of Section 2. The sample size is 100, $d = \{0.5, 0.6, 0.7, 0.8, 0.9, 1\}$, errors are Gaussian and the nominal level is 0.05. We consider FDF tests with several selections for d_1, namely

a) $d_1 = d$, denoted by FDF(d),

b) $d_1 = d^*$, FDF(d^*),

c) $d_1 = \widehat{d}_1^*(d)$, FDF$(\widehat{d}_1^*(d))$,

d) $d_1 = \widehat{d}_1^*(\widehat{d}_n)$, where $\widehat{d}_n$ is the Whittle parametric estimator, FDF$(\widehat{d}_1^*(\widehat{d}_n))$, and

e) $d_1 = \widehat{d}_1^*(\widetilde{d}_m)$, where $\widetilde{d}_m$ is the Gaussian semiparametric estimator with bandwidth m, FDF$(\widehat{d}_1^*(\widetilde{d}_m))$.

Regarding a) and c) notice that they represent unfeasible implementations of the FDF test that assume that the true d is known and ignore the sampling error associated with the estimation of d. Notice that in the considered pure fractional case, employing a semiparametric estimator is unnecessary and inefficient, however we also compute e) for completeness. Regarding d) and e) note that they could be calculated in two different ways since both the Whittle parametric estimator and the Gaussian semiparametric estimator could be applied to the original data or to the first differences of the data. We tried both possibilities and the results were very similar. The only apparent difference is that the size is slightly better controlled when the first differences are employed. The reason of this difference is that for $d = 1$ the estimators based on the levels are not consistent in their original form, see Velasco (1999b) and Velasco and Robinson (2000). For $\widetilde{d}_m$, the selected bandwidth is $m = n^{0.55}$, which is sufficient for (8) to hold. Finally, following DGM's suggestion for the test a) we set $d_1 = 0.99$ for computing the size results.

In addition to the FDF tests we also include the KS and the CvM tests considered in Section 2 and Robinson-Tanaka's LM test. For the KS and CvM tests we tried two types of critical values, asymptotic and bootstrap, calculated for $d_1 \in [0.51, 0.7]$, because in the considered pure fractional case any value higher than 0.69 is not optimal for any possible alternative. Since the performance with the bootstrap was slightly better, we just report the bootstrap results. Similar to Hansen (1996), for the bootstrap approximation we replace the numerator of the FDF $t(d_1)$ statistic by

$$\sum_{t=2}^{T} \Delta y_t \Delta^{d_1} y_{t-1} v_t,$$

where the $\{v_t\}$ is an i.i.d. sequence of zero mean and unit variance random variables, independent of y_t, and independent in each bootstrap replication. In these experiments 300 bootstrap replications have been computed, and a uniform grid of 30 values for $d_1 \in [0.51, 0.7]$ has been employed. Regarding the selection of v_t, Hansen employs the standard $N(0,1)$, whereas we employ a Bernoulli variate where $P(v_t = 0.5(1-\sqrt{5})) = (1+\sqrt{5})/2\sqrt{5}$ and $P(v_t = 0.5(1+\sqrt{5})) = 1-(1+\sqrt{5})/2\sqrt{5}$. This selection has been employed before (see Mammen (1993) or Stute, González-Manteiga and Presedo-Quindimil (1998)), and it presents the advantage that the third moment of v_t is equal to 1, and hence, the first three moments of the bootstrap series coincide with the first three moments of the original series. Finally, we also tried the ExpLM test of Andrews and Ploberger (1994), but these results are not reported since they were very similar to those of CvM.

Table 3 reports the Monte Carlo size ($d = 1$) and power ($d < 1$) of the previous tests based on 5% asymptotic critical values for all tests except for the KS and CvM tests whose figures are based on 5% bootstrap-based critical values. For this table and the next, the number of replications is 10,000 for the bootstrap KS and CvM, and 100,000 for all the other tests. Regarding the size results, notice that for all the variants of the FDF test the empirical rejection probabilities under the null are higher than the nominal size. In particular, FDF($\widehat{d}_1^*(\widetilde{d}_m)$) presents a severe size distortion for this sample size. On the contrary, all the non-FDF tests appear to be conservative. Hence, in Table 4 we also report the figures based on empirical critical values (size-adjusted power) for all tests. The main findings from these tables are the following. First, all the FDF tests appear to be more powerful than the non-FDF tests. Second, compared to FDF(d), FDF(d^*) fares relatively well when the alternatives are close to the null, as could be expected. Third, comparing the different implementations of the FDF test, it is very interesting to see that the unfeasible version of the FDF test proposed by DGM, FDF(d), is dominated by all the other implementations of the FDF tests in terms of power. In particular, when $d = 0.8$ or 0.9, power can increase in relative terms between 20% and 30% by using FDF($\widehat{d}_1^*(d)$) instead of FDF(d). Fourth, regarding the tests that employ all the values of the nuisance parameters, KS and CvM, the results are very similar with a slight advantage of KS over CvM, as it is shown in Table 3. In this respect, note that Table 4 offers a contradictory message, reflecting the difficulties of interpretation associated to the unrealistic "size-adjusted" figures.

These findings are based on a very simple DGP. Monte Carlo experiments using alternate parent distributions can be additionally informative. However, two preliminary conclusions arise from this finite sample evidence. First, the FDF(d^*) test may be a very sensible option in practical applications since it performs fairly well in finite samples and has the advantage of being very simple to implement. Second, if a consistent estimator for d is available, more power in finite samples can be achieved by using alternative versions of the FDF test.

4 Conclusions

Similar to the Dickey-Fuller test, the FDF correlation test is likely to become very popular among applied researchers. However, the implementation of the FDF test proposed by DGM is not supported by any statistical principle or optimality criterion and, apart from not being theoretically optimal, it leads to a loss of power in finite samples. DGM obtained the basic asymptotic properties of their arbitrary and complicated implementations of the FDF test statistics, but they did not provide any formal justification of their test procedures. In particular, they did not address many interesting issues, such as the interpretation of the nuisance parameter d_1, the optimal implementation of the FDF test, the consistency of the FDF test for local alternatives for any $d_1 \geq 0.5$, and the possible use of d_1 based on semiparametric estimators of d.

In this article we have analyzed the FDF test with a model where a nuisance or auxiliary parameter, d_1, is not identified under the null hypothesis. By using this framework, we have been able to provide rationale to some implementations of the FDF tests. In particular, we have shown that the input value d_1 that needs to be used in the test should not be interpreted as "the true value of d under the alternative hypothesis" as DGM do, but rather as an auxiliary parameter that maximizes the power of the FDF test. Contrary to DGM who arbitrarily selected d_1, we have addressed the issue of optimally selecting the value of d_1. In a local alternative framework, we have proved that the FDF test is consistent against local alternatives for any $d_1 \geq 0.5$, and most importantly, we have derived an optimal selection for this auxiliary parameter. In addition, we have compared this optimal FDF test with Robinson-Tanaka's optimal test. In the pure fractional framework, we have shown that, even in the composite alternative case, there is no need to estimate d, contrary to what DGM state. Hence, in this context, the FDF test should not be considered as a Wald type test on d, as DGM do. In the context of fixed alternatives, we have defined the true optimal d_1 using a criterion based on the population squared correlation between the dependent and independent variables of regression (1). In this framework, we have derived optimal tests that are consistent against alternatives that converge to the null at the parametric rate, and where d_1 can be based on semiparametric estimators of d.

We stress that we have just analyzed the case where the DGP is a pure fractionally integrated process and where the employed test is the FDF test. Similarly to DGM, in practical applications it is important to allow for more complicated DGP's where the errors ε_t may be weakly serially correlated. In this situation, we could follow DGM's recommendation and apply the augmented FDF test. Following the arguments in DGM and our Theorem 2, it is straightforward to show that in the augmented FDF regression model, the asymptotic null distribution of the t ratio statistic associated to the coefficient of the regressor $\Delta^{d_1} y_{t-1}$ is still the standard normal, but the asymptotic covariance function of the $t(d_1)$ process is more complicated, reflecting the short run properties of $\Delta^{d_1} y_{t-1}$. In particular, the covariance kernel is given by

$$C_A^W(d_1^a, d_1^b) := \left(\sum_{i=0}^{\infty} \pi_i^A(d_1^a - 1) \pi_i^A(d_1^b - 1) \right) \left(V_A(d_1^a) V_A(d_1^b) \right)^{-1/2},$$

where $V_A(d_1) = \sum_{i=0}^{\infty} \pi_i^A(d_1 - 1)^2$ and the filter $\pi_i^A(d_1 - 1)$ is defined by the convolution of the fractional filter $\pi_i(d_1 - 1)$ and the filter ψ_i

$$\pi_i^A(d_1 - 1) = \sum_{j=0}^{i} \pi_j(d_1 - 1)\psi_{j-i},$$

where the coefficients $\{\psi_j\}$ reflect the short run properties of $\Delta^d y_t$, that is, $\Delta^d y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, so that the pure fractional model is recovered when $\psi_0 = 1$ and $\psi_i = 0$ for all $i \neq 0$. In a similar way, under local alternatives the drift parameter also changes, in particular $h(d_1)$ is replaced by

$$h_A(d_1) = \left(\sum_{i=0}^{\infty} \pi_i^A(d_1 - 1)^2 \right)^{-1/2} \sum_{i=0}^{\infty} \frac{\pi_i^A(d_1 - 1)}{i + 1}.$$

The optimal d_1 in this case is obtained by maximizing $h_A(d_1)$ which depends on the whole sequence ψ_i through π_i^A. Hence, in this general case, the true optimal d_1 is going to be different to the true optimal d_1 for the pure fractional case. Though it would be possible to estimate the true d_1 by using parametric and semiparametric estimates for the true d and the ψ_i, this procedure would be rather complicated in practice, so we believe that the simplest and most natural testing procedures would be the KS or CvM tests with bootstrap based critical values. However, a careful analysis of the general case is beyond the scope of the current article.

ACKNOWLEDGMENT

We thank Josu Arteche, Miguel Delgado, Luis Gil-Alaña, Jesús Gonzalo, Javier Hidalgo, Joel Horowitz, Pierre Perron and Werner Ploberger for useful conversations. Lobato acknowledges financial support from Asociación Mexicana de Cultura and from the mexican Consejo Nacional de Ciencia y Tecnología (CONACYT) under project grant 41893-S. Velasco acknowledges financial support from the spanish Dirección General de Enseñanza Superior (DGES) reference number BEC2001-1270.

$d_1 \setminus d$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
0	100	100	100	100	100	91.7	67.2	37.2	15.9	5.34
0.1	100	100	100	100	100	98.1	81.8	48.2	19.3	5.38
0.2	100	100	100	100	100	99.7	91.9	59.7	23.1	5.55
0.3	100	100	100	100	100	100	96.3	68.3	26.1	5.53
0.4	100	100	100	100	100	100	97.8	74.1	29.0	5.53
0.5	100	100	100	100	100	99.9	98.1	78.4	33.3	6.31
0.6	100	100	100	100	100	99.7	96.4	75.4	32.5	5.91
0.7	100	100	100	100	99.9	99.1	93.5	70.2	30.0	5.51
0.8	100	100	100	100	99.7	97.7	89.4	64.1	27.3	5.37
0.9	100	100	100	99.8	99.0	95.6	83.6	57.8	24.5	5.27

Table 1. Monte Carlo size ($d = 1$) and power ($d < 1$) of the FDF $t(d_1)$ tests: Percentage of rejections based on 5% asymptotic critical value. Series follow a FI(d) with Gaussian errors. Sample size is 100. Number of replications is 30,000.

$d_1 \setminus d$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0	100	100	100	100	99.4	91.1	66.1	36.2	15.2
0.1	100	100	100	100	100	97.9	80.9	46.8	18.5
0.2	100	100	100	100	100	99.7	90.9	57.2	21.5
0.3	100	100	100	100	100	99.9	95.8	66.2	24.5
0.4	100	100	100	100	100	99.9	97.3	72.0	27.1
0.5	100	100	100	100	100	99.9	97.1	74.3	29.1
0.6	100	100	100	100	100	99.6	95.7	72.8	29.6
0.7	100	100	100	100	99.9	99.0	92.9	68.2	28.1
0.8	100	100	100	100	99.6	97.6	88.6	62.5	26.0
0.9	100	100	100	99.8	98.9	95.4	82.9	56.8	23.7

Table 2. Monte Carlo power of the FDF $t(d_1)$ tests: Percentage of rejections based on 5% empirical critical values. Series follow a FI(d) with Gaussian errors. Sample size is 100. Number of replications is 30,000.

d	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
FDF(d)	100	99.7	93.7	64.1	25.1	5.37
FDF(d^*)	99.9	99.3	94.0	71.1	30.6	5.59
FDF($\widehat{d}_1^*(d)$)	100	100	98.2	77.6	32.2	5.57
FDF($\widehat{d}_1^*(\widehat{d}_n)$)	100	99.5	94.8	71.8	30.6	5.65
FDF($\widehat{d}_1^*(\widetilde{d}_m)$)	99.8	99.8	97.0	76.5	34.3	6.93
KS	100	99.0	90.7	58.6	19.3	4.70
CvM	99.8	98.5	89.4	58.3	19.3	4.42
LM	99.9	99.0	91.2	62.9	24.5	4.64

Table 3. Monte Carlo size ($d = 1$) and power ($d < 1$) of the FDF tests with several selections for d_1, the KS, the CvM and LM test: Percentage of rejections based on 5% asymptotic critical values for LM and all FDF tests and bootstrap-based critical values for KS and CvM. Series follow a FI(d) with Gaussian errors. Sample size is 100. The number of replications is 10,000 for KS and CvM and 100,000 for all the other tests.

d	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
FDF(d)	100	99.7	93.4	63.5	24.4
FDF(d^*)	99.9	99.1	93.2	68.9	28.4
FDF($\widehat{d}_1^*(d)$)	100	100	97.8	75.5	30.0
FDF($\widehat{d}_1^*(\widehat{d}_n)$)	100	99.4	94.1	69.6	28.5
FDF($\widehat{d}_1^*(\widetilde{d}_m)$)	99.8	99.6	95.2	70.4	28.1
KS	99.9	99.5	93.1	62.4	21.2
CvM	99.9	99.2	92.4	63.4	21.8
LM	99.9	99.1	91.9	64.5	25.7

Table 4. Monte Carlo power of the FDF tests with several selections for d_1, the KS, the CvM and LM test: Percentage of rejections based on 5% empirical critical values. Series follow a FI(d) with Gaussian errors. Sample size is 100. The number of replications is 10,000.

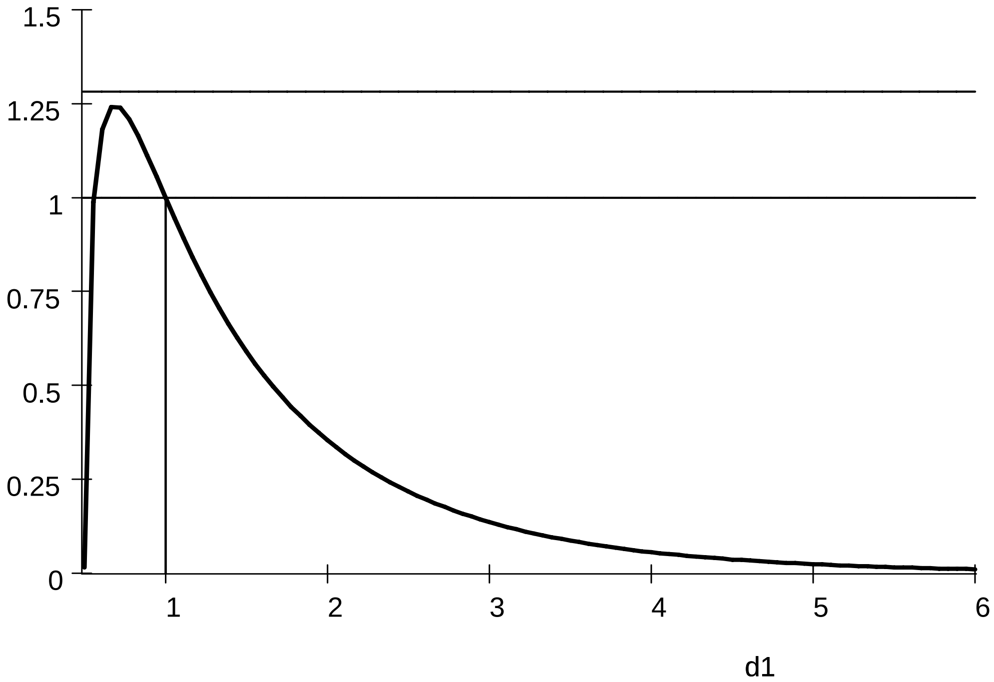

Figure 1. Asymptotic efficiency of the FDF and LM tests: plots of $h(d_1)$, $h(1) = 1$ and $\sqrt{\pi^2/6}$. The point $(1, 1)$ corresponds to DGM proposal.

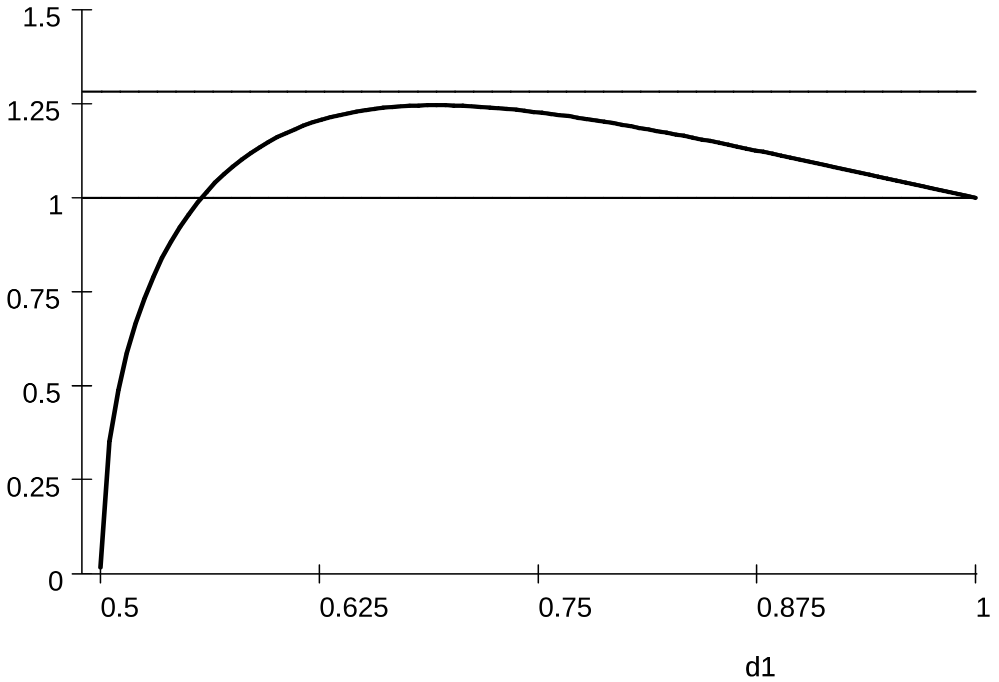

Figure 2. Asymptotic efficiency of the FDF and LM tests: plots of $h(d_1)$, $h(1) = 1$ and $\sqrt{\pi^2/6}$.

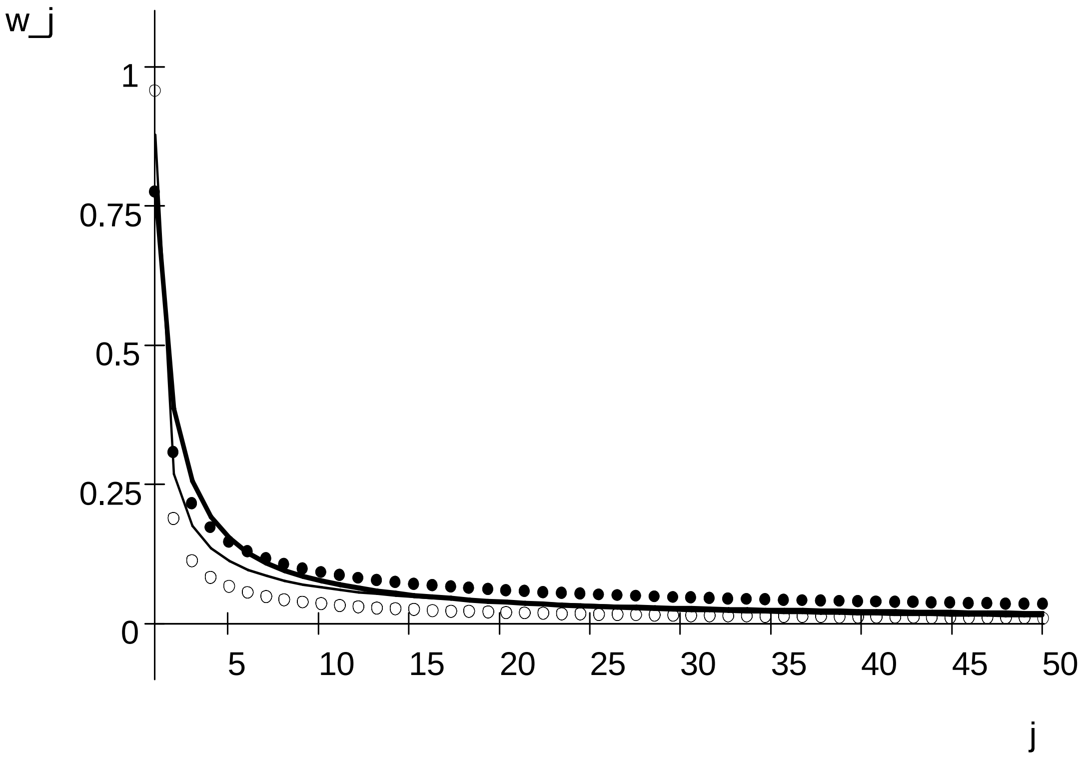

Figure 3. The weights ω_j^{RT} (bold solid line) and $\omega_j(d_1)$ for $d_1 = d^*$ (thin solid line), $d_1 = 0.8$ (dotted line) and $d_1 = 0.6$ (circled dotted line).

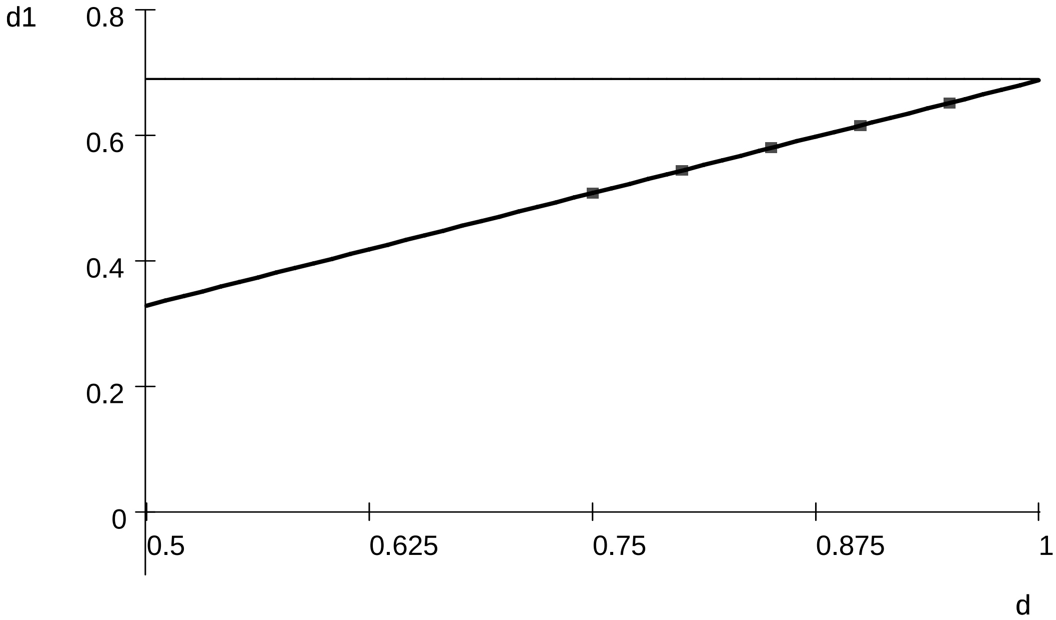

Figure 4. Plots of the points $(d, d_1^*(d))$, and the lines $d_1 = \widehat{d_1^*}(d)$ and $d_1 = d^* \equiv 0.69$.

5 Appendix

For simplicity, in this appendix we assume that the variance of ε_t is one.

Proof of Theorem 1. We begin by introducing some notation. Let

$$\Delta y_t = \Delta^{-\theta_T} \varepsilon_t 1_{t>0} = \varepsilon_t + \sum_{i=1}^{t-1} \pi_i(-\theta_T)\varepsilon_{t-i},$$

where $\theta_T := -\delta T^{-1/2}$, and $\pi_1(-\theta_T) = \theta_T$, $\pi_2(-\theta_T) = 0.5\theta_T(1+\theta_T) \approx -0.5\delta T^{-1/2}$, and in general $\pi_j(-\theta_T) \approx -j^{-1}\delta T^{-1/2}$, where the symbol $\approx$ means that as the sample size tends to infinity the ratio of the LHS and the RHS tends to one. Also,

$$\Delta^{d_1} y_{t-1} = \Delta^{-\eta_T} \varepsilon_{t-1} 1_{t>1} = \varepsilon_{t-1} + \sum_{i=1}^{t-2} \pi_i(-\eta_T)\varepsilon_{t-1-i},$$

where $\eta_T = 1 - d_1 - \delta T^{-1/2}$, so that $\pi_1(-\eta_T) = \eta_T \approx 1 - d_1$, $\pi_2(-\eta_T) = 0.5\eta_T(1+\eta_T) \approx 0.5(1-d_1)(2-d_1)$ and so on.

First, consider the numerator of $t(d_1)$ scaled by $T^{-1/2}$,

$$\begin{aligned} Q_T(d_1) &= T^{-1/2} \sum_{t=2}^{T} \Delta y_t \Delta^{d_1} y_{t-1} \\ &= T^{-1/2} \sum_{t=2}^{T} \left(\varepsilon_t + \sum_{i=1}^{t-1} \left(\frac{1}{i} \frac{-\delta}{\sqrt{T}} \right) \varepsilon_{t-i} \right) \left(\varepsilon_{t-1} + \sum_{i=1}^{t-2} \pi_i(-\eta_T)\varepsilon_{t-1-i} \right) \qquad (10) \\ &\quad + T^{-1/2} \frac{\delta^2}{2T} \sum_{t=2}^{T} \left(\sum_{i=1}^{t-1} \pi_i'(-\theta^*)\varepsilon_{t-i} \right) \left(\varepsilon_{t-1} + \sum_{i=1}^{t-2} \pi_i(-\eta_T)\varepsilon_{t-1-i} \right) \qquad (11) \end{aligned}$$

where π_i' is the first derivative of π_i and θ^* is some point between 0 and θ_T. Note that $|\pi_i'(-\theta^*)| \le C\ i^{-1}\log i$ by Lemma 1 of Delgado and Velasco (2003). Since (10) is $O_p(1)$ as it is showed next, it is straightforward to show that (11) is $o_p(1)$.

The leading term (10) of $Q_T(d_1)$ can be written as

$$T^{-1/2} \sum_{t=2}^{T} \left(\varepsilon_t + \left(\frac{-\delta}{\sqrt{T}} \right) \varepsilon_{t-1} + \sum_{i=1}^{t-2} \left(\frac{1}{(i+1)} \frac{-\delta}{\sqrt{T}} \right) \varepsilon_{t-i-1} \right) \left(\varepsilon_{t-1} + \sum_{i=1}^{t-2} \pi_i(-\eta_T)\varepsilon_{t-1-i} \right)$$

$$
\begin{aligned}
= \quad & T^{-1/2}\sum_{t=2}^{T}\left(\left(\frac{-\delta}{\sqrt{T}}\right)\varepsilon_{t-1}^2+\sum_{i=1}^{t-2}\left(\frac{1}{(i+1)}\frac{-\delta}{\sqrt{T}}\right)\pi_i(-\eta_T)\varepsilon_{t-i-1}^2\right) & (12)\\
& +T^{-1/2}\sum_{t=2}^{T}\varepsilon_t\left(\varepsilon_{t-1}+\sum_{i=1}^{t-2}\pi_i(-\eta_T)\varepsilon_{t-1-i}\right) & (13)\\
& +T^{-1/2}\sum_{t=2}^{T}\left(\left(\frac{-\delta}{\sqrt{T}}\right)\varepsilon_{t-1}\right)\left(\sum_{i=1}^{t-2}\pi_i(-\eta_T)\varepsilon_{t-1-i}\right) & (14)\\
& +T^{-1/2}\sum_{t=2}^{T}\left(\sum_{i=1}^{t-2}\left(\frac{1}{(i+1)}\frac{-\delta}{\sqrt{T}}\right)\varepsilon_{t-i-1}\right)\left(\sum_{j=1,j\neq i}^{t-2}\pi_i(-\eta_T)\varepsilon_{t-1-j}\right). & (15)
\end{aligned}
$$

The last two terms, (14) and (15), in the previous expression are $o_p(1)$ using similar reasoning to that in the proof of Theorem 4 in DGM. The term (12) is

$$\frac{-\delta}{T}\sum_{t=2}^{T}\left(\varepsilon_{t-1}^2+\sum_{i=1}^{t-2}\frac{1}{(i+1)}\pi_i(-\eta_T)\varepsilon_{t-i-1}^2\right)\to_p -\delta K(d_1)$$

where

$$K(d_1)=\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\left(\sum_{i=0}^{t-2}\frac{\pi_i(-\eta_T)}{i+1}\right)=\sum_{i=0}^{\infty}\frac{\pi_i(d_1-1)}{i+1}.$$

Using a standard central limit theorem, the term (13) is

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\left(\varepsilon_t\varepsilon_{t-1}+\sum_{i=1}^{t-2}\pi_i(-\eta_T)\varepsilon_t\varepsilon_{t-1-i}\right)\to_d N(0,V)$$

where

$$
\begin{aligned}
V &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}E\left(\varepsilon_t\varepsilon_{t-1}+\sum_{i=1}^{t-2}\pi_i(-\eta_T)\varepsilon_t\varepsilon_{t-1-i}\right)^2\\
&= \lim_{t\to\infty}E\left(\sum_{i=0}^{t-2}\pi_i(d_1-1)\varepsilon_t\varepsilon_{t-1-i}\right)^2\\
&= \sum_{i=0}^{\infty}\pi_i(d_1-1)^2<\infty
\end{aligned}
$$

because $1-d_1<0.5$. Hence, $Q_T(d_1)\to_d N(-\delta K(d_1),\sum_{i=0}^{\infty}\pi_i(d_1-1)^2)$.

Second, consider the denominator of $t(d_1)$ scaled by $T^{-1/2}$. It is straightforward to show that $\widehat{S}_T^2(d_1)\to_p 1$, and, given the above expression for $\Delta^{d_1}y_{t-1}$, by a law of large numbers it is easy to see that

$$\frac{1}{T}\sum_{t=2}^{T}\left(\Delta^{d_1}y_{t-1}\right)^2\to_p \lim_{t\to\infty}E\left(\varepsilon_{t-1}+\sum_{i=1}^{t-2}\pi_i(d_1-1)\varepsilon_{t-1-i}\right)^2=\sum_{i=0}^{\infty}\pi_i(d_1-1)^2.$$

Hence,

$$t(d_1) \to_d N\left(\frac{-\delta K(d_1)}{\sqrt{\sum_{i=0}^{\infty} \pi_i(-\eta_T)^2}}, 1\right).$$

Finally, direct calculations lead to $K(d_1) = 1/d_1$ and to

$$\sum_{i=0}^{\infty} \pi_i(-\eta_T)^2 = \frac{\Gamma(2d_1 - 1)}{\Gamma(d_1)^2}.$$

Proof of Theorem 2. First, we consider the numerator of $t(d_1)$ scaled by $T^{-1/2}$. We want to show that for $d_1 \in D_1$

$$Q_T(d_1) \Rightarrow Q(d_1) - \delta/d_1,$$

where $Q(d_1)$ is a zero mean Gaussian process with covariance kernel given by

$$C^Q(d_1^a, d_1^b) = \sum_{i=0}^{\infty} \pi_i(d_1^a - 1)\pi_i(d_1^b - 1).$$

The finite dimensional distributions of each of the terms in which we decomposed Q_T, (12) to (15), have been analyzed in Theorem 1. Thus, it only remains to check their tightness. We start by analyzing in detail the second component of $Q_T(d_1)$. Using the Cramer-Wold device, (13) converges in distribution for each finite set J of values of d_1 to $N(0, V(J))$, where $V_{ab}(J) = C^Q(d_1^a, d_1^b)$. Now, define $X_t(d_1) = \varepsilon_t\varepsilon_{t-1} + \sum_{i=1}^{t-2} \pi_i(-\eta_T)\varepsilon_t\varepsilon_{t-1-i}$. In order to prove tightness, it is sufficient to show that, for and any $d_1^a, d_1^b \in D_1$,

$$E\left(\frac{1}{\sqrt{T}} \sum_{t=2}^{T} \left(X_t(d_1^a) - X_t(d_1^b)\right)\right)^2 \leq K|d_1^a - d_1^b|^\gamma \tag{16}$$

for some $\gamma > 1$, where $K > 0$ is a generic constant that does not depend on T or $\left(d_1^a, d_1^b\right)$. Then, using the i.i.d. property of the ε_t, the left hand side of (16) equals

$$\frac{\sigma^4}{T} \sum_{t=2}^{T} \sum_{i=1}^{t-2} \left(\pi_i(d_1^a - 1 - \delta/\sqrt{T}) - \pi_i(d_1^b - 1 - \delta/\sqrt{T})\right)^2. \tag{17}$$

Using the Mean Value Theorem for $\pi_i(\cdot)$, (17) is bounded by

$$\frac{K}{T} \sum_{t=2}^{T} \sum_{i=1}^{t-2} \frac{|d_1^a - d_1^b|^2}{i^2} \log^2 i \leq K|d_1^a - d_1^b|^2.$$

Next, define $Z_t(d_1) = \varepsilon_{t-1}^2 + \sum_{i=1}^{t-2}(i+1)^{-1}\pi_{i-1}(-\eta_T)\varepsilon_{t-i-1}^2$. The first term of $Q_T(d_1)$, (12), converges in probability to 0 uniformly in D_1, because it is $o_p(1)$ for each d_1, and it is tight using that

$$E\left(\frac{-\delta}{T}\sum_{t=2}^{T}\left(Z_t(d_1^a) - E\left[Z_t(d_1^a)\right] - Z_t(d_1^b) + E\left[Z_t(d_1^b)\right]\right)\right)^2 \leq K|d_1^a - d_1^b|^2,$$

and because

$$\sup_{d_1 \in D_1}\left|E\left[Z_t(d_1)\right] + K(d_1)\right| = o(1).$$

The last two components of $Q_T(d_1)$, (14) and (15), are also $o_p(1)$ uniformly in $d_1 \in D_1$, using similar arguments and the proof of Theorem 4 in DGM.

In addition, it is straightforward to show that $\sup_{d_1 \in D_1}\left|\widehat{S}_T^2(d_1) - 1\right| \to_p 0$, and that

$$\sup_{d_1 \in D_1}\left|\frac{1}{T}\sum_{t=2}^{T}\left(\Delta^{d_1}y_{t-1}\right)^2 - V(d_1)\right| \to_p 0.$$

Finally, the theorem follows by the Continuous Mapping Theorem.

REFERENCES

Andrews, D.W.K. and W. Ploberger (1994): "Optimal Tests when a Nuisance Parameter is Present Only under the Alternative," *Econometrica*, 62, 1383-1414.

Davies, R.B. (1977): "Hypothesis Testing when a Nuisance is Present Only under the Alternative," *Biometrika*, 64, 247-254.

_____ (1987): "Hypothesis Testing when a Nuisance is Present Only under the Alternative," *Biometrika*, 74, 33-43.

Delgado, M.A. and C. Velasco (2003): "Sign tests for long memory time series," Preprint, Universidad Carlos III de Madrid.

Dolado, J.J., J. Gonzalo and L. Mayoral (2002): "A Fractional Dickey-Fuller Test for Unit Roots," *Econometrica*, 70, 1963-2006.

Hansen, B.E. (1996): "Inference when a Nuisance Parameter is not Identified under the Null Hypothesis," *Econometrica*, 64, 413-430.

Mammen, E. (1993): "Bootstrap and Wild Bootstrap for High Dimensional Linear Models", *Annals of Statistics*, 21, 255-285.

Robinson, P. M. (1991), "Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression", *Journal of Econometrics*, **47**, 67-84.

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

Stute, W., W. González-Manteiga and M. Presedo-Quindimil (1998): "Bootstrap approximations in model checks for regression", *Journal of the American Statistical Association*, 83, 141-149.

Tanaka, K. (1999): "The Nonstationary Fractional Unit Root," *Econometric Theory*, 15, 249-264.

Velasco, C. (1999a): "Non-stationary log-periodogram regression," *Journal of Econometrics*, 91, 325-371.

_____ (1999b): "Gaussian semiparametric estimation of non-stationary time series," *Journal of Time Series Analysis*, 20, 87-127.

Velasco, C. and P.M. Robinson (2000): "Whittle Pseudo-Maximum Likelihood Estimates of Non-Stationary Time Series", *Journal of the American Statistical Association*, 95, 1229-1243.