OPTIMAL FRACTIONAL DICKEY-FULLER TESTS FOR UNIT ROOTS

By 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 considers the fractional Dickey-Fuller test for unit roots introduced recently by Dolado, Gonzalo and Mayoral (2002). The implementation of this test depends on a nuisance parameter that affects the power of the test. Since the arbitrary selection proposed by these authors is not optimal, in this article we investigate optimality aspects of the class of tests indexed by this parameter and propose feasible tests with good asymptotic and finite sample properties.

1 Introduction

Recently, Dolado, Gonzalo and Mayoral (2002, hereinafter DGM) have introduced a fractional Dickey-Fuller (hereinafter, FDF) test for unit roots which extends the popular Dickey-Fuller approach to fractionally integrated processes. In DGM's simplest framework y_t denotes a fractionally integrated process whose true order of integration is d, $\Delta^d y_t = \varepsilon_t$, where ε_t is i.i.d. with zero mean and finite variance. 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 the sequence $\pi_i(\alpha) = (i + \alpha - 1)!/(i!(\alpha - 1)!)$ are the coefficients of the binomial expansion.

The FDF test for the null hypothesis $d = 1$ versus either a simple alternative ($d = d_A$) or a composite alternative ($d < 1$), is based on the OLS estimation of the following model

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

where $d_1 < 1$ is a value chosen by the researcher. The FDF test statistic proposed by DGM is the t ratio associated with the OLS estimate of ϕ,

$$t(d_1) = \frac{\sum_{2}^{T} \Delta y_t \Delta^{d_1} y_{t-1}}{\hat{\sigma}_T(d_1) \sqrt{\sum_{2}^{T} (\Delta^{d_1} y_{t-1})^2}}, \tag{2}$$

where

$$\hat{\sigma}_T^2(d_1) = \frac{1}{T} \sum_{2}^{T} (\Delta y_t - \hat{\phi} \Delta^{d_1} y_{t-1})^2,$$

and T denotes the sample size. We emphasize that expression (2), which is equation (9) in DGM, represents a class of test statistics by writing explicitly the input value d_1 as an argument of the test statistic. DGM arbitrarily recommend to use $d_1 = d_A$ when the alternative hypothesis is simple, and to use $d_1 = \hat{d}$ when the alternative is composite, where $\hat{d}$ is a trimmed version of a $\sqrt{T}$-consistent estimator for d, such that $\hat{d}$ is strictly smaller than 1 with probability one.

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 $\Delta^{d_1} y_{t-1}$, that is,

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

The previous limits will be evaluated in Section 3. The key insight is that the parameter ϕ is different for each d_1, so its dependence on d_1 has been stressed by writing it as $\phi(d_1)$. Notice that the null hypothesis implies $\phi(d_1) = 0$ for all $d_1 \in [0, 1)$, so the testing framework is similar, for instance, to that considered in Section 2 in Hansen (1996). This situation is nonstandard since d_1 does not appear in the regression under the null hypothesis, so that d_1 can be regarded as a nuisance parameter that is not identified under the null hypothesis. 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 nuisance parameter can take. However, building on the results in DGM, it can be deduced that there is no need to employ all these values. In fact, DGM prove that the asymptotic null distribution of $t(d_1)$ is pivotal when the nuisance parameter d_1 is any constant in $[1/2, 1)$ or a (truncated) $\sqrt{T}$-consistent estimator for $d = 1$, and that the $t(d_1)$ test is consistent when the nuisance parameter d_1 takes any constant value in $[0, 1)$, (see Theorems 2, 3 and 5 in DGM). Hence, these results suggest that simple and powerful tests can be derived by carefully choosing a single nuisance parameter d_1 (although DGM simply propose an arbitrary choice of d_1 instead of pursuing an optimal selection).

In this article we analyze optimality aspects of the FDF test within the framework of econometric problems where a nuisance parameter is not identified under the null hypothesis. More precisely, we compute the asymptotic power function of the FDF test for each d_1 under a sequence of local alternatives that converge to the null at the parametric rate and derive an optimal selection for the d_1 parameter, $d^* \simeq 0.69$, which is consistent against these alternatives. For fixed alternatives we introduce the maximal squared correlation as a criterion function to select 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. We show that model-free semiparametric estimators can be used, since a parametric rate is not necessary as claimed by DGM. In practice, this result is important since semiparametric estimators do not demand the correct specification of a parametric model.

The plan of the article is the following. In Section 2 we derive an optimal selection for the d_1 parameter under a sequence of local alternatives that converges to the null at the parametric rate. Based on this framework, we also analyze testing procedures that consider all the values of d_1 in a given interval. Section 3 studies optimal tests for fixed alternatives and contains a brief Monte Carlo exercise that compares the finite sample performance of the considered tests. Section 4 concludes.

2 An optimal FDF test for local alternatives

In order to motivate the problem, 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 $d_1 = \{0, 0.1, ..., 0.9\}$. The sample size is 100 and the number of replications is 30,000. Simulations have been carried out in Fortran 90 double precision. 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. 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 choosing optimally 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^*. Since DGM's results are based on just 1,000 replications, a first impression from DGM's Figure 2 could be that sampling error is causing the power variations. However, our Table 1 shows that sampling error cannot explain the fact that the optimal d_1 is not the true d.

It could be argued that the previous 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 choosing d_1 optimally.

From the previous simulation results it is clear that some criterion to select d_1 optimally 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 choose the value of d_1 that 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 possible values of d_1.

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

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

the asymptotic distribution of the test statistic $t(d_1)$ is given by:

(i) if $0 \leq d_1 < 0.5$

$$t(d_1) \xrightarrow{w} -\frac{\int_0^1 W_{-d_1}(r)dB(r)}{\sqrt{\int_0^1 W_{-d_1}^2(r)dr}}, \tag{3}$$

where $W_{-d_1}(r)$ and $B(r)$ are defined as in DGM;

(ii) if $0.5 \leq d_1 < 1$

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

where

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

Γ represents the gamma function and $h(d_1)$ achieves its maximum at $d_1 = d^* \simeq 0.69145$.

The proof of the theorem is in the Appendix. Note that DGM's Theorem 4 also analyzes local alternatives but it just considers the case $d_1 = d = 1 - \delta/\sqrt{T}$. The implications of this theorem are the following. First, for the case $0 \leq d_1 < 0.5$ the asymptotic distribution of $t(d_1)$ is the same as the corresponding one in Theorem 2 in DGM. Hence, choosing a value for d_1 in the interval $[0, 0.5)$ delivers a test with trivial asymptotic local power against a sequence of alternatives tending to the null at the $T^{-1/2}$ rate. In fact, inspection of the proof of Theorem 1(i) reveals that in order to have non-trivial power the sequence of local alternatives has to tend to the null at the rate T^{-d_1} that is slower than $T^{-1/2}$. Second, to our knowledge, the result that the Dickey-Fuller has trivial power against local fractional alternatives converging to the null at any rate T^{-a}, $a > 0$, is new. Third, for the $0.5 \leq d_1 < 1$ case the noncentrality parameter is a function of d_1 that achieves a maximum at d^*, see the plot of the function $h(d_1)$ in Figure 1. In the same plot we have added the horizontal line $\sqrt{\pi^2/6}$ that represents the noncentrality parameter achieved by the optimal Robinson-Tanaka test. Notice that the two lines are quite close at $d_1 = d^*$, 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 $d_1 \leq 0.5$. 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 $T^{-1/2} \log T$.

Therefore, the previous theorem indicates that an optimal, with respect to an asymptotic local power criterion, implementation of the FDF test requires selecting $d_1 = d^*$. An alternative approach to the FDF testing procedure employs test statistics which take into account simultaneously many values of the nuisance parameter d_1. The distributional theory

necessary for such analysis has been developed in other contexts for a sequence of local alternatives similar to the one considered above, see Davies (1977, 1987), Andrews and Ploberger (1994) or Hansen (1996). In the rest of this section we briefly analyze this approach.

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 $D = [0.5, 1)$, 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, $\mathbf{C}(D_1)$, 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

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

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 $\mathbf{C}(D_1)$ endowed with the uniform metric, $W(d_1)$ is a zero mean Gaussian process with covariance kernel given by

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

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

In particular, notice that under the null hypothesis

$$t(d_1) \Rightarrow W(d_1), \tag{4}$$

and that for each $d_1 \in D_1$ the asymptotic distribution of $t(d_1)$ is the standard normal, agreeing with Theorem 1(ii).

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 (4) 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. In the next section we will investigate briefly the finite sample behavior of these tests.

3 An optimal FDF test for fixed alternatives

In the previous section we have selected d_1 optimally by maximizing the asymptotic power function of the test against a sequence of local alternatives. Next, we show that an equivalent criterion to optimally set d_1 is to maximize the squared correlation between Δy_t and $\Delta^{d_1} y_{t-1}$. Recall that the criterion of maximizing the power against a sequence of local alternatives leads to restrict the attention to the range $d_1 \in (0.5, 1)$, for which the asymptotic null distribution of the $t(d_1)$ statistic does not depend on d_1. Hence, maximizing the power is equivalent to finding the value of d_1 that maximizes $t(d_1)^2$ for each sample. Equivalently, denoting by $R^2(d_1)$ the squared sample correlation between Δy_t and $\Delta^{d_1} y_{t-1}$, that is,

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

the basic relation of simple regression theory,

$$t(d_1)^2 = T\frac{R^2(d_1)}{1 - R^2(d_1)}, \tag{5}$$

implies that maximizing $t(d_1)^2$ is equivalent to maximizing $R(d_1)^2$.

Therefore, in order to introduce a criterion to select optimally d_1, the natural one is the population analog of $R^2(d_1)$, that is, the squared population correlation between Δy_t and $\Delta^{d_1} y_{t-1}$, that we call $\rho^2(d_1)$. Denote the optimal d_1 by $d_1^* = \arg\max_{d_1} \rho^2(d_1)$. Since d_1 does not appear on the variance of Δy_t,

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

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 $\Delta^{d_1-d}\varepsilon_{t-1} = \sum_{i=0}^{t} \pi_i(d_1-d)\varepsilon_{t-1-i}$,

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

if

$$d_1 - d > -0.5, \tag{6}$$

and in this case, $\sum_{i=0}^{\infty} \pi_i(d_1-d)^2 = \Gamma(2d_1-2d+1)/\Gamma(d-d_1-1)^2$. Now, 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, as long as $d_1 - d > -0.5$

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

where

$$L(d, d_1) = \frac{\left(\sum_{i=1}^{\infty} \pi_i(1-d)\pi_{i-1}(d_1-d)\right)^2}{\Gamma(2d_1-2d+1)/\Gamma(d-d_1-1)^2}. \tag{7}$$

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 optimal value of d_1 for each value of d. Unfortunately, this is quite complicated for a general d. When $d = 0$, it is easy to see from the original equations that the optimal selection of d_1 is 0, justifying the optimality of the standard Dickey-Fuller test over the Fractional Dickey-Fuller for the case $d = 0$. 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 (7). However, we can approximate $d_1^* = d_1^*(d)$ numerically with any level of precision, and in Table 3 we report the d_1^* function for some values of d and a truncation at $i = 10^5$ in the infinite sum in (7). Confirming the results of the previous Monte Carlo experiments, Table 3 shows that the optimal value for d_1 is always below the true d. Table 3 also indicates that the relation between d_1^* and d is essentially linear when $d \geq 0.5$, whereas for lower values of d there is some curvature. In Figure 2 we have plotted

the results of Table 3, and we have added the regression line of $d_1^*(d)$ on d for $d \geq 0.5$. This fit is given by $\hat{d}_1^*(d) = -0.030 + 0.717d$ using 10 values of d, 0.5(0.05)0.95, and a truncation at $i = 10^5$ in the infinite sum in (7). The standard error of this estimation is 0.0004. Notice that $\hat{d}_1^*(d) - d > -0.5$, so that the condition (6) is always satisfied. In addition, in agreement with the results of the previous section, $\hat{d}_1^*(1)$ is very close to d^*, the discrepancy is just due to the numerical error in the approximation.

For the simple alternative case, the previous analysis indicates that $d_1^* = 0$ when $d_A = 0$, whereas when $d_A > 0$ our simple proposal is to select $\hat{d}_1^*(d_A) = -0.030 + 0.717d_A$. For the more interesting composite alternative case, notice that $\hat{d}_1^*(d)$ is still unfeasible, but similarly to DGM we could replace d by a $\sqrt{T}$-consistent estimator in the expression of $\hat{d}_1^*(d)$. In fact, a careful inspection of the proof of Theorem 5 in DGM shows that their results still hold when the estimator $\hat{d}$ of the true d satisfies $T^{1/4} \log T \quad \hat{d} - d \quad = o_p(1)$. This condition holds for many semiparametric estimators for an appropriate choice of the bandwidth parameter, see Velasco (1999a, b). DGM overlooked this result but it is very important because it means that the FDF test is consistent even if the researcher does not specify correctly the parametric model. Obviously, for the simple framework where y_t follows a pure fractional process, a parametric estimator, such as the Whittle estimator that we employ in the simulations at the end of this section, is more efficient than a semiparametric estimator. The role of a semiparametric estimator appears when this restrictive framework is relaxed and more complicated data generating processes are allowed. In these cases, instead of the FDF test, an Augmented FDF test that employs a semiparametric estimator should be used.

Hence, in order to derive a simple operational modification of the FDF test our proposal is to implement the FDF test with

$$\hat{d}_1^*(\hat{d}) = -0.030 + 0.717\hat{d} \tag{8}$$

where $\hat{d}$ can be any consistent semiparametric estimator of d such as those proposed in Velasco (1999a, b). Similarly to DGM, in practice, for a moderate sample size $\hat{d}_1^*(\hat{d})$ will always be below 1, but for very small sample sizes, a trimming such as the one proposed by DGM in their equation (33) would be necessary. From now on, $\hat{d}_1^*(\hat{d})$ will denote the trimmed version.

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

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

where $\hat{d}_1^*(\hat{d})$ is a trimmed version of equation (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. The only difference is to realize that expression (A.48) in DGM is bounded as $O_p(T^{1/4}\log T)$ rather than $o_p(T^{1/2})$, justifying the use of a semiparametric estimator. The intuition of the lemma is also similar to DGM: under the null $\hat{d}_1^*(\hat{d})$ will be close to d^*, and hence the asymptotic standard distribution is the standard normal.

Our choice of d_1 is asymptotically optimal for $d \geq 0.5$, 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 considered FDF tests with several selections for d_1, namely a) $d_1 = d$, denoted by FDF(d), b) $d_1 = d^*$, FDF(d^*), c) $d_1 = \hat{d}_1^*(d)$, FDF($\hat{d}_1^*(d)$), d) $d_1 = \hat{d}_1^*(\hat{d}_n)$, where $\hat{d}_n$ is the Whittle parametric estimator, FDF($\hat{d}_1^*(\hat{d}_n)$), and e) $d_1 = \hat{d}_1^*(\hat{d}_m)$, where $\hat{d}_m$ is the Gaussian semiparametric estimator with bandwidth m, FDF($\hat{d}_1^*(\hat{d}_m)$). 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. 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). In addition, similarly to DGM, a trimming rule was introduced for calculating $\hat{d}_n$ and $\hat{d}_m$, such that these estimators are always less or equal than 0.99. For $\hat{d}_m$, the selected bandwidth is $m = n^{0.55}$, which is sufficient for our asymptotic theory to hold. Finally, for the tests a) and c) 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 Tanaka's (1999) 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 our framework 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

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

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 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 4 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($\hat{d}_1^*(\hat{d}_m)$) presents a severe size distortion for this sample size. On the contrary, all the non-FDF tests appear to be conservative. Hence, in Table 5 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($\hat{d}_1^*(d)$) 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. These results suggest that there is no need to employ this type of tests that considers all the range of d_1, since an optimal selection for d_1 is available.

These findings are based on a very simple DGP. Additional Monte Carlo experiments for more complicated DGP's are needed. However, two preliminary conclusions arise from this finite sample evidence. First, the FDF(d^*) test is a very sensible option 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 proposed by DGM is likely to become very popular among applied researchers. In this article we have analyzed the FDF test with a model where a nuisance parameter, d_1, is not identified under the null hypothesis. In this framework two approaches can be considered. The first approach employs tests statistics that use a unique value of d_1. The second approach employs all values in an interval.

The FDF test belongs to the first approach. In their original proposal, DGM arbitrarily select the value of d_1, so that their implementation of the FDF test is not theoretically optimal, and in practice, their arbitrary selection of d_1 may lead to a severe loss of power in finite samples. In this article we have addressed the issue of optimally selecting the value of d_1. Optimality considerations demand the introduction of a criterion function. Since one of the main advantages of the FDF approach is that it does not require the introduction of a particular parametric model or distribution, a natural criterion is $\rho^2(d_1)$, the population squared correlation between the dependent and independent variables of regression (1). Using this criterion function, we have derived optimal tests that are consistent against alternatives that converge to the null at the parametric rate.

In the spirit of the second type of tests, we have analyzed two popular test statistics, the Kolmogorov-Smirnov and the Cramer-von Mises. However, in our case there is no apparent theoretical reason to justify the use of these statistics because the asymptotic null distribution of the statistic $t(d_1)$ is pivotal, leading to an optimal solution for the d_1 parameter in terms of d. In addition, we have shown that a semiparametric estimator of d can be plugged into the FDF test, rendering unnecessary the full specification of a parametric model. An optimality analysis similar to Andrews and Ploberger (1994) merits further study, but it is beyond the scope of this article.

Finally, 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, we presume that the asymptotic null distribution of the t ratio statistic associated to the coefficient of the regressor $\Delta^{d_1^*(d)} y_{t-1}$ is still the standard normal, as long as the number of included lags of Δy_t in the augmented regression is large enough to capture asymptotically all the serial correlation in the errors. Notice that in this case a semiparametric optimal implementation of the augmented FDF test would demand, in principle, the introduction of two user-chosen numbers: one reflects the number of lags included in the augmented FDF regression, and the other necessary to estimate consistently d, and hence, to estimate consistently the optimal $d_1^*(d)$. We stress that a semiparametric estimator is especially relevant for this general framework, since specifying a correct model for the autocorrelation of the errors ε_t can be particularly difficult. In this case, optimal augmented FDF tests are consistent against alternatives that converge to the null at the parametric rate. This feature is shared by the optimal tests developed by Robinson (1994) or Tanaka (1999). However, compared to them, optimal augmented FDF tests present the advantage of not requiring the correct specification of a parametric model.

ACKNOWLEDGMENT

We thank Josu Arteche, Miguel Delgado, Luis Gil-Alaña, Jesús Gonzalo, Javier Hidalgo and Pierre Perron for useful conversations. Lobato acknowledges financial support from Asociación Mexicana de Cultura. 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	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d_1^*	0.000	0.0619	.1261	.1921	.2600	.3294	.4001	.4716	.5437	0.6158	0.6515

Table 3. The function $d_1^*(d)$ computed from (7) with a truncation at $i = 10^5$.

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($\hat{d}_1^*(d)$)	100	100	98.2	77.6	32.2	5.57
FDF($\hat{d}_1^*(\hat{d}_n)$)	100	99.5	94.8	71.8	30.6	5.65
FDF($\hat{d}_1^*(\tilde{d}_m)$)	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
TAN	99.9	99.0	91.2	62.9	24.5	4.64

Table 4. Monte Carlo size ($d = 1$) and power ($d < 1$) of the FDF tests with several selections for d_1, the KS, the CvM and Tanaka's test: Percentage of rejections based on 5% asymptotic critical values for all 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($\hat{d}_1^*(d)$)	100	100	97.8	75.5	30.0
FDF($\hat{d}_1^*(\hat{d}_n)$)	100	99.4	94.1	69.6	28.5
FDF($\hat{d}_1^*(\tilde{d}_m)$)	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
TAN	99.9	99.1	91.9	64.5	25.7

Table 5. Monte Carlo power of the FDF tests with several selections for d_1, the KS, the CvM and Tanaka's 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 100,000.

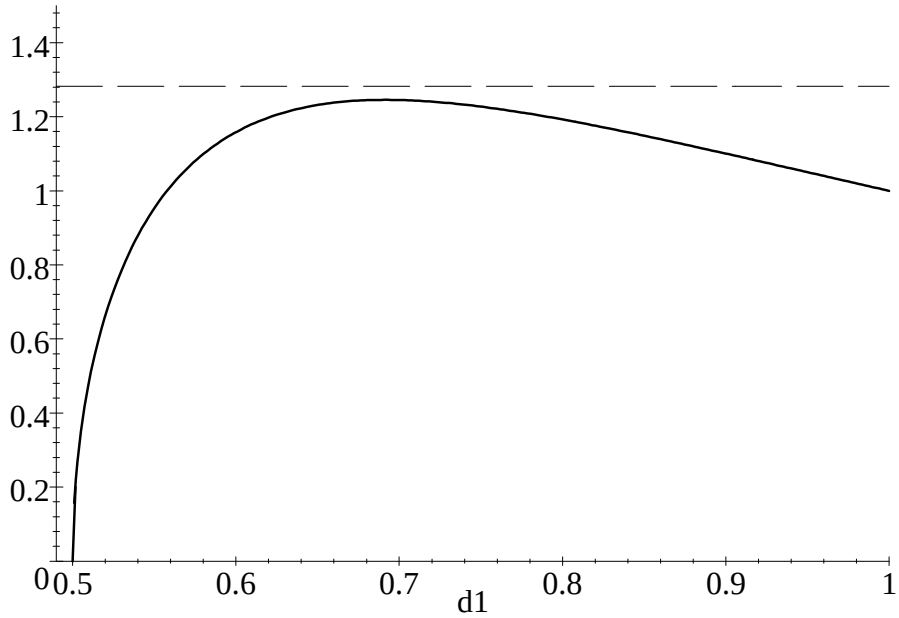

Figure 1. Asymptotic efficiency of the FDF and LM tests: plots of $h(d_1)$ and $\sqrt{\pi^2/6}$.

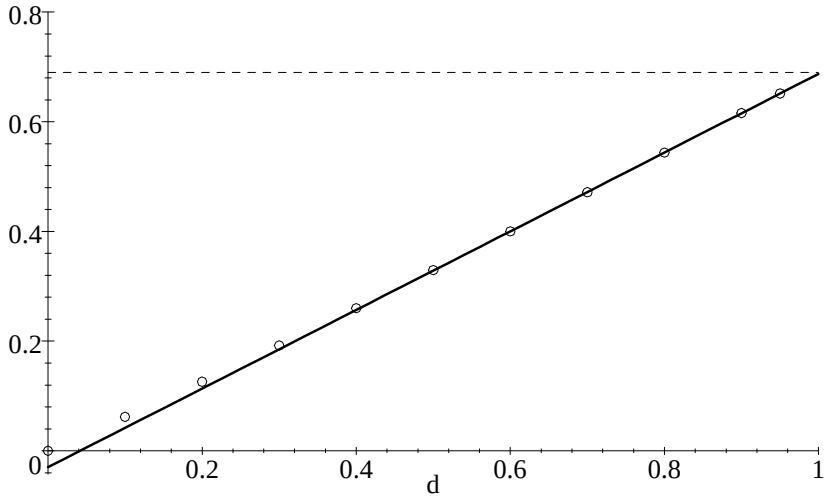

Figure 2. Plots of the points $(d, d_1^*(d))$ of Table 3, and the lines $d_1 = \sqrt{h}\, d_1^*(d)$ and $d_1 = d^* \equiv 0.69$.

5 Appendix

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

Proof of Theorem 1. We start with the proof of (ii).

We start introducing some notation. Let

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

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 $\pi_j(-\theta_T) \approx -j^{-1}\delta T^{-1/2}$. Also,

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

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 $T^{-1/2}$,

$$\begin{aligned}
Q_T(d_1) &= T^{-1/2} \sum_{2}^{T} \Delta y_t \Delta^{d_1} y_{t-1} \\
&= T^{-1/2} \sum_{2}^{T} \left(\varepsilon_t + \sum_{1}^{t-1} \left(\frac{1}{i}\frac{-\delta}{\sqrt{T}} \right) \varepsilon_{t-i} \right) \left(\varepsilon_{t-1} + \sum_{1}^{t-2} \pi_i(-\eta_T)\varepsilon_{t-1-i} \right) + o_p(1)
\end{aligned}$$

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

$$\begin{aligned}
&= T^{-1/2} \sum_{t=2}^{T} \left(\left(\frac{-\delta}{\sqrt{T}} \right) \varepsilon_{t-1}^2 + \sum_{i=1}^{t-2} \left(\frac{1}{(i+1)}\frac{-\delta}{\sqrt{T}} \right) \pi_i(-\eta_T)\varepsilon_{t-i-1}^2 \right) && (10) \\
&+ T^{-1/2} \sum_{2}^{T} \left(\varepsilon_t \left(\varepsilon_{t-1} + \sum_{1}^{t-2} \pi_i(-\eta_T)\varepsilon_{t-1-i} \right) \right) && (11) \\
&+ T^{-1/2} \sum_{2}^{T} \left(\left(\frac{-\delta}{\sqrt{T}} \right) \varepsilon_{t-1} \left(\sum_{1}^{t-2} \pi_i(-\eta_T)\varepsilon_{t-1-i} \right) \right) && (12) \\
&+ T^{-1/2} \sum_{2}^{T} \left(\sum_{1}^{t-2} \left(\frac{1}{(i+1)}\frac{-\delta}{\sqrt{T}} \right) \varepsilon_{t-i-1} \left(\sum_{i \neq j=1}^{t-2} \pi_i(-\eta_T)\varepsilon_{t-1-j} \right) \right) + o_p(1). && (13)
\end{aligned}$$

The last two terms, (12) and (13), in the previous expression are $o_p(1)$ using similar reasoning to that in the proof of Theorem 4 in DGM. The term (10) 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 (11) is

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

where

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

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 $\hat{b}_T^2(d_1)\to_p 1$, and, given the above expression for $\Delta^{d_1}y_{t-1}$, by a law of large numbers it is easy to see that

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

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

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

Next, we prove part (i). As above, it is straightforward to show that $\hat{S}_T^2(d_1) \to_p 1$. First, we analyze the numerator of $t(d_1)$ scaled by T^{d_1-1},

$$T^{d_1-1} \sum_{t=2}^{T} \Delta y_t \Delta^{d_1} y_{t-1} = T^{d_1-1} \sum_{t=2}^{T} \left(\varepsilon_t + \sum_{i=1}^{t-1} \left(\frac{1}{i} \frac{-\delta}{\sqrt{T}} \right) \varepsilon_{t-i} \right) (w_{t-1} + z_{t-2}) \tag{14}$$

where

$$w_{t-1} = \varepsilon_{t-1} + \sum_{i=1}^{t-2} \pi_i(d_1 - 1)\varepsilon_{t-1-i}$$

and

$$z_{t-2} = \sum_{i=1}^{t-2} \left(\pi_i(-\eta_T) - \pi_i(d_1 - 1) \right) \varepsilon_{t-1-i}.$$

There are four terms in (14). The first is the same that was derived in Theorem 2 in DGM

$$T^{d_1-1} \sum_{t=2}^{T} \varepsilon_t w_{t-1} \xrightarrow{w} \int_0^1 W_{-d_1}(r) dB(r).$$

The second has zero mean and verifies that

$$T^{d_1-1} \sum_{t=2}^{T} \varepsilon_t z_{t-2} \to_p 0.$$

The third term is

$$\begin{aligned} T^{d_1-1} \sum_{t=2}^{T} \sum_{i=1}^{t-1} \left(\frac{1}{i} \frac{-\delta}{\sqrt{T}} \right) \varepsilon_{t-i} w_{t-1} &= T^{d_1-1} \sum_{t=2}^{T} \sum_{i=1}^{t-1} \left(\frac{1}{i} \frac{-\delta}{\sqrt{T}} \right) \varepsilon_{t-i} \left(\varepsilon_{t-1} + \sum_{i=1}^{t-2} \pi_i(d_1 - 1)\varepsilon_{t-1-i} \right) \\ &= -\delta T^{d_1-3/2} \sum_{t=2}^{T} \sum_{i=1}^{t-2} \frac{\pi_{i+1}(d_1 - 1)}{i}. \end{aligned}$$

If $d_1 > 0$ this term is $O_p(T^{d_1-1/2}) = o_p(1)$, while if $d_1 = 0$ the term is $O_p((\log T)T^{d_1-1/2}) = o_p(1)$. We emphasize the rate of convergence to zero, since it is the main input to determine the rate of convergence at which a sequence of local alternatives has to tend to the null so that the $t(d_1)$ has non-trivial power in the $0 \leq d_1 < 0.5$ case.

Finally, the fourth term

$$T^{d_1-1} \sum_{t=2}^{T} \left(\sum_{i=1}^{t-1} \left(\frac{1}{i} \frac{-\delta}{\sqrt{T}} \right) \varepsilon_{t-i} \right) z_{t-2}$$

is $O_p(T^{d_1-1/2}) = o_p(1)$ if $d_1 > 0$, and when $d_1 = 0$ it is $O_p((\log T)T^{d_1-1/2}) = o_p(1)$, as above. Hence, the numerator of $t(d_1)$ scaled by T^{d_1-1} verifies that

$$T^{d_1-1} \sum_{t=2}^{T} \Delta y_t \Delta^{d_1} y_{t-1} \xrightarrow{w} \int_0^1 W_{-d_1}(r) dB(r).$$

Second, we examine the denominator of $t(d_1)$ scaled by T^{d_1-1}. As above, we write $\Delta^{d_1} y_{t-1} = w_{t-1} + z_{t-2}$. Then,

$$T^{2d_1-2} \sum_{2}^{T} \left(\Delta^{d_1} y_{t-1}\right)^2 = T^{2d_1-2} \sum_{2}^{T} \left(w_{t-1} + z_{t-2}\right)^2 . \tag{15}$$

Expression (15) has three terms. The first one is the one that appears in Theorem 2 in DGM

$$T^{2d_1-2} \sum_{2}^{T} w_{t-1}^2 \xrightarrow{w} \int_0^1 W_{-d_1}^2(r)dr.$$

The second term is equal to

$$T^{2d_1-2} \sum_{2}^{T} z_{t-2}^2 = T^{2d_1-2} \sum_{2}^{T} \left(\sum_{1}^{t-2} \left(\pi_i(-\eta_T) - \pi_i(d_1-1)\right) \varepsilon_{t-1-i} \right)^2 .$$

Using the mean value theorem $\pi_i(-\eta_T) - \pi_i(d_1-1) = \delta \pi_i'(d_1-1)/\sqrt{T} + O(T^{-1} i^{-d_1} \log^2 i)$ as $T \to \infty$ and $i \to \infty$, and then

$$\begin{aligned} T^{2d_1-2} \sum_{2}^{T} z_{t-2}^2 &= \delta^2 T^{2d_1-3} \sum_{2}^{T} \left(\sum_{1}^{t-2} \pi_i'(d_1-1)\varepsilon_{t-1-i} \right)^2 + o_p(1) \\ &= O_p\left(T^{-1} \log^2 T\right) = o_p(1) \end{aligned}$$

since $\pi_i'(d_1-1) \sim C\ i^{-d_1} \log i$.

Finally, the last term in expression (15) is

$$\begin{aligned} & T^{2d_1-2} \sum_{2}^{T} w_{t-1} z_{t-2} \\ = \; & T^{2d_1-2} \sum_{2}^{T} \left(\varepsilon_{t-1} + \sum_{1}^{t-2} \pi_i(d_1-1)\varepsilon_{t-1-i} \right) \left(\sum_{1}^{t-2} \left(\pi_i(-\eta_T) - \pi_i(d_1-1)\right) \varepsilon_{t-1-i} \right) \\ = \; & O_p\left(T^{2d_1-3/2} \sum_{1}^{T} \pi_i(d_1-1)\pi_i'(d_1-1) \right) = O_p\left(T^{-1/2} \log T\right) = o_p(1). \end{aligned}$$

Hence, the theorem follows.

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, (10) to (13), 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, (11) 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 $\{d_1^a, d_1^b\}$. 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 $Q_T(d_1)$,(10), 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 $Q_T(d_1)$, (12) and (13), 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 $\sup_{d_1 \in D_1} \hat{\sigma}_T^2(d_1) \to_p 1$, 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.

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. (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.