Optimal Fractional Dickey–Fuller tests

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 Economía, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe (Madrid), Spain
E-mail:carlos.velasco@uc3m.es

Received: November 2005

Summary This article analyzes the fractional Dickey–Fuller (FDF) test for unit roots recently introduced by Dolado, Gonzalo and Mayoral (2002 *Econometrica 70*, 1963–2006) within a more general setup. These authors motivate their test with a particular analogy with the Dickey–Fuller test, whereas we interpret the FDF test as a class of tests indexed by an auxiliary parameter, which can be chosen to maximize the power of the test. Within this framework, we investigate optimality aspects of the FDF test and show that the version of the test proposed by these authors is not optimal. For the white noise case, we derive simple optimal FDF tests based on consistent estimators of the true degree of integration. For the serial correlation case, optimal augmented FDF (AFDF) tests are difficult to implement since they depend on the short-term component. Hence, we propose a feasible procedure that automatically optimizes a prewhitened version of the AFDF test and avoids this problem.

Key words: serial correlation, long memory, unit roots.

1. INTRODUCTION

In a recent paper, Dolado, Gonzalo and Mayoral (2002, hereinafter DGM) have introduced a fractional Dickey–Fuller (hereafter FDF) test for testing the null of unit root against the alternative of fractional integration. In DGM's simplest framework, y_t denotes a fractionally integrated process whose true order of integration is d,

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

where ε_t are independent and identically distributed (i.i.d.) random variables with zero mean and finite variance, L is the lag operator and the fractional difference operator is defined as in DGM.

DGM considered 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 Δy_t on $\Delta^{d_1} y_{t-1}$. That is, DGM considered the OLS estimation of the model

$$\Delta y_t = \phi \Delta^{d_1} y_{t-1} + u_t, \quad t = 1, \ldots, T, \tag{2}$$

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

$$t(d_1) = \frac{\sqrt{T}\sum_{t=2}^{T}\Delta y_t \Delta^{d_1} y_{t-}}{\sqrt{\sum_{t=2}^{T}\left(\Delta y_t - \widehat{\phi}\Delta^{d_1} y_{t-1}\right)^2 \sum_{t=2}^{T}\left(\Delta^{d_1} y_{t-1}\right)^2}}, \tag{3}$$

where T denotes the sample size. DGM established this test based on an analogy with the Dickey–Fuller (hereafter DF) test, and they interpreted d_1 as ‘the true value of d under the alternative hypothesis’. By contrast, we argue that the parameter d_1 in (2) has a very concrete statistical meaning, since it defines a class of tests indexed by d_1, as it is emphasized in expression (3) 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 and asymptotically more powerful implementations of the FDF test. In Section 2, we will see that maximizing the power of the FDF test is achieved by maximizing the correlation between Δy_t and $\Delta^{d_1} y_{t-1}$. Hence, we interpret d_1 as ‘a parameter that determines the power of the FDF test’ and the optimal d_1 is the value that maximizes the power of the FDF test. The important difference with DGM is that d_1 is not some arbitrary value derived from DGM’s analogy with the DF test, but a parameter that the researcher should choose to maximize the correlation between Δy_t and $\Delta^{d_1} y_{t-1}$, and hence to maximize the power of the FDF test.

Although the optimal d_1 is precisely defined under the alternative hypothesis (as the parameter that maximizes the correlation between Δy_t and $\Delta^{d_1} y_{t-1}$), it is not identified under the null hypothesis since in this case Δy_t and $\Delta^{d_1} y_{t-1}$ are uncorrelated for any d_1. This fact hampers the use of $d_1 = \overline{d}_1$, where $\overline{d}_1$ is the argument that maximizes the squared sample correlation between Δy_t and $\Delta^{d_1} y_{t-1}$, as we will discuss in Section 4. Instead of following this approach, we will directly pursue optimal implementations of the FDF test. When ε_t is white noise, we will show that there is a simple optimal selection for d_1 as a function of the true d. However, in the serial correlation case the optimal value of d_1 also depends on the short-term component. In order to arrive at an optimal implementation, we will propose a feasible procedure that automatically optimizes a prewhitened version of the augmented FDF (AFDF) test. This test procedure is based on an algorithm that avoids the lack of identification of the auxiliary parameter under the null hypothesis, because it employs differentiated versions of the original series.

The plan of the article is as follows. Section 2 studies in detail the case where the data generating process (DGP) is given by (1) where ε_t is white noise and derive optimal FDF tests both in a local alternative framework and in a fixed alternative framework where a consistent estimator for d is available. Sections 3 and 4 consider the serial correlation case. Section 3 introduces the prewhitened AFDF (PAFDF) test and derives the asymptotic local power of both tests, the AFDF and the PAFDF. Section 4 proposes the automatic optimal implementation of the PAFDF test. Section 5 reports a brief Monte Carlo exercise to compare the finite sample performance of the considered tests. Finally, Section 6 concludes. For simplicity, we have followed the notation in DGM as close as possible.

2. OPTIMAL FDF TESTS: THE WHITE NOISE CASE

DGM’s FDF test depends on the choice of the parameter d_1 to run regression (2), but there is not an obvious selection of such value. Since the choice of d_1 has important implications on the power properties of the test, as it is clear from several simulation results in DGM's paper, in this section we will derive optimal selections for d_1 in the white noise case. We first derive the value of d_1 that maximizes the asymptotic 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 null distribution of $t(d_1)$ is the standard normal only for these values of d_1, and therefore power comparisons are analytically tractable. Note that this analysis includes the case $d_1 = 1$ that would not make sense under DGM's analysis since $d_1 = 1$ is the value of d under the null hypothesis. DGM's theorem 4 also studies local alternatives but, following their interpretation, they just consider the case $d_1 = d = 1 - \delta/\sqrt{T}$.

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

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

where ε_t is i.i.d. with finite fourth 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 non-centrality 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 $d_1 = d_1^* \simeq 0.69145$, $h(d_1^*) \simeq 1.2456$, 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. This theorem is remarkable because it shows that there exists a unique optimal d_1 independent of δ for testing against local hypotheses. Hence, since the optimal d_1 is a fixed number greater than 0.5, the asymptotic null distribution of the t-statistic evaluated at d_1^* is the standard normal (cf. theorem 2 in DGM). Note that the asymptotic relative efficiency of the original DGM's FDF test with respect to this locally optimal implementation is 0.81.

In Figure 1, we have added a horizontal line at $1 = h\ (1)$, which is the non-centrality parameter for DGM's original proposal, $d_1 = d = 1 - \delta/\sqrt{T} \to 1$. Note that employing any value of d_1 between 0.5578 and 1 leads to an FDF test with more asymptotic local power than DGM's original proposal. Also note 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 $T^{-1/2} \log T$. 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.

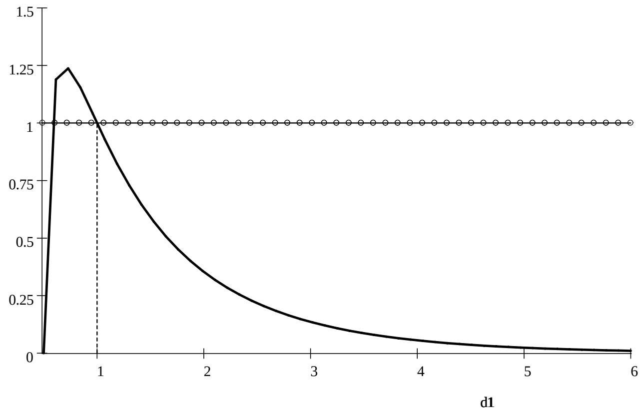

Figure 1. Asymptotic efficiency of the FDF tests: plots of $h(d_1)$ and $h(1) = 1$. The point (1, 1) corresponds to DGM proposal.

We now consider a complementary criterion to select optimally d_1 in a fixed alternative framework where $d_A \in (0.5, 1)$. Since the asymptotic null distribution of the $t(d_1)$ statistic is the standard normal for any $d_1 \geq 0.5$, maximizing the power for this range of values for d_1 is equivalent to finding the value of d_1 that maximizes the probability limit of $t(d_1)^2$, properly standardized. In addition, recall the basic relation of simple regression theory,

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

where $R^2(d_1)$ denotes the squared sample correlation between Δy_t and $\Delta^{d_1} y_{t-1}$, that is,

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

Equation (4) establishes a monotonic increasing relation between $R^2(d_1)$ and $t(d_1)^2$, which 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 optimal 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 optimal d_1 by

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

Since d_1 does not appear on the variance of Δy_t,

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

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

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

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

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

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

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

Hence,

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

where

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

In agreement with the previous results, when $d = 1 - \delta/\sqrt{T}$ the optimal selection of d_1 is $d_1^* = d_1^*(d) \simeq 0.69$. 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 2, we report the $d_1^*(d)$ for some values of d and a truncation at $i = 10^5$ in the infinite sum in (6). Figure 2 shows that d_1^* is always below the true d. Figure 2 also indicates that the relation between d_1^* and d is essentially linear for the range $d_1^* \geq 0.5$. In Figure 2, we have added the regression line of $d_1^*(d)$ on d. This fit is given by $\widehat{d_1^*}(d) = -0.031 + 0.719d$ using a truncation at $i = 10^5$ in the infinite sum in (6). The standard error of this regression estimation is 0.0004. Note that $\widehat{d_1^*}(d) - d > -0.5$, so that the condition (5) is always satisfied. In addition, and

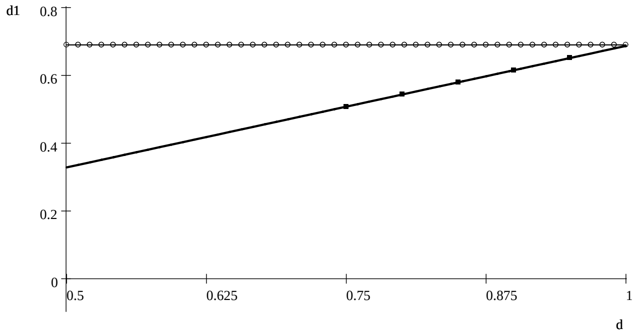

Figure 2. Plots of the points $(d, d_1^*(d))$, and the lines $d_1 = \widehat{d}_1^*(d)$ and $d_1 = d^* \equiv 0.69$.

in agreement with the previous results, $\widehat{d}_1^*(1)$ is very close to d_1^*, the discrepancy can be attributed to the numerical error in the approximation.

In particular, for the simple alternative case, we can use $d_1 = -0.031 + 0.719 d_A$, and for the more interesting composite alternative case, we can employ as d_1

$$\widehat{d}_1^*(\widetilde{d}_T) := -0.031 + 0.719 \widetilde{d}_T, \tag{7}$$

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

$$T^\tau \left(\widetilde{d}_T - d\right) = o_p(1) \text{ for some } \tau > 0, \tag{8}$$

and $|\widetilde{d}_T| \leq K$ for some $K < \infty$. Note that condition (8) holds for parametric estimators of d as required by DGM, and also for many semiparametric estimators. The use of consistent (at a power rate) estimators for d can be easily justified using similar techniques to Robinson and Hualde (2003, proposition 9). The previous condition (8) holds for many semiparametric estimators for an appropriate choice of the bandwidth parameter (see Robinson 1995a,b, Velasco 1999a,b).

In addition, note that contrary to DGM's discussion, trimming is not necessary since d_1 is allowed to be equal or larger than 1, although these values are not optimal for any case under the DGP of Theorem 1. The following lemma justifies this implementation of the FDF test. The proof is omitted since it is similar to the proof of theorem 5 in DGM.

Lemma 1. *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 $N(0, 1)$.*

3 OPTIMAL FDF TESTS: THE SERIAL CORRELATION CASE

The analysis in the previous section imposes that the DGP is $\Delta^d y_t = \varepsilon_t$, where ε_t is white noise. Practically, it is more appropriate to assume that ε_t is serially correlated, so that the DGP of y_t is given by

$$\alpha(L)\,\Delta^d y_t = \varepsilon_t 1_{\{t>0\}}, \tag{10}$$

where we assume that $\alpha(L) = 1 - \alpha_1 L - \cdots - \alpha_p L^p$ is a polynomial in the lag operator with all its roots outside the unit circle and ε_t has finite fourth moment. Note that in this situation, the FDF test is not valid because it cannot control the type I error. In order to control the type I error, DGM proposed the use of the AFDF test that is based on the t-statistic associated to the coefficient of the regressor $\Delta^{d_1} y_{t-1}$ in a regression of Δy_t on $\Delta^{d_1} y_{t-1}$ and p lags of Δy_t

$$\Delta y_t = \phi \Delta^{d_1} y_{t-1} + \alpha_1 \Delta y_{t-1} + \cdots + \alpha_p \Delta y_{t-p} + u_t, \quad t = 1, \ldots, T, \tag{11}$$

where regression (2) has been augmented by adding the lags of Δy_t .

DGM showed that the AFDF test can properly control the type I error when the DGP is an ARFIMA(p, 1, 0), p is known and p lags of Δy_t are included in the augmented regression. DGM also showed that the test is consistent. However, DGM did not consider the behaviour under local alternatives. The next theorem complements DGM by studying this case. Introduce the following notation. Define $\kappa = (\kappa_1, \ldots, \kappa_p)'$ with $\kappa_k = \sum_{j=k}^{\infty} j^{-1} c_{j-k},\ k = 1, \ldots, p$, where c_j denotes the coefficient of L^j in the expansion of $1/\alpha\ (L)$. Also, denote the Fisher information matrix for α under Gaussianity by $\Phi = \left[\Phi_{k,j}\right],\ \Phi_{k,j} = \sum_{t=0}^{\infty} c_t c_{t+|k-j|}$, for $k, j = 1, \ldots, p$. Finally, call $\kappa(d_1) = (\kappa_1(d_1),\ \kappa_2(d_1), \ldots,\ \kappa_p(d_1))$, where $\kappa_k(d_1) = \sum_{j=k}^{\infty} \pi_j (d_1 - 1)\, c_{j-k}$, for $k = 1, \ldots, p$.

Theorem 2. *Under the assumption that the DGP is an ARFIMA (p,d,0) model defined by (10), the asymptotic distribution of the t-ratio test statistic* $t(d_1)$*, for testing* $\phi = 0$ *in* (11)*, for any* $d_1 \geq 0.5$*, under local alternatives* ($d = 1 - \delta/\sqrt{T}$, $\delta \geq 0$)*, is given by*

$$t(d_1) \to_d N\left(-\delta\mu(d_1, \alpha, p), 1\right),$$

where

$$\mu(d_1, \alpha, p) = \frac{\left(\sum_{j=1}^{\infty} \pi_j (d_1 - 1)\, j^{-1} - \kappa'(d_1)\, \Phi^{-1} \kappa\right)}{\left(\sum_{j=0}^{\infty} \pi_j (d_1 - 1)^2 - \kappa'(d_1)\, \Phi^{-1} \kappa(d_1)\right)^{1/2}}.$$

The proof of this theorem is in the Appendix. The importance of this result resides in the fact that the non-centrality parameter $\mu(d_1, \alpha, p)$ depends on the serial correlation parameters α. Hence, a simple expression for the optimal d_1^* (the analogous of $d_1^* \simeq 0.69$) cannot be derived in this case, and optimal AFDF tests cannot be simply implemented. In order to recover the optimal selections obtained in Section 2, the natural approach would be to employ a pre-whitening procedure to attempt to get back to the white noise framework by filtering (pre-whitening) the serial correlation of the original series. The main idea is to perform the augmented regression (11) where both the dependent variable and the independent variable, whose significance is tested, have been pre-whitened.

Table 1. Optimal d_1^* for PAFDF tests.

p	0	1	2	3	4	5
d_1^*	0.691	0.846	0.901	0.927	0.942	0.951

Optimal d_1^* for ARFIMA(p, d, 0) for the PAFDF test.

The prewhitened augmented FDF (PAFDF) test consists of two steps:

(1) Fit an AR(p) to the differenced series Δy_t. Call $\widehat{\alpha}(L)$ to the estimated filter and denote the prewhitened series by $\widetilde{y}_t$, that is, $\widetilde{y}_t = \widehat{\alpha}(L) y_t$. Note that the prewhitened series $\widetilde{y}_t$ are, asymptotically, purely fractional under the null hypothesis and under local alternatives, since the short-term dependence has been eliminated.

(2) The PAFDF test is the t-test statistic associated to ϕ in

$$\Delta \widetilde{y}_t = \phi \Delta^{d_1} \widetilde{y}_{t-1} + \alpha_1 \Delta y_{t-1} + \cdots + \alpha_p \Delta y_{t-p} + u_t, \quad t = 1, \ldots, T. \tag{12}$$

Note that, in spite of employing the filtered series $\widetilde{y}_t$, the p lags of Δy_t still need to be included in order to control the size of the text, see Agiakloglou and Newbold (1994) and Breitung and Hassler (2002) for a similar approach in a related context.

It is straightforward to show that the asymptotic null distribution of the PAFDF test is the standard normal and that the test is consistent. The next theorem, that is stated without proof, establishes the asymptotic behaviour of the PAFDF test under local alternatives.

Theorem 3. *Under the assumption that the DGP is an ARFIMA (p,d,0) model defined by (10), with* $d = 1 - \delta/\sqrt{T}$, $\delta \geq 0$, *the asymptotic distribution of the t-ratio test statistic* $t(d_1)$ *for testing* $\phi = 0$ *in* (12), *is given by*

$$t(d_1) \to_d N(-\delta \mu(d_1, p), 1),$$

where

$$\mu(d_1, p) = \frac{\left(\sum_{j=p+1}^{\infty} \pi_j (d_1 - 1) j^{-1}\right)}{\left(\sum_{j=p}^{\infty} \pi_j (d_1 - 1)^2\right)^{1/2}}.$$

The drift expression, $\mu(d_1, p)$, is easily obtained using similar arguments as the corresponding drift expression in Theorem 2. It is simple to see that both drifts coincide when $\alpha(L) = 1$. The important point is that, contrary to the noncentrality parameter of Theorem 2, $\mu(d_1, p)$ depends only on p, and not on the serial correlation parameters α. In Table 1, we give the expressions for the optimal d_1^* for $p = 0, 1, 2, \ldots, 5$.

So far, we have studied the asymptotic local properties of the AFDF and PAFDF tests. For fixed alternatives, the main practical problem is that the analogous of d_1^* (d) does not exist. For instance, consider an ARFIMA (1,d,0) with autoregressive parameter denoted by α. For this case, we could employ the AFDF test with one lag of Δy_t. For this test, the optimal d_1 turns out to be $-0.027 + 0.86d$ when $\alpha = 0$, whereas it equals $0.020 + 0.59d$, when $\alpha = 0.6$. Hence, it is different for each value of α. For the PAFDF test, a similar situation occurs. Despite the fact that asymptotic local power of the PAFDF test does not depend on the serial correlation parameters, for fixed alternatives the optimal selection of d_1 as a function of d still depends on α .

Hence, both the AFDF and the PAFDF tests present the problem that their optimal implementations depend on the DGP for fixed alternatives. Therefore, optimal expressions of d_1 as a function of d are of limited practical interest. In order to overcome this problem, in the next section we introduce an algorithm that automatically selects the optimal d_1 for either the AFDF and the PAFDF tests. The only drawback of this algorithm is that it can be computationally involved because in order to estimate the relation between the optimal d_1 and d, the data need to be fractionally differenced a number of times and a corresponding optimization problem has to be solved. However, in Section 5 we will show by simulations that there are significant empirical power improvements associated to the use of this algorithm.

4. AUTOMATIC OPTIMAL IMPLEMENTATION OF THE AFDF AND PAFDF TESTS

In the previous section, we have seen that the optimal selections of d_1 as a function of d hardly have any practical use, since they depend on the serial correlation parameters for fixed alternatives, for both the original AFDF test and the PAFDF test. In practice, a natural selection for d_1 is

$$\overline{d}_1 = \arg\max_{d_1 \in D_1} t(d_1)^2,$$

where $D_1 = [\underline{d}, \overline{d}]$ is any closed interval that belongs to the interior of $D = [0.5, \infty)$. This choice for d_1 comes intuitively from the discussion in Section 2 where we saw that maximizing the power entailed maximizing the value of the squared t-statistic. This choice is behind the spirit of some empirical applications, such as Heravi and Patterson (2005), who report the AFDF test for a grid of values of d_1. However, this selection for d_1 should be carefully considered. Note that our previous theory, as well as DGM's, is only valid when d_1 is either a fixed value or a consistent estimator of a fixed value (see DGM's theorems 2 and 5). The problem is that under the null $\overline{d}_1$ does not converge to a fixed limit, but to a random variable. The underlying reason is that the optimal value of d_1 is not identified under the null hypothesis, as commented in the Introduction. Therefore, the established theory is not applicable for $t(\overline{d}_1)$ since critical values from the standard normal distribution cannot be employed. Hence, for practical purposes, tests based on $t(\overline{d}_1)$ are of limited interest.

An alternative to using $\overline{d}_1$ is to employ a consistent estimator of the optimal d_1. Note that, as commented above, this optimal d_1 depends on the DGP. The important point to realize is that, despite the optimal d_1 not being identified under the null hypothesis, it is identified (and hence it can be consistently estimated) under the alternative. However, in practice the researcher does not know whether the data have been generated under the alternative or under the null. Therefore, in order to obtain consistent estimators of the optimal d_1, we need to design a procedure that guarantees that the employed data are generated under the alternative and, at the same time, it does not alter the short term properties of the data. We propose to fractionally difference the given sequence y_t by an amount $\delta > 0$, so that we can be sure that the data behave under the alternative (except when $d = 1 + \delta$) without altering the short-term behaviour of the data. Given that the relation between the optimal d_1 and the true d is linear (this is true for the ARFIMA(p, d, 0) process, at least), by considering several values of δ, we can approximate with any degree of accuracy the relationship between any $d - \delta$ value and the corresponding optimal d_1. Finally, we can extrapolate this relation at $\delta = 0$ so that we consistently estimate the optimal d_1 for the data at hand. It is simple to show that, when the data are generated under the null, the projected value for d_1 converges to a fixed value (that would be d_1^* when there is no serial correlation), and, hence, employing critical values from the standard normal guarantees that the type I error is properly controlled. In addition, when the data are generated under the alternative, this procedure estimates consistently the optimal d_1 for the given data.

In practice, the procedure consists of the following steps.

(1) Select a set of $0 < \delta_1 < \delta_2 < \ < \delta_q$ with $q \geq 2$. Note that for ARFIMA(p, d, 0) processes $q = 2$ is enough given that the optimal relation is linear.

(2) For each δ_i, obtain

$$\overline{d}_1(\delta_i) = \arg\max_{d_1} t_\delta^*(d_1)^2,$$

where t_δ^* (d_1) denotes the t-statistic associated to the coefficient ϕ in the regression (11), where the input series y_t is replaced by the fractional differenced series $\Delta^\delta y_t$.

(3) Use the pairs $(\delta_i, \overline{d}_1(\delta_i))$ to fit a simple polynomial model (typically linear) by OLS or other estimation procedure and denote the fitted model by $\widetilde{d}_1^*(\delta)$.

(4) The proposed t-test uses the original y_t and employs

$$d_1 = \max(\widetilde{d}_1^*(0), 0.5) \tag{13}$$

in the augmented regression (11).

Note that the fitted $\widetilde{d}_1^*(0)$ converges to a constant both under the null (the optimal d_1 in the local alternative case) and under the alternative (the optimal d_1 in the fixed alternative case). However, this optimal d_1 could be below 0.5. In order to retain the asymptotic null standard distribution, we need to assure that the input d_1 is greater than 0.5 and this is the reason for the truncation in (13). Then, it can be justified, as in the white noise case, that the asymptotic null distribution of the t-statistic that employs (13) as d_1 is the standard normal. Note that an additional feature of the previous algorithm is that it avoids the pre-estimation of d, and, therefore, the introduction of the bandwidth parameter necessary for obtaining the semiparametric estimator of d.

The previous algorithm has been presented for the original AFDF test, but a similar algorithm can be employed to derive optimal implementations of the prewhitened AFDF. The difference is that in Step 2 the t-statistic is now associated to the coefficient ϕ in the regression (12). In the next section, we will study the finite sample properties of these tests.

5. SIMULATIONS

In this section, we comment on the results of a small Monte Carlo study. We consider two DGPs, a pure fractionally integrated Gaussian process and a Gaussian ARFIMA(1, d, 0). Tables 2 and 3 report the results for the first DGP for a nominal level of 0.05 and two samples sizes, 100 and 500, respectively. The number of replications is 50,000 for Table 2, and 10,000 for Table 3. The parameter d takes values from 0.5 to 1 with increments of 0.05 in Table 2, and it takes values from 0.8 to 1 with increments of 0.025 in Table 3. These tables report the results of the FDF test with five selections for d_1, namely $d_1 = d$, FDF(d), $d_1 = \widehat{d_1^*(d)} = -0.031 + 0.719d$, FDF($d_1^*$ (d)), $d_1 = d_1^* \simeq 0.69$, FDF(d_1^*), $d_1 = \widehat{d}_{SP}$, FDF($\widehat{d}_{SP}$) and $d_1 = d_1^*(\widehat{d}_{SP}) = -0.031 + 0.719\widehat{d}_{SP}$, FDF($d_1^*(\widehat{d}_{SP})$), where $\widehat{d}_{SP}$ denotes Robinson's (1995b) semiparametric estimator. Regarding the first and second selections of d_1, 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.

Table 2. Size and size-adjusted power. White noise case. $T = 100$.

Test \ d	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**	**0.95**	**1**
FDF(d)	100	99.9	99.6	98.2	93.6	82.1	64.2	43.0	24.5	11.9	5.27
FDF(d_1^* (d))	100	100	99.9	99.7	97.8	91.0	75.7	52.7	30.0	13.6	5.52
FDF(d_1^*)	100	99.7	99.2	97.4	93.3	84.3	69.0	48.9	28.6	13.4	5.50
FDF($\widehat{d}_{SP}$)	99.9	99.7	98.9	96.7	91.1	80.1	62.8	42.0	23.9	11.6	6.65
FDF(d_1^* ($\widehat{d}_{SP}$))	100	100	99.9	99.6	97.0	90.0	73.4	50.2	28.6	13.0	6.87
LM	99.9	99.7	99.1	97.1	92.1	81.5	64.6	44.6	25.8	12.6	4.53
W	100	99.9	99.5	97.8	92.7	81.5	63.7	43.0	24.7	11.8	10.82

Monte Carlo size ($d = 1$) and (size-adjusted) power ($d < 1$). Percentage of rejections based on 5% nominal level. Series follow a FI(d) with Gaussian errors. Sample size is 100. The number of replications is 50,000.

Table 3. Size and size-adjusted power. White noise case. $T = 500$.

Test \ d	**0.8**	**0.825**	**0.85**	**0.875**	**0.9**	**0.925**	**0.95**	**0.975**	**1**
FDF(d)	99.9	99.3	96.4	88.6	73.0	51.4	29.0	13.2	4.86
FDF(d_{1^*}(d))	100	99.9	99.4	96.8	86.3	64.7	37.3	15.7	5.14
FDF(d_{1^*})	100	99.8	99.1	95.2	84.3	63.3	36.7	15.6	5.13
FDF($\widehat{d}_{SP}$)	99.9	99.3	96.9	87.9	78.0	50.6	30.9	11.7	6.25
FDF(d_{1^*}($\widehat{d}_{SP}$))	100	99.9	98.8	95.0	84.5	62.2	35.5	14.3	5.60
LM	100	99.9	99.1	93.9	83.9	62.5	35.8	15.1	5.50
W	100	100	100	97.0	84.7	63.1	35.8	13.5	7.21

Monte Carlo size ($d = 1$) and (size-adjusted) power ($d < 1$). Percentage of rejections based on 5% nominal level. Series follow a FI(d) with Gaussian errors. Sample size is 500. The number of replications is 10,000.

In addition, we report two feasible parametric tests: the Lagrange Multiplier (LM) test considered by Robinson (1991, 1994) and Tanaka (1999) and the Wald (W) test derived from the results in Fox and Taqqu (1986). Note that the W test is applied to the differenced series. These tables report the size results ($d = 1$) and size-adjusted power instead of raw power ($d < 1$) since the W test presents severe size distortions.

These two tables indicate that the proposed optimal implementation of the FDF test in some cases can improve up to 30% the size-adjusted power with respect to DGM's original proposal, for both the unfeasible and the feasible versions of the test. Also note that the loss of empirical power of the FDF(d_1^*) and the LM tests is larger as the alternative is further from the null, reflecting the local character of these tests. The W test presents severe size distortions, especially for $n = 100$. From these tables, we can conclude that the advantage of the optimal implementation of the FDF test with respect to W is to better control the size, whereas compared to LM it offers a superior empirical power.

In Tables 4 and 5, we consider the case where the DGP is a Gaussian ARFIMA(1, d, 0) with autoregressive parameter $\alpha_1 = \{-0.5, 0, 0.3, 0.6, 0.8\}$. In these tables, we only report the results for one negative value for α_1 because for other negative values of α_1 the results were similar, contrary to the $\alpha_1 > 0$ case where finite sample power depends greatly on α_1. In addition, the most relevant empirical case is when $\alpha_1 > 0$. The parameter d takes values from 0.5 to 1 with increments of 0.05. As above, we use 0.05 as the nominal level, and consider same sample sizes

Table 4. Size and size-adjusted power. Serial correlation case. $T = 100$.

α_1	d	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**	**0.95**	**1**
	AFDF($\widehat{d}_{SP}$)	98.6	97.4	93.7	87.7	79.0	63.3	44.4	28.9	16.5	8.7	9.47
	Auto-AFDF	100	99.9	98.3	96.0	87.3	74.3	56.7	38.9	21.8	11.5	8.44
-0.5	PAFDF($\widehat{d}_{SP}$)	99.3	98.1	95.4	89.5	79.1	64.3	46.6	30.2	17.8	9.5	9.13
	Auto-PAFDF	99.9	99.4	98.6	96.2	88.5	75.2	56.9	39.9	23.0	11.9	8.07
	LM	99.3	98.1	95.2	89.5	79.8	66.1	49.4	33.2	19.9	10.5	5.09
	W	99.5	98.5	95.8	89.7	78.9	63.7	46.6	30.8	18.4	10.0	14.41
	AFDF($\widehat{d}_{SP}$)	97.3	93.9	87.5	76.7	62.6	46.7	31.8	20.2	12.1	7.1	9.81
	Auto-AFDF	98.5	95.3	89.9	79.6	66.3	52.2	36.6	24.5	14.8	8.1	9.55
0.0	PAFDF($\widehat{d}_{SP}$)	95.6	90.8	83.2	72.6	59.4	44.9	31.2	20.1	12.1	7.1	9.70
	Auto-PAFDF	98.6	95.3	89.9	81.2	67.4	53.8	38.2	26.1	15.9	9.2	8.78
	LM	93.1	87.9	80.1	69.7	57.7	44.8	32.5	22.2	14.1	8.6	4.06
	W	96.6	92.3	84.7	73.9	60.5	46.6	33.5	22.8	14.3	8.8	13.62
	AFDF($\widehat{d}_{SP}$)	83.9	73.7	62.1	50.3	37.5	26.4	18.6	12.0	8.6	6.0	9.94
	Auto-AFDF	91.2	84.6	75.0	63.4	49.3	37.0	26.6	18.6	12.6	8.5	9.65
0.3	PAFDF($\widehat{d}_{SP}$)	84.2	75.7	64.9	52.5	39.7	28.1	18.9	12.5	8.4	6.1	8.43
	Auto-PAFDF	92.7	85.9	77.2	65.3	52.2	39.2	28.0	18.7	11.9	8.3	7.26
	LM	79.5	71.5	61.8	51.7	41.5	31.9	23.4	16.7	11.5	7.7	2.04
	W	94.4	89.3	81.7	71.6	59.9	47.3	35.0	24.1	15.6	9.3	3.35
	AFDF($\widehat{d}_{SP}$)	40.0	30.9	24.0	18.6	13.7	11.0	8.6	7.0	6.0	5.4	8.70
	Auto-ADGM	49.5	39.2	30.8	24.3	17.9	13.8	10.4	8.4	6.9	5.5	10.12
0.6	PAFDF($\widehat{d}_{SP}$)	58.8	41.1	39.0	28.8	21.7	15.5	11.0	8.1	6.4	5.5	4.66
	Auto-PAFDF	59.7	49.6	40.2	32.1	24.8	18.7	15.0	12.1	8.9	6.3	7.25
	LM	43.4	36.1	29.2	23.2	18.0	13.9	10.5	8.2	6.4	5.5	1.11
	W	88.1	80.9	72.5	62.2	50.3	38.6	28.1	19.3	12.8	8.2	0.44
	ADGM($\widehat{d}_{SP}$)	13.1	11.8	11.3	10.8	10.6	10.3	9.6	8.9	7.7	6.4	4.68
	Auto-AFDF	14.3	10.8	8.0	5.7	4.9	4.3	4.1	4.1	4.3	4.4	9.52
0.8	PAFDF($\widehat{d}_{SP}$)	20.2	17.0	14.6	13.0	11.9	11.1	10.3	9.1	7.9	6.5	4.39
	Auto-PAFDF	24.2	20.3	18.3	16.9	14.7	13.8	12.8	10.7	8.7	6.6	5.90
	LM	2.9	2.3	2.1	2.4	3.1	4.4	5.8	6.9	7.2	6.4	3.75
	W	52.5	45.1	37.9	31.6	26.3	21.8	17.5	13.7	10.4	7.5	1.35

Monte Carlo size ($d = 1$) and (size-adjusted) power ($d < 1$). Percentage of rejections based on 5% nominal level. Series follow an ARFIMA(1,d,0) with Gaussian errors. The autoregressive parameter is α_1. The number of lags of Δy_t included in the augmented regression is 1. Sample size is 100. The number of replications is 50,000.

and number of replications. These tables report the results for six feasible tests. The first four are related to the AFDF test. In particular, they are: DGM's initial proposal of the augmented FDF test that uses $d_1 = \widehat{d}_{SP}$, AFDF($\widehat{d}_{SP}$), the automatic optimal implementation of the AFDF test discussed in Section 4, Auto-AFDF, the prewhitened AFDF that employs $d_1 = \widehat{d}_{SP}$, PAFDF($\widehat{d}_{SP}$), and the automatic optimal implementation of the PAFDF test, Auto-PAFDF. These tables also include the previously mentioned LM and W tests with $p = 1$.

Table 5. Size and size-adjusted power. Serial correlation case. $T = 500$.

α_1	d	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**	**0.95**	**1**
	AFDF($\widehat{d}_{SP}$)	100	100	100	99.9	99.8	99.1	96.3	87.9	58.9	21.0	8.48
	Auto-AFDF	100	100	100	99.7	98.5	96.7	93.7	92.5	73.3	29.5	7.67
-0.5	PAFDF($\widehat{d}_{SP}$)	100	100	100	100	100	99.9	98.5	88.7	58.0	22.5	7.17
	Auto-PAFDF	100	100	100	100	100	100	99.7	97.4	75.7	30.8	7.06
	LM	100	100	100	100	100	100	99.8	95.4	70.6	28.5	5.65
	W	100	100	100	100	100	100	99.9	96.0	70.6	28.2	7.91
	AFDF($\widehat{d}_{SP}$)	100	100	100	100	99.8	98.8	92.4	72.5	37.9	13.2	8.77
	Auto-AFDF	100	100	99.9	99.6	98.9	98.1	97.0	85.8	53.4	21.9	7.83
0.0	PAFDF($\widehat{d}_{SP}$)	100	100	100	100	99.8	98.1	91.2	70.0	38.7	14.1	7.98
	Auto-PAFDF	100	100	100	100	100	99.6	97.6	87.1	55.7	23.3	6.60
	LM	100	100	100	100	100	99.5	95.7	80.2	48.8	19.5	5.47
	W	100	100	100	100	100	99.7	96.7	80.9	47.9	18.4	7.81
	AFDF($\widehat{d}_{SP}$)	100	100	100	99.9	99.0	92.5	75.6	48.6	24.3	10.2	8.67
	Auto-AFDF	100	100	99.9	99.8	99.4	98.1	88.8	64.7	36.1	16.0	7.41
0.3	PAFDF($\widehat{d}_{SP}$)	100	100	100	99.7	98.5	91.4	69.9	39.9	17.6	7.6	7.84
	Auto-PAFDF	100	100	100	100	99.6	98.8	89.2	65.6	36.2	16.3	6.09
	LM	100	100	100	99.8	98.7	94.0	80.7	57.3	32.4	14.3	4.62
	W	100	100	100	100	99.3	94.6	79.1	51.7	26.4	11.4	8.36
	AFDF($\widehat{d}_{SP}$)	99.6	98.4	92.8	82.1	65.4	44.1	28.5	16.4	9.9	6.3	8.82
	Auto-AFDF	99.9	99.3	97.5	93.4	82.6	64.4	44.6	26.8	15.6	8.6	6.99
0.6	PAFDF($\widehat{d}_{SP}$)	99.9	98.9	95.1	85.4	69.2	49.4	31.2	18.6	10.4	6.9	5.40
	Auto-PAFDF	99.9	99.2	98.7	95.6	87.5	70.0	51.3	32.7	18.8	9.3	6.02
	LM	98.6	96.2	90.6	81.3	68.2	52.5	36.9	23.6	14.6	8.2	4.29
	W	100	99.7	98.1	92.3	79.2	59.2	38.3	23.3	13.6	8.3	6.43
	AFDF($\widehat{d}_{SP}$)	42.1	28.8	18.6	12.1	10.1	9.3	9.5	9.6	8.7	7.1	5.59
	Auto-AFDF	58.1	42.5	29.1	20.1	13.8	9.9	6.7	5.5	5.0	4.9	6.94
0.8	PAFDF($\widehat{d}_{SP}$)	80.4	70.8	60.9	51.6	43.1	35.4	28.3	22.1	15.6	9.8	4.47
	Auto-PAFDF	85.8	76.3	65.8	54.2	44.4	35.5	28.4	21.7	15.4	9.5	5.83
	LM	22.2	17.5	15.0	14.7	16.2	18.5	20.0	19.2	15.3	10.1	5.97
	W	95.2	90.9	84.6	76.5	67.0	55.5	43.5	30.9	20.2	11.4	4.60

Monte Carlo size ($d = 1$) and (size-adjusted) power ($d < 1$). Percentage of rejections based on 5% nominal level. Series follow an ARFIMA(1,d,0) with Gaussian errors. The autoregressive parameter is α_1. The number of lags of Δy_t included in the augmented regression is 1. Sample size is 500. The number of replications is 10,000.

Similarly to the white noise case, we report size-adjusted power. For all employed tests, we have correctly controlled for the serial correlation by including one lag in the augmented regression. Regarding the automatic tests note that, in order to stabilize the fit of $\widetilde{d}_1^*(\delta)$, we consider positive and negative values for δ. The reason for including negative values for δ is that the function $d_1^*(d)$ is smooth for $d \geq 1$. Note that under the alternative many of these fractionally differenced series with $\delta < 0$ will have memory less than 1. The only potential problem is that for some particular δ_j the fractionally differenced series $\Delta^{\delta_j} y_t$ has memory equal to 1, rendering inconsistent $\overline{d}_1(\delta_j)$. Note that this can happen for at most one δ_j (when $d = 1 + \delta_j$). In order to control the possible distortion caused by this single point, we can either use robust estimation procedures (such as LAD) instead of OLS or use a number of $\delta's$ that increase with the sample size. In these simulations, we have followed the first option and use LAD-linear regression. The chosen set of possible values for d_1 is [0.2, 1.2]. A Fortran code with the program is available from the authors. In these simulations, we have chosen $q = 8$ and $\delta \in \{0.5, 0.4, 0.3, 0.2, 0.1, -0.1, -0.2, -0.3\}$.

The main conclusions from Tables 4 and 5 are the following. Compared to the white noise case, there is a significant loss of power for any value of α_1. It is also especially notable that power is higher when the serial correlation is negative, and decreases fast for positive α_1. For the AFDF tests, the main effect of pre-whitening is to help controlling the type I error. It is also apparent that the automatic optimal implementations for both the AFDF and the PAFDF tests improve, sometimes substantially, the empirical power. For instance, compared to DGM's initial proposal, the automatic PAFDF test improves power by around 40% for moderate serial correlation ($\alpha = 0.3$) when d equals 0.85 or 0.9. Note also that the LM test presents low power for alternatives away from the null. This is especially notable for the strong serial correlation cases, $\alpha = 0.6, 0.8$. The W test cannot properly control the type I error for moderate sample sizes, similarly to the white noise case, hence, the size-corrected power figures should be interpreted with care. However, for large samples, it presents the highest empirical power, as could be anticipated, given its parametric nature. Similarly to the white noise case, these simulations indicate that the automatic optimal implementation of the PAFDF test presents some advantage over the existing simpler parametric tests, in terms of power with respect to the LM, and in terms of size with respect to the W.

6. CONCLUSIONS

This article has provided a new interpretation for the FDF test that has led to the development of more powerful tests. In particular, the input value d_1 that needs to be used in the FDF test is not interpreted as 'the true value of d under the alternative hypothesis' as DGM do, but as an auxiliary parameter that maximizes the power of the FDF test. Contrary to DGM's arbitrary selection of d_1, we have addressed the issue of optimally selecting the value of d_1 .

For the white noise case, 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 derived the optimal selection for d_1. 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 (2). In this framework, we have derived optimal FDF 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.

For the serial correlation case, it is also possible to establish optimal selections for d_1 as a function of the long memory parameter. However, these optimal expressions are of limited practical use since they depend on the short memory parameters. For practical purposes, we have proposed an automatic approach that optimizes the PAFDF test. The test procedure is based on an algorithm that fractionally difference the series by various degrees, so it avoids the use of these optimal rules and it also avoids the lack of identification of the auxiliary parameter under the null hypothesis. This algorithm automatically optimizes the AFDF and the PAFDF tests, and, as an additional feature, it avoids the introduction of the bandwidth parameter necessary for estimating semiparametrically d.

We finish with a suggestion for further research. In this article, we have considered the case where the number of lags in the augmented regression is correctly set by the researcher. An alternative, which should be appealing for applied researchers, consists of applying a data based information criterion to select automatically the lag length. However, note that a routine application of typical criteria, such as AIC, does not guarantee the delivery of tests with good power properties, since these criteria are typically designed for an estimation framework, rather than for a testing problem, (see e.g. Ng and Perron 2001). Another possibility is to consider testing the significance of the lags sequentially. A careful analysis of the asymptotic properties and the finite sample behaviour of these tests procedures when an automatic lag selection is employed merits further research.

ACKNOWLEDGMENTS

We thank the co-editor and two referees for useful comments. We also thank J. Arteche, M. Avarucci, M. Delgado, J. Dolado, L. Gil-Alaña, J. Gonzalo, J. Hidalgo, L. Mayoral, P. Perron, W. Ploberger and P. Robinson for useful conversations. Part of this research was carried out while Lobato was visiting Universidad Carlos III de Madrid thanks to the Spanish Secretaría de Estado de Universidades e Investigación, Ref. no. SAB2004-0034. 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 Ministerio de Educación y Ciencia, Ref. no. SEJ2004-04583/ECON.

REFERENCES

Agiakloglou, Ch. and P. Newbold (1994). Lagrange multiplier tests for fractional difference, *Journal of Time Series Analysis 15*, 253–62.

Breitung, J. and U. Hassler (2002). Inference on the cointegration rank in fractionally integrated processes, *Journal of Econometrics 110*, 167–85.

Delgado, M. A. and C. Velasco (2005). Sign tests for long memory time series, *Journal of Econometrics 128*, 215–51.

Dolado, J. J., J. Gonzalo and L. Mayoral (2002). A fractional Dickey–Fuller test for unit roots, *Econometrica 70*, 1963–2006.

Fox, R. and Taqqu, M. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Annals of Statistics 14*, 517–32.

Heravi, S. and K. Patterson (2005). Optimal and adaptive semi-parametric narrowband and broadband and maximum likelihood estimation of the long-memory parameter for real exchange rates, *The Manchester School 73*, 165–213.

Ng, S. and P. Perron (2001). Lag length selection and the construction of unit root tests with good size and power, *Econometrica 69*, 1519–54.

Robinson, P.M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression, *Journal of Econometrics 47*, 67–84.

Robinson, P. M. (1994). Efficient tests of nonstationary hypotheses, *Journal of the American Statistical Association 89*, 1420–37.

Robinson, P. M. (1995a). Log-periodogram regression of time series with long range dependence, *Annals of Statistics*, *23*, 1048–72.
Robinson, P. M. (1995b). Gaussian semiparametric estimation of long range dependence, *Annals of Statistics 23*, 1630–61.
Robinson, P. M. and J. Hualde (2003). Cointegration in fractional systems with unknown integration orders, *Econometrica 71*, 1727–66.
Tanaka, K. (1999). The Nonstationary Fractional Unit Root, *Econometric Theory 15*, 249–64.
Velasco, C. (1999a). Non-stationary log-periodogram regression, *Journal of Econometrics 91*, 325–71.
Velasco, C. (1999b). Gaussian semiparametric estimation of non-stationary time series, *Journal of Time Series Analysis 20*, 87–127.

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 $T^{-1/2}$,

$$\begin{aligned} Q_T(d_1) &= T^{-1/2}\sum_{t=2}^{T}\Delta y_t \Delta^{d_1} y_{t-1} \\ &= T^{-1/2}\sum_{t=2}^{T}\left(\varepsilon_t + \sum_{i=1}^{t-1}\left(\frac{1}{i}\frac{-\delta}{\sqrt{T}}\right)\varepsilon_{t-i}\right)\left(\varepsilon_{t-1} + \sum_{i=1}^{t-2}\pi_i(-\eta_T)\varepsilon_{t-1-i}\right) \qquad \text{(A1)} \end{aligned}$$

$$+\, T^{-1/2}\frac{\delta^2}{2T}\sum_{t=2}^{T}\left(\sum_{i=1}^{t-1}\pi_i'(-\theta^*)\varepsilon_{t-i}\right)\left(\varepsilon_{t-1} + \sum_{i=1}^{t-2}\pi_i(-\eta_T)\varepsilon_{t-1-i}\right) \qquad \text{(A2)}$$

where π_i' is the first derivative of π_i and θ^* is some point between 0 and θ_T. Note that $\left|\pi_i'(-\theta^*)\right| \leq C\, i^{-1}\log i$ by lemma 1 of Delgado and Velasco (2005). Since (14) is $O_p(1)$ as it is showen next, it is straightforward to show that (15) is $o_p(1)$.

The leading term (14) of $Q_T(d_1)$ can be written as

$$= -\delta T^{-1} \sum_{t=2}^{T} \left(\varepsilon_{t-1}^2 + \sum_{i=1}^{t-2} \frac{\pi_i(-\eta_T)\varepsilon_{t-i-1}^2}{(i+1)} \right) \tag{A3}$$

$$+T^{-1/2} \sum_{t=2}^{T} \varepsilon_t \left(\varepsilon_{t-1} + \sum_{i=1}^{t-2} \pi_i(-\eta_T)\varepsilon_{t-1-i} \right) \tag{A4}$$

$$-\delta T^{-1} \sum_{t=2}^{T} \varepsilon_{t-1} \left(\sum_{i=1}^{t-2} \pi_i(-\eta_T)\varepsilon_{t-1-i} \right) \tag{A5}$$

$$-\delta T^{-1} \sum_{t=2}^{T} \sum_{i=1}^{t-2} \frac{1}{(i+1)} \varepsilon_{t-i-1} \left(\sum_{j=1, j \neq i}^{t-2} \pi_j(-\eta_T)\varepsilon_{t-1-j} \right). \tag{A6}$$

The last two terms, (18) and (19), in the previous expression are $o_p(1)$ using similar reasoning to that in the proof of theorem 4 in DGM. The term (16) 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 for martingale difference sequences, the term (17) converges in distribution to a $N(0, V)$ 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 $T^{-1}\sum_{t=2}^{T}(\Delta y_t - \widehat{\phi}\Delta^{d_1} y_{t-1}) \to_p 1$, and, given the above expression for $\Delta^{d_1} y_{t-1}$, by a law of large numbers it is easy to see that the $p \lim$ of $T^{-1}\sum_{t=2}^{T}(\Delta^{d_1} y_{t-1})^2$ is given by

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

Direct calculations lead to $K(d_1) = 1/d_1$ and to

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

Hence, using

$$\lim_{T\to\infty} \sum_{i=0}^{\infty} \pi_i(-\eta_T)^2 = \sum_{i=0}^{\infty} \pi_i(d_1 - 1)^2$$

we derive that

$$t(d_1) \to_d N\left(-\delta h(d_1), 1\right).$$

Proof of Theorem 2. In order to derive the noncentrality parameter of the asymptotic distribution of $t(d_1)$, the key idea is to use the basic equation of multivariate regression

$$t(d_1) = \sqrt{T}\frac{R_T}{\sqrt{1 - R_T^2}}, \tag{A7}$$

where R_T denotes the sample partial correlation coefficient between $Y_t := \alpha\,(L)\Delta y_t$ and $X_t := \Delta^{d_1} y_{t-1}$ given the p lags of Δy_t, $Z_t := (Z_{t,1}, \ldots, Z_{t,p})'$ with $Z_{t,k} = \Delta y_{t-k}$, $k = 1, \ldots, p$. Note that the denominator in (20) tends to 1 in probability under local alternatives for which the DGP is given by

$$\Delta y_t = \alpha(L)^{-1}\Delta^{\delta/\sqrt{T}}\varepsilon_t 1_{\{t>0\}},$$

and where the operator $\Delta^{\delta/\sqrt{T}}$ can be written as

$$\Delta^{\delta/\sqrt{T}} = 1 - \frac{\delta}{\sqrt{T}} J\,(L) + \frac{1}{T} H_T\,(L),$$

where $J(L) = \sum_{j=1}^{\infty} j^{-1}L^j$ and $H_T(L) = \sum_{j=1}^{\infty} h_{T,j}L^j$, with $|h_{T,j}| \le Cj^{-1}\log^2 j$, $j \ge 1$, uniformly in T. Then, we can write the series involved in $t(d_1)$ in terms of the i.i.d. variables ε_t, as follows: $Y_t = \Delta^{\delta/\sqrt{T}}\varepsilon_t$, $X_t = \left[\alpha\,(L)\,\Delta^{d_1-1}L\right]\Delta y_t = \Delta^{d_1-1}\Delta^{\delta/\sqrt{T}}L\varepsilon_t$ and $Z_{t,k} = \alpha(L)^{-1}\Delta^{\delta/\sqrt{T}}L^k\varepsilon_t$, $k = 1, \ldots, p$.

Next, we obtain the residuals Y_t^* and X_t^* of projecting Y_t and X_t, respectively, on the vector Z_t. It is simple to show that $Y_t^* = \Delta^{\delta/\sqrt{T}}\varepsilon_t$, plus a term due to the estimation of the projection on Z_t that contributes to the drift of $t(d_1)$ at a smaller order of magnitude because it is orthogonal to the residuals X_t^*. In order to study X_t^*, note that

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} X_t Z_{t,k} &= E\left[\Delta^{d_1-1}L\varepsilon_t \cdot \alpha(L)^{-1}\varepsilon_{t-k}\right] \\
&= \sum_{j=k}^{\infty} \pi_j\,(d_1 - 1)\,c_{j-k} = \kappa_k\,(d_1), \quad k = 1, \ldots, p,
\end{aligned}$$

whereas

$$\begin{aligned}\operatorname*{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}Z_{t,k}Z_{t,j} &= E\left[\alpha(L)^{-1}\varepsilon_{t-k}\cdot\alpha(L)^{-1}\varepsilon_{t-j}\right]\\ &= \sum_{t=0}^{\infty}c_t c_{t+|k-j|}=\Phi_{k,j},\quad k,j=1,\ldots,p.\end{aligned}$$

Then, the (population) least-squares projection coefficients of X_t onto Z_t are given by $\Phi^{-1}\ \kappa$, and, therefore, $X_t^* = \Delta^{d_1-1}L\varepsilon_t - \kappa'(d_1)\,\Phi^{-1}\alpha(L)^{-1}\varepsilon_{t,p}$, where $\varepsilon_{t,p}=\left(\varepsilon_{t-1},\ldots,\varepsilon_{t-p}\right)'$, plus smaller order terms. Next, we have that $T^{1/2}\sum_{t=1}^{T}Y_t^* X_t^*$ converges in distribution to a normal variate with mean equal to

$$\begin{aligned}&\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E\left[-\delta J(L)\,\varepsilon_t\cdot\left\{\Delta^{d_1-1}L\varepsilon_t-\kappa'(d_1)\,\Phi^{-1}\alpha(L)^{-1}\varepsilon_{t,p}\right\}\right]\\ &= -\delta\left(\sum_{j=1}^{\infty}\pi_j(d_1-1)\,j^{-1}-\kappa'(d_1)\,\Phi^{-1}\kappa\right),\end{aligned}$$

and variance $\sum_{j=0}^{\infty}\pi_j(d_1-1)^2-\kappa'(d_1)\Phi^{-1}\kappa(d_1)$. Note also that $\operatorname{plim}_{T\to\infty}T^{-1}\sum_{t=1}^{T}(Y_t^*)^2 = \operatorname{Var}[\varepsilon_t]=1$. Therefore, $\operatorname{plim}_{T\to\infty}T^{-1}\sum_{t=1}^{T}(X_t^*)^2$ is given by

$$\begin{aligned}&Var\left(\Delta^{d_1-1}L\varepsilon_t-\kappa'(d_1)\,\Phi^{-1}\alpha(L)^{-1}\varepsilon_{t,p}\right)\\ &= Var(\Delta^{d_1-1}L\varepsilon_t)+Var(\kappa'(d_1)\Phi^{-1}\alpha(L)^{-1}\varepsilon_{t,p})-2Cov(\Delta^{d_1-1}L\varepsilon_t,\kappa'(d_1)\Phi^{-1}\alpha(L)^{-1}\varepsilon_{t,p})\\ &= \sum_{j=0}^{\infty}\pi_j(d_1-1)^2+\kappa'(d_1)\,\Phi^{-1}\kappa(d_1)-2\kappa'(d_1)\,\Phi^{-1}\kappa(d_1)\\ &= \sum_{j=0}^{\infty}\pi_j(d_1-1)^2-\kappa'(d_1)\,\Phi^{-1}\kappa(d_1),\end{aligned}$$

and the theorem follows.