UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 05-69
Economics Series 35
November 2005

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624 98 75

EFFICIENT WALD TESTS FOR FRACTIONAL UNIT ROOTS *

Ignacio N. Lobato [1] y Carlos Velasco [2]

Abstract

In this article we introduce efficient Wald tests for testing the null hypothesis of unit root against the alternative of fractional unit root. In a local alternative framework, the proposed tests are locally asymptotically equivalent to the optimal Robinson (1991, 1994a) Lagrange Multiplier tests. Our results contrast with the tests for fractional unit roots introduced by Dolado, Gonzalo and Mayoral (2002) which are inefficient. In the presence of short range serial correlation, we propose a simple and efficient two-step test that avoids the estimation of a nonlinear regression model. In addition, the first order asymptotic properties of the proposed tests are not affected by the pre-estimation of short or long memory parameters

* **Acknowledgements:** We thank the co-editor and two referees for very useful comments, 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, and seminar participants at the London School of Economics and at the 2005 Econometric Society World Congress. 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.

[1] 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
[2] Departamento de Economía .Universidad Carlos III de Madrid. Calle Madrid 126, 28903 Getafe (Madrid), Spain. e-mail: carlos.velasco@uc3m.es

EFFICIENT WALD TESTS FOR FRACTIONAL UNIT ROOTS*

Ignacio N. Lobato and Carlos Velasco
Centro de Investigación Económica
Instituto Tecnológico Autónomo de México (ITAM)
Av. Camino Sta. Teresa 930, México D.F. 10700, Mexico
e-mail: ilobato@itam.mx
Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid 126, 28903 Getafe (Madrid), Spain
e-mail: carlos.velasco@uc3m.es

Abstract

In this article we introduce efficient Wald tests for testing the null hypothesis of unit root against the alternative of fractional unit root. In a local alternative framework, the proposed tests are locally asymptotically equivalent to the optimal Robinson (1991, 1994a) Lagrange Multiplier tests. Our results contrast with the tests for fractional unit roots introduced by Dolado, Gonzalo and Mayoral (2002) which are inefficient. In the presence of short range serial correlation, we propose a simple and efficient two-step test that avoids the estimation of a nonlinear regression model. In addition, the first order asymptotic properties of the proposed tests are not affected by the pre-estimation of short or long memory parameters.

*We thank the co-editor and two referees for very useful comments, 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, and seminar participants at the London School of Economics and at the 2005 Econometric Society World Congress. 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.

1 Introduction

Testing for nonstationarity of a time series is routinely performed as a first step in econometric modeling. For instance, in the traditional I(0)/I(1) framework, unit-root tests have been applied frequently. Recently, there has been considerable interest in studying long memory series where the degree of nonstationarity is characterized by a fractional integration parameter that takes values in a continuum. Analysis with long memory series has posed new problems and led to the development of new asymptotic and optimality theory. For instance, Robinson (1991, 1994a) have proposed Lagrange Multiplier (LM) tests both in the frequency and time domain, and Dolado, Gonzalo and Mayoral (2002, hereinafter DGM) have introduced a test based on an auxiliary regression for the null of unit root against the alternative of fractional integration.

In the basic framework y_t denotes a fractionally integrated process whose true order of integration is d, denoted as $I(d)$,

$$\Delta^d y_t 1\{t > 0\} = \varepsilon_t, \quad t = 1, 2, \ldots, \tag{1}$$

where ε_t are independent and identically distributed (i.i.d.) random variables with zero mean and finite variance, and $1\{\cdot\}$ denotes the indicator function. The fractional difference operator $\Delta^d = (1-L)^d$ is defined in terms of the lag operator L by the formal expansion,

$$\Delta^\alpha := \sum_{i=0}^{\infty} \pi_i(\alpha) L^i,$$

for any real α, where for $\alpha \neq 1, 2, \ldots,$

$$\pi_i(\alpha) = \frac{\Gamma(i-\alpha)}{\Gamma(i+1)\Gamma(-\alpha)},$$

and Γ is the Gamma function, with $\Gamma(0)/\Gamma(0) = 1$, so the first coefficients are $\pi_0(\alpha) = 1$ and $\pi_1(\alpha) = -\alpha$. From now on, in the notation we will suppress the truncation in (1) for nonpositive t, assuming implicitly that $y_t = \varepsilon_t = 0$, $t \leq 0$.

We consider testing the null hypothesis

$$H_0 : d = 1,$$

versus either a simple alternative

$$H_A : d = d_A < 1,$$

or a composite alternative

$$H_1 : d < 1.$$

DGM proposed to test the null hypothesis by means of the t-statistic of the coefficient of $\Delta^{d_1} y_{t-1}$ in the ordinary least squares (OLS) regression

$$\Delta y_t = \phi_1 \Delta^{d_1} y_{t-1} + u_t, \qquad t = 1, \ldots, T, \tag{2}$$

where T denotes the sample size. DGM called this t-ratio the fractional Dickey-Fuller test, based on a particular analogy with the Dickey and Fuller's (1979) test that led them to interpret d_1 as "the true value of d under the alternative hypothesis", and hence, they propose to use $d_1 = d_A$ when testing against H_A, and a consistent estimator of d when testing against H_1.

Notice that, in model (2), the null and alternative hypotheses can be expressed in terms of ϕ_1, defined as the probability limit of the OLS coefficient of $\Delta^{d_1} y_{t-1}$. Under H_0, $\phi_1 = 0$ because Δy_t is white noise, and hence, it is uncorrelated with $\Delta^{d_1} y_{t-1}$ for any value of d_1. By contrast, under the alternative, using that $\Delta y_t = \Delta^{1-d} \varepsilon_t = \varepsilon_t + (d-1)\,\varepsilon_{t-1} + \cdots$ and that $\Delta^{d_1} y_{t-1} = \Delta^{d} y_{t-1} = \varepsilon_{t-1}$ when $d_1 = d$ is employed, it is simple to show that $\phi_1 = d - 1 < 0$. Since ϕ_1 is also negative for any $d_1 > 0.5$, the regression model (2) can be used for testing the null hypothesis by checking the significance of the regressor $\Delta^{d_1} y_{t-1}$ with a one sided t-ratio test.

However, note that the null hypothesis could also be tested by testing the significance of alternative regressors. In fact, given that Δy_t is i.i.d. under the null, $\Delta^{d_1} y_{t-1}$ could be replaced in (2) by any function of the past, and the associated coefficient would still be zero; whereas, under the alternative, this coefficient would be negative for any function of the past with negative covariance with Δy_t.

This article questions the use of the regressor $\Delta^{d_1} y_{t-1}$ proposed by DGM, and examines carefully the optimal selection of the regressor in a regression model like (2) to conduct inference on the degree of integration of y_t. We argue that $\Delta^{d_1} y_{t-1}$ is not the best class of regressors one can choose. In order to grasp the intuition behind it, consider all the regressors which lead to a test statistic whose asymptotic null distribution is the standard normal (for instance, $\Delta^{d_1} y_{t-1}$, with $d_1 > 0.5$). Note that the test that maximizes the power among this group is the one that maximizes the correlation between the regressand and the regressor, and thus, it is based on a regression model where the errors are serially uncorrelated and uncorrelated to the regressor. Therefore, a regressor such as $\Delta^{d_1} y_{t-1}$ can not be optimal because, under the alternative hypothesis, there does not exist any values of ϕ_1 and d_1 that guarantee that the error term u_t in model (2) is serially uncorrelated and orthogonal with the regressor $\Delta^{d_1} y_{t-1}$. In this sense, model (2) is misspecified because it does not include the data generating process (DGP) defined by (1) as a particular case under the alternative hypothesis. In particular, the errors of the model, u_t, are different from the innovations of the process, ε_t, defined in (1). This misspecification implies that OLS estimation and the resulting t-test based on regression (2) are inefficient, even when d_1 is optimally chosen.

In this article we propose the use of an alternative regression model based on (1), which leads to an efficient t-test that can also be interpreted as a Wald test since the relevant slope coefficient in the estimated regression is linearly related to the parameter of interest. The proposed Wald test is asymptotically efficient against local alternatives since it is asymptotically equivalent to Robinson's (1991, 1994a) LM test, which is optimal in a Gaussian framework. In particular, we show that our t-test statistic is locally asymptotically equivalent to Robinson's time-domain LM test statistic

$$LM = T^{1/2}\left(\frac{\pi^2}{6}\right)^{-1/2}\sum_{j=1}^{T-1} j^{-1}\widehat{\rho}_{\Delta y}(j), \tag{3}$$

where $\widehat{\rho}_{\Delta y}(j)$ denotes the sample autocorrelation of order j of Δy_t. This statistic has also been attributed to Tanaka (1999), but note that it already appears in Robinson (1991), see also Robinson (1994b).

The plan of the article is the following. Section 2 proposes and analyzes the new efficient fractional regression test. Section 3 studies the consequences of allowing for serial correlation in ε_t in (1) and proposes a simple and efficient two-step test. Section 4 reports a Monte Carlo exercise on the finite sample performance of the considered tests. Section 5 concludes and proposes some lines of further research.

2 An optimal Wald test

In this section we study carefully the optimal selection of the regressor and develop an efficient Wald type test. In order to motivate the selection of the proposed regressor, note that for any d we can rewrite the DGP (1) as

$$\Delta y_t = \left(\Delta - \Delta^d\right) y_t + \varepsilon_t = \left(1 - \Delta^{d-1}\right)\Delta y_t + \varepsilon_t, \tag{4}$$

where the error term ε_t is truly i.i.d. under (1), both under the null and under the alternative hypotheses, and where the variable $\left(1 - \Delta^{d-1}\right)\Delta y_t$ does not contain Δy_t because

$$\left(1 - \Delta^{d-1}\right)\Delta y_t = (d-1)\Delta y_{t-1} + \sum_{j=2}^{t-1}\pi_j(d-1)\Delta y_{t-j}. \tag{5}$$

Equation (4) can also be written as

$$\Delta y_t = \varphi_2\left(\Delta^{d-1} - 1\right)\Delta y_t + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{6}$$

where $\varphi_2 = 0$ under the null and $\varphi_2 = -1$ under the alternative. Equation (6) suggests the use of the regressor $\left(\Delta^{d_2-1} - 1\right)\Delta y_t$ where d_2 denotes the input of the new test to distinguish it from the input d_1 in DGM's test. However, note that d_2 can not take the value $d_2 = 1$, which would make the regressor equal to zero, as (5) indicates. In addition, note that a one-sided t-ratio test statistic for the significance of φ_2 in (6) should not be used with values of $d_2 > 1$ because the sign of the coefficients of $\left(\Delta^{d_2-1} - 1\right)\Delta y_t$ in the expansion (5) changes depending on whether $d_2 - 1$ is positive or negative. Therefore, in order to make the regressor continuous at $d_2 = 1$, instead of (6) we propose to employ the following rescaled regression model

$$\Delta y_t = \phi_2 z_{t-1}(d_2) + u_t, \qquad t = 1, \ldots, T, \tag{7}$$

where

$$z_{t-1}(d_2) = \frac{\left(\Delta^{d_2-1} - 1\right)}{1 - d_2}\Delta y_t. \tag{8}$$

We propose to test the null hypothesis by testing the significance of the coefficient of $z_{t-1}(d_2)$, with $d_2 > 0.5$, in (7) by means of a left-sided test based on the t-ratio test statistic, denoted by t_ϕ.

Note that, when $d_2 = d$ in (7), the true value of ϕ_2 is obtained immediately by $\phi_2 = \varphi_2(1 - d) = d - 1$, which maps the hypotheses on the parameter d continuously into ϕ_2. That is, under the null, $\phi_2 = \varphi_2 = 0$, and, under the alternative, ϕ_2 takes negative values, the larger in absolute value the further d is from the null. Note the analogy with the original Dickey-Fuller test based on model $\Delta y_t = \phi y_{t-1} + u_t$, where $\phi = \rho - 1$ and ρ denotes the first order autocorrelation. In this case $\phi = 0$ (or $\rho = 1$) is the null and $\phi < 0$ (or $\rho < 1$) is the alternative. Both tests are Wald because of the relation between the slope coefficient in the auxiliary regression and the parameter of interest.

The model (7) is obviously related to the DGP (4) as we analyze next. Under the null hypothesis, Δy_t is i.i.d. and so, $\phi_2 = 0$ for any value of d_2, and model (7) is properly specified, with $u_t = \varepsilon_t$. Under the alternative hypothesis, when d_2 is chosen equal to d, $\phi_2 = d - 1$ and model (7) is again properly specified, with $u_t = \varepsilon_t$. However, when d_2 is chosen differently from the true value of d, this property is lost because the errors u_t are not i.i.d. and, in consequence, ϕ_2 (defined as the probability limit of the OLS estimator of the coefficient of $z_{t-1}(d_2)$) is no longer $d - 1$. Therefore, under the alternative, in order to maximize the correlation between the regressand and the regressor (and hence, to maximize the power of the corresponding t-test), the researcher should set $d_2 = d$. Other selections of d_2 would render consistent but inefficient tests compared to the selection $d_2 = d$.

Comparing models (2) and (7), we see that the only difference with DGM's test is the use of the regressor $z_{t-1}(d_2)$ instead of the regressor $\Delta^{d_1} y_{t-1}$. Both regressors can be expressed as a linear combination of past values of y_t, and if we denote by c_j^z and c_j^o the coefficients of y_{t-j} for $z_{t-1}(d)$ and $\Delta^d y_{t-1}$, respectively, it is simple to see that $c_j^z = (d-1)c_{j+1}^o$ for $j \geq 2$. However, the use of regressor $z_{t-1}(d_2)$ instead of $\Delta^{d_1} y_{t-1}$ leads to an important difference.

Whereas for model (7) there exist a value of the pair (d_2, ϕ_2), namely $(d_2, \phi_2) = (d, d-1)$, that leads to errors which are i.i.d. and independent of the regressor under the alternative hypothesis, for model (2) there does not exist any value of the pair (d_1, ϕ_1) with that property. Therefore, the t-test based on the OLS estimation of (2) is inefficient (for any selection of d_1) compared to the t-test based on the OLS estimation of (7) that uses $d_2 = d$. The intuition behind this inefficiency is straightforward: the regressor $z_{t-1}(d)$ contains all relevant past information to forecast Δy_t, whereas $\Delta^{d_1} y_{t-1}$ does not, irrespective of the value of d_1.

In addition, note that in the $d_2 = 1$ case, the indetermination 0/0 in (8) is solved using L'Hôpital rule since, as $d_2 \to 1$, the ratio $\left(\Delta^{d_2-1} - 1\right)/(1-d_2)$ tends to the derivative of the fractional filter $(1-L)^{-\delta}$ evaluated at $\delta = 0$, that is, to the linear filter $J(L) = -\log(1-L) = \sum_{j=1}^{\infty} j^{-1} L^j$. In this case the regression (7) can be rewritten as

$$\Delta y_t = \phi_2 \sum_{j=1}^{t-1} j^{-1} \Delta y_{t-j} + u_t, \qquad t = 1, \ldots, T. \tag{9}$$

Interestingly, the t-test for the significance of ϕ_2 in (9) is Robinson's LM test statistic given in (3), apart from a different, but asymptotically equivalent (under local alternatives) normalization. In order to see that, note that the sample covariance between the dependent and independent variable in (9) is given by $\sum_{j=1}^{T-1} j^{-1} \widehat{\gamma}_{\Delta y}(j)$, where $\widehat{\gamma}_{\Delta y}(j)$ denotes the sample autocovariance of order j of Δy_t. The t-test for the significance of ϕ_2 in (9) has been considered by Agiakloglou and Newbold (1994) and Breitung and Hassler (2002). Although the t-test based on (9) is asymptotically locally equivalent to the t-test based on (7), in a fixed alternative framework the t-test based on (7) should be preferred to one based on (9). The reason is that there does not exist any value for ϕ_2 that makes u_t in (9) to be both i.i.d. and independent of the regressor for fixed alternatives, and hence, the regressor $\sum_{j=1}^{t-1} j^{-1} \Delta y_{t-j}$ does not maximize the correlation with the regressand Δy_t.

The next theorem establishes the asymptotic properties of t_ϕ where d_2 is allowed to be stochastic with limit not necessarily equal to d. In particular, under local alternatives it shows that the test is asymptotically equivalent to the optimal Robinson's LM test when d_2 is optimally chosen. The proof is in Appendix 1. Introduce the function h,

$$h(d_2) = \frac{\sum_{j=1}^{\infty} j^{-1} \pi_j (d_2 - 1)}{\sqrt{\sum_{i=1}^{\infty} \pi_i (d_2 - 1)^2}}, \quad d_2 > 0.5, \quad d_2 \neq 1,$$

and $h(1) = \sqrt{\sum_{j=1}^{\infty} j^{-2}} = \sqrt{\pi^2/6}$.

Theorem 1. *Under the assumption that the DGP is given by*

$$\Delta^d y_t 1\{t > 0\} = \varepsilon_t,$$

where ε_t is i.i.d. with finite fourth moment, the asymptotic properties of the t-test statistic t_ϕ for testing $\phi_2 = 0$ in (7), where the input $\widehat{d_2}$ of z_{t-1} satisfies

$$\widehat{d_2} = d_2 + o_p\left(T^{-\tau}\right), \text{ with } \tau > 0, \text{ and } \widehat{d_2} > 0.5, \tag{10}$$

for some fixed $d_2 > 0.5$, are given by:

a) Under the null ($d = 1$),

$$t_\phi \rightarrow_d N(0, 1).$$

b) Under fixed alternatives ($d < 1$), the test based on t_ϕ is consistent.

c) Under local alternatives ($d = 1 - \delta/\sqrt{T}$, $\delta > 0$),

$$t_\phi \rightarrow_d N(-\delta h(d_2), 1).$$

Remark 1.1. The drift function h is plotted in Figure 1. Note that h achieves an absolute maximum at $d_2 = 1$, and that $h(1)$ equals the noncentrality parameter of the locally optimal Robinson's LM test, so the new test is locally asymptotically equivalent to this test when a consistent estimator of d, which satisfies condition (10), is employed as the input d_2. Also note that the drift of DGM's test statistic is 1, so the asymptotic relative efficiency of DGM test is 0.79.

Remark 1.2. Notice that the first part of condition (10) holds with $d_2 = d$ for any estimator of d that is consistent at a power rate, so that not only parametric $\sqrt{T}$-consistent estimators of d as proposed by DGM (e.g. Velasco and Robinson, 2000) are allowed but also many semiparametric estimators for an appropriate choice of the bandwidth parameter can be employed, such as those of Velasco (1999a, b). The condition $\widehat{d_2} > 0.5$ can be imposed naturally for implicitly defined memory estimators, such as the Gaussian semiparametric procedure of Robinson (1995), whereas for other estimators this condition could be replaced by the condition $|\widehat{d_2}| \leq K$, for some $K > 0$, as in Robinson and Hualde's (2003) Assumption 3. The purpose of these conditions is to guarantee that the use of estimated regressors does not alter the asymptotic distribution of the test statistic, given that $\phi_2 = 0$ under the null, see, for instance, the discussion in Wooldridge (2002, Chapter 6).

3 Short run dynamics

The analysis in the previous sections imposes that the DGP is $\Delta^d y_t = \varepsilon_t$, where ε_t is white noise. Practically, it is more appropriate to allow for $\Delta^d y_t$ to be serially correlated. In this section we consider that the DGP of y_t is given by the ARFIMA$(p, d, 0)$ model

$$\alpha(L)\,\Delta^d y_t 1\{t > 0\} = \varepsilon_t, \quad t = 1, 2, \ldots, \tag{11}$$

where $\alpha(L) = 1 - \alpha_1 L - \cdots - \alpha_p L^p$ is a polynomial in the lag operator with all its roots outside the unit circle. Note that this DGP can be written as

$$\alpha(L)\Delta y_t = \alpha(L)\left(1 - \Delta^{d-1}\right)\Delta y_t + \varepsilon_t,$$

or equivalently, by letting the same dependent variable on the left as in the pure fractional case,

$$\Delta y_t = \alpha(L)\left(1 - \Delta^{d-1}\right)\Delta y_t + (1 - \alpha(L))\Delta y_t + \varepsilon_t. \quad (12)$$

Note that none of the t-tests considered in the previous sections can properly control the type one error because of the short run correlation induced by $\alpha(L)$ on $\Delta^d y_t$. DGM proposed the use of an augmented test 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. Similarly, in order to keep the linearity of the regression model, we could simplify equation (12) by suppressing the factor $\alpha(L)$ in the first regressor, and consider the regression of Δy_t on $z_{t-1}(d_2)$ and p lags of Δy_t. It is simple to show that this test can properly control the type I error but it is inefficient due to the deletion of the factor $\alpha(L)$ in the first regressor of (12). Hence, we prefer to analyze the following two-step approach that leads to efficient tests.

Note that equation (12) motivates the nonlinear regression model

$$\Delta y_t = \varphi_2\left\{\alpha(L)\left(\Delta^{d_2-1} - 1\right)\Delta y_t\right\} + \sum_{j=1}^{p}\alpha_j \Delta y_{t-j} + u_t,$$

which is similar to (6), except for the inclusion of the lags of Δy_t, and for the filter $\alpha(L)$ in the regressor whose significance is tested. Similar to the white noise case, for continuity reasons, we propose to use the rescaled regression model

$$\Delta y_t = \phi_2\left\{\alpha(L) z_{t-1}(d_2)\right\} + \sum_{j=1}^{p}\alpha_j \Delta y_{t-j} + u_t, \quad (13)$$

with $z_{t-1}(d_2)$ defined in (8). As in the white noise case, the DGP (12) is a particular case of model (13). Under the null hypothesis, $\Delta y_t - \sum_{j=1}^{p}\alpha_j \Delta y_{t-j}$ is i.i.d. and, therefore, $\phi_2 = 0$ for any value of d_2, with $u_t = \varepsilon_t$. Under the alternative hypothesis, when d_2 is chosen equal to d, $\phi_2 = d - 1$ (so that the DGP (12) is recovered), model (13) is properly specified, with regressors $\alpha(L) z_{t-1}(d)$ and $\{\Delta y_{t-j}\}_{j=1}^{p}$ independent of the i.i.d. error term $u_t = \varepsilon_t$. This is not true when d_2 is chosen differently from the true value of d, indicating that an appropriate selection of the input d_2 is needed for deriving optimal tests.

Estimation of model (13) is complicated because of the nonlinearity in the parameters ϕ_2 and $\alpha = (\alpha_1, \ldots, \alpha_p)'$. Compared to the white noise case, note that the practical problem arises because the vector α is unknown, and so, the regressor $\alpha(L) z_{t-1}(d_2)$ is unfeasible.

Hence, first we need to obtain a consistent estimate of α. We propose the following two step procedure.

First, estimate by OLS the equation

$$\Delta^{\widehat{d}_2} y_t = \sum_{j=1}^{p} \alpha_j \Delta^{\widehat{d}_2} y_{t-j} + u_t \tag{14}$$

where the input $\widehat{d}_2$ is any consistent estimator of d that satisfies

$$\widehat{d}_2 = d + O_p\left(T^{-\tau}\right), \ \tau > 0, \text{ and } |\widehat{d}_2| \leq K, \text{ for some } K > 0. \tag{15}$$

The OLS estimator of α is consistent with a convergence rate that depends on the convergence rate of the estimator of d, cf. the proof of Theorem 1 in Appendix 1.

Second, estimate by OLS the equation

$$\Delta y_t = \phi_2 \left[\widehat{\alpha}\left(L\right) z_{t-1}(\widehat{d}_2)\right] + \sum_{j=1}^{p} \alpha_j \Delta y_{t-j} + v_t, \tag{16}$$

where $\widehat{\alpha}\left(L\right)$ denotes the estimator of $\alpha\left(L\right)$ from the first step, and $\widehat{d}_2$ takes the same value as in the first step. The asymptotic null distribution of the resulting t-statistic associated to ϕ_2 is still the standard normal, as if $\widehat{\alpha}$ (and $\widehat{d}_2$) were fixed, because $\phi_2 = 0$ under H_0, see Remark 1.2 in Section 2. Under the alternative, since $\widehat{\alpha}$ converges to the true α and $\widehat{d}_2$ converges to the true d, a t-test for $\phi_2 = 0$ based on (16) has asymptotic properties similar to one based on model (13) with $d_2 = d$.

The next theorem establishes the asymptotic properties of the t-test statistic, t_ϕ, for testing $\phi_2 = 0$ in (16). The proof is in Appendix 2. Introduce the notation

$$\omega^2 = \frac{\pi^2}{6} - \kappa' \Phi^{-1} \kappa,$$

$\kappa = (\kappa_1, ..., \kappa_p)'$ with $\kappa_k = \sum_{j=k}^{\infty} j^{-1} c_{j-k}, \quad k = 1, \ldots, p$, where the c_j are the coefficients of L^j in the expansion of $1/\alpha(L)$, and where $\Phi = \left[\Phi_{k,j}\right], \ \Phi_{k,j} = \sum_{t=0}^{\infty} c_t c_{t+|k-j|}, \ \ k, j = 1, \ldots, p$, denotes the Fisher information matrix for α under Gaussianity.

Theorem 2. *Under the assumption that the DGP is an ARFIMA (p,d,0) model defined as*

$$\alpha\left(L\right) \Delta^d y_t 1\left\{t > 0\right\} = \varepsilon_t,$$

where ε_t is i.i.d. with finite fourth moment, and $\alpha\left(L\right) = 1 - \alpha_1 L - \cdots - \alpha_p L^p$ is a polynomial in the lag operator with all its roots outside the unit circle, the asymptotic properties of the t-ratio test statistic t_ϕ for testing $\phi_2 = 0$ in (16), where the $\widehat{\alpha}$ used in the regressor

$\left\{\widehat{\alpha}(L)\, z_{t-1}(\widehat{d}_2)\right\}$ *is obtained from the OLS estimation of* (14) *and the input* $\widehat{d}_2$ *of* z_{t-1} *satisfies* (15), *are given by:*

a) *Under the null* $(d = 1)$,

$$t_\phi \to_d N(0,1).$$

b) *Under fixed alternatives* $(d < 1)$, *the test based on* t_ϕ *is consistent.*

c) *Under local alternatives* $(d = 1 - \delta/\sqrt{T},\ \delta > 0)$,

$$t_\phi \to_d N(-\delta\omega, 1).$$

Remark 2.1. Note that the drift of the asymptotic distribution under local alternatives coincides with that in Robinson (1994a, Theorem 4), and so, the proposed Wald test is asymptotically locally equivalent to the optimal LM test, similarly to the white noise case. Comparing ω with $h(1) = \pi/\sqrt{6}$ given in Theorem 1, we can observe the asymptotic loss of efficiency due to the estimation of the short memory parameters.

Remark 2.2. In Theorem 2, for simplicity, we have just considered the case where consistent estimators of d are employed as $\widehat{d}_2$, because, as in Theorem 1, these are the only values that lead to efficient tests in this framework. Under condition (10), when $d_2 \neq d$, t-tests are asymptotically standard normal under the null, but inefficient.

Remark 2.3. In a framework similar to the one of this section, Breitung and Hassler (2002) have also proposed a two step procedure that presents two main differences with the one described in this section. First, it is based on the local regressor $z_{t-1}(1)$, and second, in their first step the $\alpha's$ are estimated consistently only under the null hypothesis. However, note that these selections for the long and short term parameters lead to a regression model where the regressor whose significance is tested does not maximize the partial correlation with Δy_t given the p lags of Δy_t for fixed alternatives.

4 Simulations

Next, we examine the finite sample performance of the considered tests by means of a small Monte Carlo study. We consider two Gaussian DGP's, a pure fractionally integrated process and an ARFIMA$(1, d, 0)$. Tables I and II report the results for the first DGP for a nominal level of 0.05 and two samples sizes, 100 and 500, respectively. For Table I the number of replications is 50,000 and for Table II it is 10,000. The parameter d takes values from 0.5 to 1 with increments of 0.05 in Table I, and it takes values from 0.8 to 1 with increments of 0.025 in Table II. These tables report the results of the time domain version of Robinson's LM test, of DGM's test, and of the new efficient Wald test. Regarding DGM's test and the new efficient test, note that the reported results correspond to unfeasible implementations of the tests because they assume that the true d is known and ignore the sampling error associated with the estimation of d. We also have computed these tests using parametric and semiparametric estimators of d with similar results, which are omitted for brevity, the only noticeable difference is some slight additional size distortion when $T = 100$. These tables report the size results ($d = 1$) and size-adjusted power instead of raw power ($d < 1$) because Robinson's LM test is somewhat conservative compared to DGM's test and the new efficient test for $T = 100$.

The main messages from these two tables are the following. First, as expected, the most powerful test is the proposed efficient test which can improve the size-adjusted power up to 30% with respect to DGM's original proposal. Second, compared to the efficient test, the loss of power of the LM test is larger the further from the null the alternative is, reflecting the local character of this test.

In Table III we consider the case where the DGP is a Gaussian ARFIMA$(1, d, 0)$ with autoregressive parameter $\alpha_1 = \{-0.5, 0, 0.3, 0.6, 0.8\}$. We only report the results for one negative value for α_1 because for other negative values the results were similar, contrary to the $\alpha_1 > 0$ case, where finite sample power depends greatly on α_1. In addition, the most empirically relevant 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 two samples sizes, 100 and 500, with 50,000 and 10,000 replications, respectively.

We report results for three tests: a) the original unfeasible augmented DGM's test that uses $d_1 = d$, b) the unfeasible two step efficient test that ignores the sampling variation associated with the estimation of d, and c) the feasible two step test that uses as d_2 the Gaussian semiparametric estimator of Velasco (1999b) with bandwidth $m = T^{0.55}$. In Table III these tests are denoted by ADGM, 2S and 2SSP, respectively. For the three tests we have included one lag in the augmented regression.

Next, we comment on the results from Table III. Note that under the null hypothesis, for any value of α_1, the empirical rejection probabilities are above the nominal level for all tests. This size distortion is especially apparent for the feasible 2SSP test, as we could expect, because the estimation of d leads to an increase in the sampling variation of the test statistic. Hence, we report size-adjusted power instead of raw power. The most noticeable feature of Table III is that power is higher when the serial correlation is negative, and deteriorates substantially, and rapidly, as α_1 becomes positive and large. For instance, it is interesting to observe the enormous loss of power associated to an increase of α_1 from 0.6 to 0.8. When $\alpha_1 = 0.8$ and $T = 100$, the three tests report very low size-adjusted power, indicating that, in the presence of moderate or strong positive correlated innovations, long time series are needed in order to discriminate reasonably well between fractional integration and weak dependence.

Table III also indicates that the unfeasible efficient 2S test presents higher size-adjusted power than the unfeasible ADGM, as expected, and that this difference is especially relevant when positive serial correlation is present, the case of most practical interest. In particular, for $\alpha_1 = 0.8$ and $T = 500$, the 2S test presents twice as much power as ADGM test for values of d between 0.6 and 0.7. In addition, note that the loss of power of the feasible 2SSP test compared to the unfeasible 2S test is rather moderate, except for the $\alpha_1 = 0.8$ case. Also, the case $\alpha_1 = 0$ is interesting for comparing the loss of power of introducing an irrelevant regressor in the augmented regression. Comparing Tables I and II with Table III, it is noticeable that this loss of power is substantial, up to 50%, indicating that a careful selection of the number of lags included in the augmented regression is crucial to balance the trade-off between size and power that a researcher faces in practice. Finally, notice that the non-monotonic behavior for the power figures, when $\alpha_1 = 0.8$, could be due to the fact that the high persistence of the AR(1) makes difficult to distinguish a unit root from long memory for high values of d and relatively small sample sizes.

5 Conclusions and Further Research

In this article we have introduced efficient Wald tests for fractional unit roots by using a model based auxiliary regression. The proposed tests are locally asymptotically equivalent to the locally optimal LM tests of Robinson (1991, 1994a). In addition, the first order asymptotic properties of the proposed tests are not affected by the estimation of short or long memory parameters. We finish with some suggestions on further research. Since our test presents a clear analogy with the original Dickey-Fuller test, it can be interesting to study the cases where deterministic trends or structural breaks may appear in the data generating process. In addition, note that the techniques employed in this paper can also be applied in a multivariate framework for testing simply and efficiently for (fractional) cointegration. In this article we have just considered the case where the short range correlation follows an autoregressive process of known order. An extension of practical interest is to examine the robustness of these procedures in the presence of short term serial correlation of unknown form. This analysis entails studying the behavior of these procedures when the order of the autoregression increases with the sample size. Finally, studying the effects of truncating the fractional filter is another area that deserves more attention. In this respect, Robinson (2005) provides an approach for handling this issue.

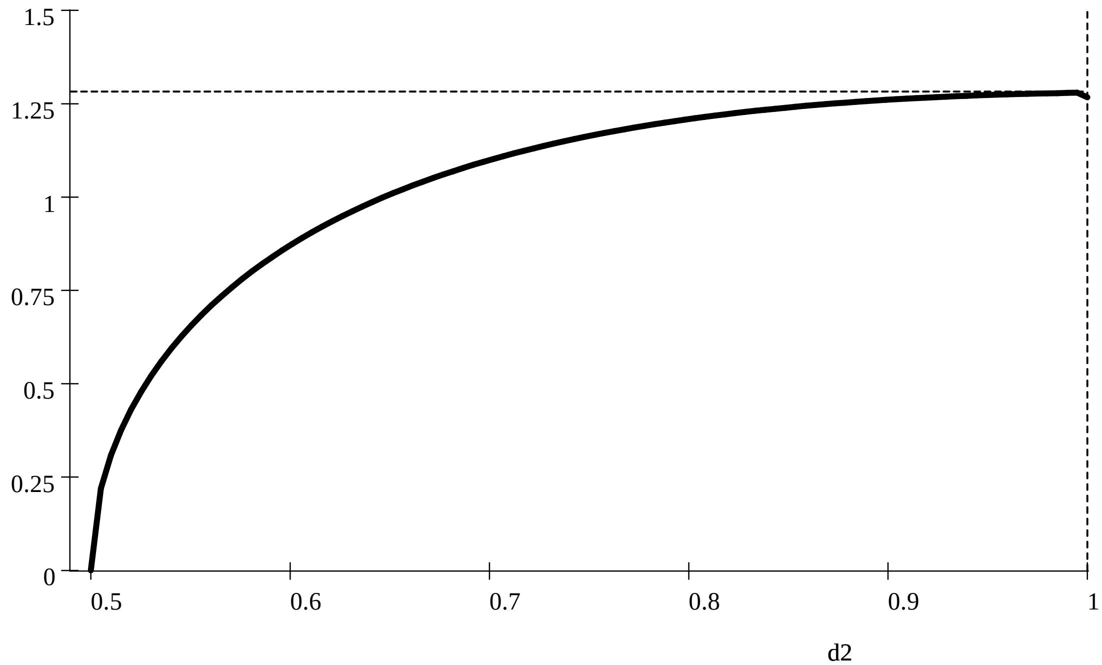

Figure 1. Plot of $h(d_1)$. The horizontal line at $\pi/\sqrt{6} \approx 1.28$ corresponds to Robinson's LM test.

$T = 100$	d	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**	**0.95**	**1**
LM		99.9	99.7	99.1	97.1	92.1	81.5	64.6	44.6	25.8	12.6	4.53
DGM	$d_1 = d$	100	99.9	99.6	98.2	93.6	82.1	64.2	43.0	24.5	11.9	5.27
EFF-W	$d_2 = d$	100	100	100	99.9	98.3	91.8	76.8	53.6	30.7	13.7	5.59

Table I. Monte Carlo size ($d = 1$) and (size adjusted) power ($d < 1$) of Robinson's time domain LM test, DGM's test and the new efficient Wald test: Percentage of rejections based on 5% nominal level. Series follow a pure Gaussian fractionally integrated process with parameter d. Sample size is 100. The number of replications is 50,000.

$T = 500$	d	**0.8**	**0.825**	**0.85**	**0.875**	**0.9**	**0.925**	**0.95**	**0.975**	**1**
LM		100	99.9	99.1	95.9	83.9	62.5	35.8	15.1	5.50
DGM	$d_1 = d$	99.9	99.3	96.4	88.6	73.0	51.4	29.0	13.2	4.86
EFF-W	$d_2 = d$	100	100	99.7	97.5	87.9	66.5	38.8	16.3	5.12

Table II. Monte Carlo size ($d = 1$) and (size adjusted) power ($d < 1$) of Robinson's time domain LM test, DGM's test and the new efficient Wald test: Percentage of rejections based on 5% nominal level. Series follow a pure Gaussian fractionally integrated process with parameter d. Sample size is 500. The number of replications is 10,000.

	d	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**	**0.95**	**1**
α_1						$T = 100$						
	ADGM	100	99.9	99.3	97.4	91.1	78.5	60.1	40.0	22.8	11.3	6.81
-0.5	2S	100	100	99.8	98.1	92.3	79.8	61.3	40.5	23.1	11.7	6.80
	2SSP	100	99.9	99.3	97.0	90.2	76.7	57.7	37.9	21.5	11.0	7.70
	ADGM	99.4	97.9	92.7	84.8	71.9	55.9	39.7	26.1	16.0	8.9	6.91
0	2S	99.8	98.7	95.0	87.0	74.2	58.0	41.4	27.1	16.5	9.4	6.73
	2SSP	99.3	97.3	93.1	84.2	70.8	54.7	38.8	25.3	15.5	9.0	7.67
	ADGM	94.2	86.9	75.9	62.5	48.7	35.9	25.4	17.3	11.3	7.5	6.87
0.3	2S	95.9	89.5	79.3	65.9	51.6	38.1	26.7	18.1	12.1	8.0	6.86
	2SSP	93.3	86.4	76.2	63.4	49.7	37.0	26.1	17.9	11.8	7.8	7.33
	ADGM	56.8	44.4	33.6	24.9	18.5	13.6	10.4	8.2	6.5	5.5	7.11
0.6	2S	63.1	50.3	38.5	28.9	21.4	15.7	11.9	9.1	7.3	6.2	6.95
	2SSP	57.9	47.2	37.0	28.4	21.6	16.1	12.1	9.3	7.4	6.1	6.79
	ADGM	11.1	8.1	5.9	4.5	3.6	3.1	2.9	3.0	3.2	3.9	7.36
0.8	2S	15.3	10.9	8.1	6.2	5.0	4.3	3.9	3.9	4.0	4.5	7.20
	2SSP	12.3	9.3	7.1	5.7	4.6	4.0	3.7	3.6	3.7	4.2	6.45
						$T = 500$						
	ADGM	100	100	100	100	100	100	99.9	96.8	73.6	28.9	5.79
-0.5	2S	100	100	100	100	100	100	100	97.6	75.0	29.4	5.74
	2SSP	100	100	100	100	100	100	99.9	96.9	73.2	28.7	6.63
	ADGM	100	100	100	100	100	99.8	97.6	83.7	51.2	19.9	5.73
0	2S	100	100	100	100	100	100	98.7	86.6	53.5	20.9	5.54
	2SSP	100	100	100	100	100	99.9	97.9	84.4	51.6	20.5	6.49
	ADGM	100	100	100	100	99.5	97.3	84.0	59.2	32.1	14.2	5.55
0.3	2S	100	100	100	100	99.9	98.5	88.8	64.6	35.8	15.6	5.43
	2SSP	100	100	100	100	99.9	97.7	86.7	63.0	34.6	15.2	6.27
	ADGM	100	99.6	98.0	90.4	75.0	54.1	35.2	21.4	12.6	7.6	5.61
0.6	2S	100	100	99.8	96.4	85.2	65.2	43.8	26.5	15.4	8.6	5.41
	2SSP	100	99.9	98.7	93.7	82.1	63.3	43.6	26.9	15.8	8.9	6.18
	ADGM	63.7	40.4	22.8	12.5	7.0	4.1	3.0	2.7	2.8	3.5	6.01
0.8	2S	85.4	64.4	41.9	24.6	14.5	8.6	5.8	4.6	4.2	4.3	5.97
	2SSP	68.9	51.9	36.1	23.9	15.0	9.5	6.2	4.8	4.4	4.5	6.22

Table III. Monte Carlo size ($d = 1$) and (size adjusted) power ($d < 1$) of the unfeasible augmented DGM's test, the unfeasible efficient two step Wald test (2S) and the feasible two step test based on a semiparametric estimator of d (2SSP). 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. The number of replications is 50,000 when $T = 100$ and 10,000 when $T = 500$.

Appendix 1

We provide here the proof of Theorem 1. The proof of b) is omitted because it is easily obtained using the same methods as DGM's Theorem 3. In addition, since a) is a particular case ($\delta = 0$) of c), we just report the proof for c). For simplicity, and without loss of generality, in this appendix we assume that the variance of ε_t is one. We start by considering the case where the input of z_{t-1}, d_2, is fixed. The case where it is stochastic (and consistent for some fixed value under (10)) is discussed at the end of this appendix.

We begin by introducing some notation. Let the t-test statistic for $\phi_2 = 0$ and be

$$t_\phi = t_\phi(d_2) = \frac{\sum_{t=2}^{T} \Delta y_t z_{t-1}(d_2)}{\widehat{S}_T(d_2)\sqrt{\sum_{t=2}^{T}\left(z_{t-1}(d_2)\right)^2}},$$

where $\widehat{S}_T^2(d_2) = T^{-1}\sum_{t=2}^{T}(\Delta y_t - \widehat{\phi}_2 z_{t-1}(d_2))^2$ and $\widehat{\phi}_2$ denotes the OLS estimator of ϕ_2 in (7). Under local alternatives we have that

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

where $\theta_T := -\delta T^{-1/2}$, $\pi_1(-\theta_T) = \theta_T$, $\pi_2(-\theta_T) = 0.5\theta_T(1+\theta_T)$, and Taylor expanding $\pi_i(\cdot)$ around $\pi_i(0) = 0$, $i > 0$, it is obtained that

$$T^{1/2}\pi_i(-\theta_T) = -i^{-1}\delta + O\left(T^{-1/2}i^{-1}\log^2 i\right), \quad i = 1, 2, \ldots, T,$$

see Delgado and Velasco (2005, Lemma 1) and Robinson and Hualde (2003, Lemma D.1). When $d_2 \neq 1$, note that

$$z_{t-1}(d_2) = \frac{\Delta^{-\eta_T} - \Delta^{-\theta_T}}{1-d_2}\varepsilon_t 1\{t > 0\} = \varepsilon_{t-1} + \sum_{i=2}^{t-1}\psi_i(\theta_T, \eta_T)\varepsilon_{t-i},$$

where $\eta_T = 1 - d_2 - \delta T^{-1/2}$ and $\psi_i(\theta_T, \eta_T) = \left(\pi_i(-\eta_T) - \pi_i(-\theta_T)\right)/(1-d_2)$.

First, consider the numerator of $t_\phi(d_2)$ scaled by $T^{-1/2}$,

$$\begin{aligned} Q_T(d_2) &:= T^{-1/2}\sum_{t=2}^{T}\Delta y_t z_{t-1}(d_2) \\ &= T^{-1/2}\sum_{t=2}^{T}\left(\varepsilon_t + \sum_{i=1}^{t-1}\left(\frac{-\delta}{i\sqrt{T}}\right)\varepsilon_{t-i}\right)\left(\varepsilon_{t-1} + \sum_{i=2}^{t-1}\psi_i(\theta_T, \eta_T)\varepsilon_{t-i}\right) \qquad (17) \\ &\quad + T^{-1/2}\frac{\delta^2}{2T}\sum_{t=2}^{T}\left(\sum_{i=1}^{t-1}\pi_i^{(2)}(-\theta^*)\varepsilon_{t-i}\right)\left(\varepsilon_{t-1} + \sum_{i=2}^{t-1}\psi_i(\theta_T, \eta_T)\varepsilon_{t-i}\right), \qquad (18) \end{aligned}$$

where $\pi_i^{(2)}$ is the second derivative of $\pi_i(\cdot)$ and θ^* is some point between 0 and θ_T. Note that $\left|\pi_i^{(2)}(-\theta^*)\right| \leq C i^{-1}\log^2 i$, $i = 1, \ldots, T$ by Lemma 1(b) of Delgado and Velasco (2005). Since (17) is $O_p(1)$, as it is showed next, it is straightforward to show that (18) is $o_p(1)$.

The leading term (17) of $Q_T(d_2)$ can be written as

$$
\begin{aligned}
& T^{-1/2}\sum_{t=2}^{T}\left(\varepsilon_t-\frac{\delta}{\sqrt{T}}\varepsilon_{t-1}-\sum_{i=2}^{t-1}\frac{\delta}{i\sqrt{T}}\varepsilon_{t-i}\right)\left(\varepsilon_{t-1}+\sum_{i=2}^{t-1}\psi_i(\theta_T,\eta_T)\varepsilon_{t-i}\right) \\
= \; & -T^{-1/2}\sum_{t=2}^{T}\left(\frac{\delta}{\sqrt{T}}\varepsilon_{t-1}^2+\sum_{i=2}^{t-1}\frac{\delta}{i\sqrt{T}}\psi_i(\theta_T,\eta_T)\varepsilon_{t-i}^2\right) \qquad (19)\\
& +T^{-1/2}\sum_{t=2}^{T}\varepsilon_t\left(\varepsilon_{t-1}+\sum_{i=2}^{t-1}\psi_i(\theta_T,\eta_T)\varepsilon_{t-i}\right) \qquad (20)\\
& -T^{-1/2}\sum_{t=2}^{T}\left(\frac{\delta}{\sqrt{T}}\varepsilon_{t-1}\right)\left(\sum_{i=2}^{t-1}\psi_i(\theta_T,\eta_T)\varepsilon_{t-i}\right) \qquad (21)\\
& -T^{-1/2}\sum_{t=2}^{T}\left[\sum_{i=2}^{t-1}\frac{\delta}{i\sqrt{T}}\varepsilon_{t-i}\left(\varepsilon_{t-1}+\sum_{j=2,j\neq i}^{t-1}\psi_i(\theta_T,\eta_T)\varepsilon_{t-j}\right)\right]. \qquad (22)
\end{aligned}
$$

The last two terms, (21) and (22), in the previous expression are $o_p(1)$ using arguments similar to those in the proof of Theorem 4 in DGM. Using the properties of the fractional difference filter, and a weak law of large numbers, see for instance, the proof of Lemma 1 in DGM, the term (19) converges in probability to $-\delta K(d_2)$, where

$$
\begin{aligned}
K(d_2) &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\left(1+\sum_{i=2}^{t-1}\frac{1}{i}\psi_i(\theta_T,\eta_T)\right)\\
&= \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\left(1+\sum_{i=2}^{t-1}\frac{\pi_i(-\eta_T)}{i(1-d_2)}\right)=\sum_{i=1}^{\infty}\frac{\pi_i(d_2-1)}{i(1-d_2)}.
\end{aligned}
$$

Using a standard central limit theorem for martingale difference sequences, the term (20) converges in distribution to a $N(0,V)$, where

$$
\begin{aligned}
V &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}E\left(\varepsilon_t\varepsilon_{t-1}+\sum_{i=2}^{t-1}\psi_i(\theta_T,\eta_T)\varepsilon_t\varepsilon_{t-i}\right)^2\\
&= \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}E\left(\sum_{i=1}^{t-1}\frac{\pi_i(d_2-1)}{(1-d_2)}\varepsilon_t\varepsilon_{t-i}\right)^2=\frac{\sum_{i=1}^{\infty}\pi_i(d_2-1)^2}{(1-d_2)^2}<\infty,
\end{aligned}
$$

because $1-d_2<0.5$ and $d_2\neq 1$. Hence, $Q_T(d_2)\to_d N(-\delta K(d_2),\sum_{i=1}^{\infty}(\pi_i(d_2-1)/(1-d_2))^2)$.

Second, consider the denominator of $t_\phi(d_2)$ scaled by $T^{-1/2}$. It is straightforward to show that $\widehat{S}_T^2(d_2)\to_p 1$, and, given the above expression for $z_{t-1}(d_2)$, by a law of large numbers it is simple to see that the limit in probability of $T^{-1}\sum_{t=2}^{T}(z_{t-1}(d_2))^2$ is given by

$$
\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}E\left(\varepsilon_{t-1}+\sum_{i=2}^{t-1}\psi_i(\theta_T,\eta_T)\varepsilon_{t-i}\right)^2=\frac{\sum_{i=1}^{\infty}\pi_i(d_2-1)^2}{(1-d_2)^2}.
$$

So far we have considered the case where $d_2 \neq 1$. The case $d_2 = 1$ follows similarly as above, the difference is that under the local alternative $z_{t-1}(1)$ is now expressed as

$$z_{t-1}(1) = J(L)\Delta^{-\theta_T}\varepsilon_t 1\{t > 0\}.$$

Note that the filter $\psi_T^*(L) := J(L)\Delta^{-\theta_T}$ can be expressed as $\psi_T^*(L) = \sum_{j=1}^{\infty}\psi_{T,i}^* L^i$ where

$$\psi_{T,i}^* = \sum_{j=1}^{i}\frac{1}{j}\pi_{i-j}(-\theta_T),\ i = 1, 2, 3, ...,$$

so that $\psi_{T,i}^* = i^{-1}\left(1 + O(\log T/\sqrt{T})\right)$ uniformly in $i = 1, ..., T$. Using this definition of $z_{t-1}(1)$, all the previous results can be easily adapted. For instance, we have that $K(1) = \lim_{T\to\infty} T^{-1}\sum_{t=2}^{T}\left(1 + \sum_{i=2}^{t-1} i^{-1}\psi_{T,i}^*\right) = \sum_{i=1}^{\infty} i^{-2} = \pi^2/6$.

Next, we analyze briefly the case of an stochastic input $\widehat{d}_2$ that satisfies condition (10) in the text. To show that $t_\phi\left(\widehat{d}_2\right) \to_p t_\phi(d_2)$, we just analyze here the most critical component of $t_\phi(d_2)$, which is the scaled numerator, $Q_T(d_2)$, the analysis for the denominator is similar but simpler. Note that under the null, for $d_2 \neq 1$, $Q_T(d_2)$ simplifies to

$$Q_T(d_2) = T^{-1/2}\sum_{t=1}^{T}\varepsilon_t\left(\frac{\Delta^{d_2-1}-1}{1-d_2}\right)\varepsilon_t = \frac{T^{-1/2}}{1-d_2}\sum_{t=1}^{T}\varepsilon_t\sum_{j=1}^{t-1}\pi_j(d_2-1)\varepsilon_{t-j}.$$

For $d_2 > 0.5$, $Q_T(d_2)$ converges to a zero mean normal variate in distribution, as we have seen above. Then, proceeding as in Robinson and Hualde (2003, Proposition 9), we just need to prove that, for $d_2 \neq 1$,

$$(1-d_2)Q_T(d_2) - \left(1-\widehat{d}_2\right)Q_T\left(\widehat{d}_2\right) \tag{23}$$

$$= T^{-1/2}\sum_{t=1}^{T}\varepsilon_t\sum_{j=1}^{t-1}\left\{\pi_j(d_2-1) - \pi_j\left(\widehat{d}_2-1\right)\right\}\varepsilon_{t-j} \tag{24}$$

is $o_p(1)$. Note that, for $j = 1, 2, \ldots, T$, the expression $\left\{\pi_j(d_2-1) - \pi_j\left(\widehat{d}_2-1\right)\right\}$ equals

$$\sum_{r=1}^{R-1}\frac{1}{r!}\left(d_2-\widehat{d}_2\right)^r\pi_j^{(r)}(d_2-1) + \frac{1}{R!}\left(d_2-\widehat{d}_2\right)^R\pi_j^{(R)}\left(\bar{d}_2-1\right), \tag{25}$$

where $\bar{d}_2$ is an intermediate point between d_2 and $\widehat{d}_2$. Using (25), (24) can be written as

$$T^{-1/2}\sum_{t=1}^{T}\varepsilon_t\sum_{j=1}^{t-1}\left\{\sum_{r=1}^{R-1}\frac{1}{r!}\left(d_2-\widehat{d}_2\right)^r\pi_j^{(r)}(d_2-1)\right\}\varepsilon_{t-j} \tag{26}$$

$$+T^{-1/2}\sum_{t=1}^{T}\varepsilon_t\sum_{j=1}^{t-1}\left\{\frac{1}{R!}\left(d_2-\widehat{d}_2\right)^R\pi_j^{(R)}\left(\bar{d}_2-1\right)\right\}\varepsilon_{t-j}. \tag{27}$$

Since $|\pi_j^{(r)}(d_2-1)| \le Cj^{-d_2}\log^r j,\ j=1,2,\ldots,T$, see Robinson and Hualde (2003, Lemma D1), it is straightforward to check that

$$T^{-1/2}\sum_{t=1}^{T}\varepsilon_t\sum_{j=1}^{t-1}\pi_j^{(r)}(d_2-1)\,\varepsilon_{t-j}=O_p(1)\,,\quad r=1,2,\ldots,R-1,$$

because it has zero mean and finite variance since the sequence $\pi_j^{(r)}(d_2-1)$ is square summable when $d_2>0.5$. Then, using condition (10), we derive that (26) is $o_p(1)$. In order to analyze (27), note that $|\pi_j^{(R)}\left(\bar{d}_2-1\right)| \le Cj^{-\bar{d}_2}\log^R j \le Cj^{-1/2},\quad j=1,2,\ldots,T$, because $\bar{d}_2>0.5$. Therefore, the remainder term

$$T^{-1/2}\sum_{t=1}^{T}\varepsilon_t\sum_{j=1}^{t-1}\pi_j^{(R)}\left(\bar{d}_2-1\right)\varepsilon_{t-j} \tag{28}$$

has first absolute moment bounded by

$$\begin{aligned}
& T^{-1/2}\sum_{t=1}^{T}\left(E|\varepsilon_t|^2\right)^{1/2}\left\{E\left[\left(\sum_{j=1}^{t-1}\pi_j^{(R)}\left(\bar{d}_2-1\right)\varepsilon_{t-j}\right)^2\right]\right\}^{1/2}\\
\le\ & CT^{-1/2}\sum_{t=1}^{T}\left\{\sum_{j=1}^{t-1}\sum_{k=1}^{t-1}E|\pi_j^{(R)}\left(\bar{d}_2-1\right)\pi_k^{(R)}\left(\bar{d}_2-1\right)\varepsilon_{t-j}\varepsilon_{t-k}|\right\}^{1/2}\\
\le\ & CT^{-1/2}\sum_{t=1}^{T}\left\{\sum_{j=1}^{t-1}\sum_{k=1}^{t-1}(jk)^{-1/2}E|\varepsilon_{t-j}\varepsilon_{t-k}|\right\}^{1/2}\le T^{-1/2}\sum_{t=1}^{T}t^{1/2}\le CT.
\end{aligned}$$

Therefore (28) is $O_p(T)$, and if we choose R such that $R\tau>1$, so that $\left(d_2-\widehat{d}_2\right)^R = o_p\left(T^{-1}\right)$, (27) is of order $o_p(1)$ and Theorem 1 follows.

Appendix 2

In this appendix we give a sketch of the proof of Theorem 2.c. The proof of part b) is omitted because it can be easily derived using the same methods as DGM's Theorem 7. We assume that the true d is known, the proof when d is consistently estimated is similar but lengthier and employ similar techniques as those explained at the end of Appendix 1.

The key idea is to use the basic equation of multivariate regression

$$t_\phi=\sqrt{T}\frac{R_T}{\sqrt{1-R_T^2}}, \tag{29}$$

where R_T denotes the sample partial correlation coefficient between $Y_t:=\Delta y_t$ and $X_t:=\alpha(L)z_{t-1}(d)$ given the p lags of Δy_t, $Z_t:=(Z_{t,1},...,Z_{t,p})'$ with $Z_{t,k}=\Delta y_{t-k}$, $k=1,...,p$, to derive the drift of the asymptotic distribution of t_ϕ. Note that the denominator in (29) tends to 1 in probability under local alternatives for which the DGP is given by

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

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

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

with $H_T(L) = \sum_{j=1}^{\infty} h_{T,j}L^j$, so that $|h_{T,j}| \leq Cj^{-1}\log^2 j$, $j \geq 1$, uniformly in T. Then, we can write the series involved in t_ϕ in terms of the i.i.d. variables ε_t, as follows: $Y_t = \alpha(L)^{-1}\Delta^{\delta/\sqrt{T}}\varepsilon_t$, $X_t = [\alpha(L)J(L)]\Delta y_t = J(L)\Delta^{\delta/\sqrt{T}}\varepsilon_t$, 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_ϕ at a smaller order of magnitude because it is orthogonal to the residuals X_t^*. In order to study X_t^*, notice that

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

whereas

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

Then, the (population) least squares projection coefficients of X_t onto Z_t are given by $\Phi^{-1}\kappa$, and, therefore, $X_t^* = J(L)\varepsilon_t - \kappa'\Phi^{-1}\alpha(L)^{-1}\varepsilon_{t,p}$, $\varepsilon_{t,p} = (\varepsilon_{t-1},\ldots,\varepsilon_{t-p})'$, plus smaller order terms. Next, we have that

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} \sqrt{T}\frac{1}{T}\sum_{t=1}^{T} Y_t^* X_t^* &= E\left[-\delta J(L)\varepsilon_t \cdot \left\{J(L)\varepsilon_t - \kappa'\Phi^{-1}\alpha(L)^{-1}\varepsilon_{t,p}\right\}\right] \\
&= -\delta\left(\sum_{j=1}^{\infty} j^{-2} - \kappa'\Phi^{-1}\kappa\right) = -\delta\omega^2,
\end{aligned}$$

and, also $\operatorname{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T}(Y_t^*)^2 = Var[\varepsilon_t] = 1$. Therefore, $\operatorname{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T}(X_t^*)^2$ is given by

$$\begin{aligned}
& Var\left(J(L)\varepsilon_t - \kappa'\Phi^{-1}\alpha(L)^{-1}\varepsilon_{t,p}\right) \\
= \; & Var(J(L)\varepsilon_t) + Var\left(\kappa'\Phi^{-1}\alpha(L)^{-1}\varepsilon_{t,p}\right) - 2Cov\left(J(L)\varepsilon_t, \kappa'\Phi^{-1}\alpha(L)^{-1}\varepsilon_{t,p}\right) \\
= \; & \pi^2/6 + \kappa'\Phi^{-1}\kappa - 2\kappa'\Phi^{-1}\kappa = \omega^2,
\end{aligned}$$

so that, the drift of t_ϕ is given by $-\delta\omega$, and the theorem follows.

References

Agiakloglou, Ch. and P. Newbold (1994): "Lagrange Multiplier tests for Fractional Difference," *J. of Time Series Analysis*, 15, 253-262.

Breitung, J. and U. Hassler (2002): "Inference on the Cointegration Rank in Fractionally Integrated Processes," *Journal of Econometrics*, 110, 167-185.

Delgado, M.A. and C. Velasco (2005): "Sign Tests for Long Memory Time Series," *Journal of Econometrics,* 128, 215-251.

Dickey, D.A., and W.A. Fuller (1979): "Distribution of Estimators of Autoregressive Time Series with a Unit Root," *J. of the American Statistical Association*, 74, 427-431.

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

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

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

_____ (1994b): "Time Series with Strong Dependence," in *Advances in Econometrics: Sixth World Congress* (C.A. Sims, ed.), vol.1, 47-95, Cambridge University Press.

_____ (1995): "Gaussian Semiparametric Estimation of Long Range Dependence," *Annals of Statistics,* 23, 1630-1661.

_____ (2005): "Efficiency Improvements in Inference on Stationary and Nonstationary Fractional Time Series," *Annals of Statistics,* 33, 1800-1842.

Robinson, P.M. and J. Hualde (2003): "Cointegration in fractional systems with unknown integration orders," *Econometrica*, 71, 1727-1766.

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,"*J. of the American Statistical Association*, 95, 1229-1243.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, the MIT Press, Cambridge, Massachusetts.