

# AN OVERVIEW OF NEW RESULTS ON SIMPLE WALD TESTS FOR THE FRACTIONAL INTEGRATION PARAMETER

BY JUAN J. DOLADO<sup>a</sup>, JESUS GONZALO<sup>a</sup>, AND LAURA MAYORAL<sup>b</sup> \*

<sup>a</sup>Dept. of Economics, Universidad Carlos III de Madrid.

<sup>b</sup>Institut d'Anàlisi Econòmica (CSIC).

December 28, 2007

## Abstract

This paper presents an overview of some results regarding an easily implementable Wald test-statistic (EFDF test) for the null hypotheses that a time-series process is  $I(1)$  or  $I(0)$  against fractional  $I(d)$  alternatives, with  $d \in (0, 1)$ , allowing for unknown deterministic components and serial correlation in the error term. Specifically, we argue that the EFDF test has better power properties under fixed alternatives than other available tests for fractional roots, as well as analyze how to implement this test when the deterministic components or the long-memory parameter are subject to structural breaks.

JEL Clasification: C12, C22

Keywords: Fractional processes, Deterministic components, Power, Structural breaks.

---

\*Corresponding E-mail: dolado@eco.uc3m.es. This paper was prepared for the conference in honour of David F. Hendry (Oxford, august 23-25, 2007) whose influence on our research has been much more persistent than the effects of the fractional integration considered here. We are grateful to two anonymous referees and Carlos Velasco for many insightful comments. Financial support from the Spanish Ministry of Education through grants SEJ2006-00369 and SEJ2004-04101, and also from the Barcelona Economics Program of CREA is gratefully acknowledged. The usual disclaimer applies.

## 1. INTRODUCTION

It is well known that tests of  $I(1)$  vs.  $I(0)$ , or  $I(0)$  vs.  $I(1)$ , processes reject the corresponding null hypotheses very occasionally when the true DGP for a time series  $\{y_t\}$  is a fractionally integrated,  $I(d)$ , process. This is often the case for the Dickey-Fuller (DF)- type tests if  $0.5 < d < 1$  and for the KPSS-type test if  $0 < d < 0.5$ . Given that the microfoundations of  $I(d)$  processes make them quite plausible in practice, this issue can have serious consequences when studying the medium and long- run properties of the variables of interest.<sup>1</sup> To mention only a few: (i) shocks could be identified as permanent when in fact they die out eventually, and (ii) two series could be considered as spuriously cointegrated when they are independent at all leads and lags (see, e.g., Gonzalo and Lee, 1998). These mistakes are more likely to occur in the presence of deterministic components like, e.g., in the case of trending economic variables. Additionally, if the true DGP is an  $I(0)$  process subject to structural breaks in its deterministic components, then it could be misinterpreted as a long-memory process, or viceversa.

In view of these caveats, the goal of this paper is four-fold. First, we illustrate the advantages, in terms of power, of recently proposed Wald tests of  $I(d_0)$  vs.  $I(d)$ ,  $d \neq d_0$ , with  $d_0 = 1$  or  $d_0 = 0$ , relative to well-known LM and semiparametric tests; for simplicity, we do this in a setup when the time series has *i.i.d.* error terms and is free of deterministic components. Secondly, we extend the previous procedures to allow for these components, possibly subject to structural breaks. Next, we derive new results both for the LM and Wald tests when the DGP is an  $I(d)$  series whose long-memory parameter,  $d$ , may change over the sample period. And, finally, we extend the previous testing approaches to the case of autocorrelated disturbances

Specifically, we focus on a modification of the Fractional Dickey-Fuller (FDF) test proposed by Dolado, Gonzalo and Mayoral (2002; DGM hereafter) recently developed by Lobato and Velasco (2007; LV hereafter) which achieves an improvement in efficiency over the former. Although this test, henceforth denoted as the EFDF (efficient FDF) test, was originally devised to extend the traditional DF test of  $I(1)$  against  $I(0)$  to the broader framework of  $I(1)$  against  $I(d)$  processes, with  $d \in [0, 1)$ , we show that it can be easily generalized to cover the case of  $I(0)$  vs.  $I(d)$ , with

---

<sup>1</sup>For explanations of the origin of  $I(d)$  processes based on aggregation of individual stationary series with heterogeneous persistence, see Robinson (1978) and Granger (1980), and for models which mimic some of the key properties of  $I(d)$  processes based on the existence of shocks that die out at a certain probabilistic rate, see Parke (1999) and Diebold and Inoue (2001). For persuasive macroeconomic applications of these processes, see Michelacci and Zaffaroni (2000) and Lo and Haubrich (2001).

$d \in (0, 1]$ , processes. The idea behind this approach is to construct a simple regression model where both the regressand and the regressor are filtered so as to become  $I(0)$  under the null and the alternative hypotheses, respectively.<sup>2</sup> The test is based upon the statistical significance of the coefficient on the regressor,  $\varphi$ , by means of its t-ratio,  $t_\varphi$ . Thus, when testing  $I(1)$  vs.  $I(d)$ ,  $\Delta y_t$  becomes the dependent variable, where  $\Delta = (1 - L)$ . As regards the regressor, whereas DGM choose  $\Delta^d y_{t-1}$ , LV show that  $z_{t-1}(d) = (1-d)^{-1}(\Delta^{d-1} - 1)\Delta y_t$  improves the efficiency of the test.<sup>3</sup> These tests belong to the Wald family because their underlying regression models are estimated under the alternative hypothesis. Thus, non-rejection of  $H_0: \varphi = 0$  against  $H_1: \varphi < 0$ , implies that the process is  $I(1)$  and, conversely, rejection of the null implies that the process is  $I(d)$ . As will be shown below, the EFDF test for testing  $I(0)$  vs.  $I(d)$  is based on an analogous t-ratio,  $t_\psi$ , this time obtained from the regression of  $y_t$  on  $d^{-1}(1 - \Delta^d)y_t$ .

In order to compute either version of the EFDF test, an input value for  $d$  is required. One could either consider a (known) simple alternative,  $H_A : d = d_A < 1$  or, more realistically, a composite one,  $H_1 : d < 1$ . In the latter case, which is the one we focus on here, it suffices to use a  $T^\kappa$ -consistent estimate (with  $\kappa > 0$ ) of the true integration order to achieve a limiting distribution of the resulting statistic that is a  $N(0, 1)$ .

Under a sequence of local alternatives approaching  $H_0 : d = 1$  from below at a rate of  $T^{-1/2}$ , LV (2007, Theorem 1) proved that, under Gaussianity, the EFDF test of  $I(1)$  vs.  $I(d)$  is asymptotically equivalent to the uniformly most powerful invariant (UMPI) test, i.e., the LM test introduced by Robinson (1991, 1994) and later adapted by Tanaka (1999) for the time domain. This result also holds for the  $I(0)$  vs.  $I(d)$  case. Our first contribution in this paper is to analyze the properties of Wald and LM tests in the case where the alternative is fixed. Not surprisingly, although both tests are consistent and diverge at the same rate, in line with the standard result about the better power properties of Wald tests relative to LM tests (see Engle, 1984), our findings here point out that the former test fares better than the latter for a given fixed alternative in

---

<sup>2</sup>In the DF setup, these filters are  $\Delta = (1 - L)$  and  $\Delta^0 = 1$ , so that the regressand and regressor are  $\Delta y_t$  and  $y_{t-1}$ , respectively.

<sup>3</sup>As explained in DGM (Appendix A; 2002), both regressors can be constructed by filtering the series  $\{y_t\}_{t=1}^T$  with the truncated version at the origin (with pre-sample shocks set to 0) of the binomial expansion of  $(1 - L)^d$  in the lag operator  $L$ . Thus,  $\Delta_+^d y_t = \sum_{i=0}^{t-1} \pi_i(d) y_{t-i}$ , where  $\pi_i(d)$  is the  $i$ -th coefficient in that expansion (for more details, see the end of this section). This “deadstart” fractional process has been popularized, among others, by Robinson and Marinucci (2001), giving rise to Type- II fractional Brownian motion. Since the limit distributions of the EFDF tests discussed throughout this paper are always Gaussian, none of the results depend on this choice. To simplify the notation, we will omit the truncation subscript in the sequel and refer to this filter simply as  $\Delta^d$ .

both setups. The reason is that the Wald test has a larger non-centrality parameter. Moreover, when compared to other tests of  $I(1)$  or  $I(0)$  vs.  $I(d)$  which rely on direct inference about semiparametric estimators of  $d$ , in general, the EFDF test exhibits better power properties, under a correct specification of the stationary short-run dynamics of the error term in the auxiliary regression. This is due to the fact that the semiparametric estimation procedures often imply larger confidence intervals of the memory parameter, in exchange with less restrictive assumptions on the error term. By contrast, the combination of a wide range of semiparametric estimators for the input value of  $d$  with auxiliary parametric regressions, as the ones discussed above, yields a parametric rate for the Wald tests.<sup>4</sup> Thus, in a sense, the Wald tests combine the favorable features of both approaches to improve power while reducing the danger of misspecifying short-run dynamics.

Following the development of unit- root tests in the past, our second contribution is to investigate how to implement Wald tests when some deterministic components are considered in the DGP, a case which was not treated in LV (2007). We first focus on the role of a *linear trend* since many (macro) economic time series exhibit this type of trending behavior in their levels. Our main result is that, in contrast with the results for most tests for  $I(1)$  against  $I(0)$  or viceversa, the EFDF test remains efficient in the presence of deterministic components and it maintains the same asymptotic distribution, insofar as they are correctly filtered. In this respect, this result mimics the one found for LM tests when these components are present; cf. Robinson (1994), Tanaka (1999) and Gil-Alaña and Robinson (1997). Next, we analyze the cases where there are structural breaks in the deterministic components, where we devise tests for  $I(d)$  cum constant-parameter deterministic terms vs.  $I(0)$  cum breaks, or in the long-memory parameter,  $d$ , as well as other alternative time varying schemes for  $d$ .

Lastly, we show that the previous asymptotic results obtained for DGPs with *i.i.d.* disturbances remain valid when the error term is allowed to be parametrically autocorrelated, as in the (augmented) ADF setup. We provide a linear single-step estimation procedure to account for (parametric) autocorrelation which simplifies the two-step procedure proposed by LV (2007).

The rest of the paper is structured as follows. Sections 2 briefly overviews the properties of the EFDF tests when the process is either a driftless random walk or *i.i.d.* under the null, and derives new results about their power relative to the power of the LM test under fixed alternatives.

---

<sup>4</sup>LV (2006, 2007) have shown that a Gaussian semiparametric estimator, such as the one proposed by Velasco (1999), suffices to achieve consistency and asymptotic normality in the analyzed Wald tests (see sections 2 and 3 below) extending the results by DGM(2002) about parametric estimators.

Section 3 extends the previous results to processes containing trending deterministic components (e.g., a linear trend) with constant parameters. Section 4 discusses tests to distinguish between  $I(0)$  series whose deterministic terms may be subject to structural breaks and  $I(d)$  processes with constant parameters. Section 5 deals with how to test for breaks in the long-memory parameter,  $d$ , as well as some other alternative time varying structures. Section 6 explains how to modify the previous tests when the error terms are autocorrelated. Finally, Section 7 concludes.

Proofs of the main results are collected in an Appendix available as supplementary material to this paper.

In the sequel, the definition of a  $I(d)$  process that we will adopt is the one used by Akonom and Gourieroux (1987) where a fractional process is initialized at the origin. This corresponds to Type-II fractional Brownian motion (see the previous discussion in footnote 3) and is similar to the definitions of an  $I(d)$  process underlying the LM test proposed Robinson (1994) and Tanaka (1999). Moreover, the following conventional notation is adopted throughout the paper:  $\Gamma(\cdot)$  denotes the Gamma function,  $\{\pi_i(d)\}$  represents the sequence of coefficients associated to the expansion of  $(1-L)^d$  in powers of  $L$ , defined as

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

The indicator function is denoted by  $1_{\{\cdot\}}$ . Finally,  $\xrightarrow{w}$  denotes weak convergence in  $D[0,1]$  endowed with the Skorohod  $J_1$  topology, and  $\xrightarrow{p}$  means convergence in probability.

## 2. THE EFDF TEST

### 2.1 I(1) vs. I(d)

Like in Robinson (1994), we consider an additive model for the process  $y_t$  which is generated as the sum of a deterministic component,  $\mu(t)$ , and an  $I(d)$  component,  $u_t$ , so that

$$y_t = \mu(t) + u_t, \tag{1}$$

where  $u_t = \Delta^{-d}\varepsilon_t 1_{\{t>0\}}$  is a purely stochastic  $I(d)$  process, with  $d \in [0,1]$ , and  $\varepsilon_t$  is a zero-mean *i.i.d.* random variable.

When  $\mu(t) \equiv 0$ ,<sup>5</sup> DGM introduced a Wald-type (FDF) test for testing the null hypothesis  $H_0 : d = 1$  against the composite alternative  $H_1 : d \in [0,1)$ , based on the t-statistic associated to the

---

<sup>5</sup>Alternatively,  $\mu(t)$  could be considered to be known. In this case, the same arguments go through after subtracting it from  $y_t$  to obtain a purely stochastic process.

hypothesis  $\phi = 0$  in the OLS regression

$$\Delta y_t = \phi \Delta^{d^*} y_{t-1} + v_t, \quad (2)$$

where  $d^* \geq 0$  is an input value needed to perform the test. If  $d^*$  is chosen such that  $d^* = \widehat{d}_T$ , where  $\widehat{d}_T$  is a  $T^\kappa$ -consistent estimator of  $d$ , with  $\kappa > 0$ , DGM (2002) and LV (2006) have shown that the asymptotic distribution of the resulting t-statistic,  $t_\phi$ , is  $N(0, 1)$ .

Recently, LV (2007) have proposed the EFDF test based on a modification of regression (2) that permits to achieve higher efficiency while keeping the good finite-sample properties of Wald tests. More specifically, their proposal is to compute the t-statistic,  $t_\varphi$ , associated to the null hypothesis  $\varphi = 0$  in the regression

$$\Delta y_t = \varphi z_{t-1}(d^*) + \varepsilon_t, \quad (3)$$

where  $z_{t-1}(d^*)$  is defined as<sup>6</sup>

$$z_{t-1}(d^*) = \frac{(\Delta^{d^*-1} - 1)}{(1 - d^*)} \Delta y_t,$$

such that  $\varphi = (d^* - 1)$  and  $d^* > 0.5$  is an input value needed to implement the test. Note that, if  $d^*$  is the true integration order of the process,  $d$ , then  $\varphi = 0$  under  $H_0 : d = 1$  so that the model becomes a random walk, i.e.,  $\Delta y_t = \varepsilon_t$ . By contrast, under  $H_1 : d \in [0, 1)$ , it holds that  $\varphi < 0$ , and the model becomes a pure fractional process, i.e.,  $\Delta^d y_t = \varepsilon_t$ .

The insight for why the EFDF test is more efficient than the FDF test is as follows. Let  $d^* = d$ ; Then under  $H_1$ , the regression model considered in (2) can be written as  $\Delta y_t = \Delta^{1-d} \varepsilon_t = \varepsilon_t + (d-1)\varepsilon_{t-1} + 0.5d(d-1)\varepsilon_{t-2} + \dots = \phi \Delta^d y_{t-1} + \varepsilon_t + 0.5d(d-1)\varepsilon_{t-2} + \dots$  with  $\phi = d-1$ . Thus, the error term  $v_t = \varepsilon_t + 0.5d(d-1)\varepsilon_{t-2} + \dots$  in (2) is serially correlated. Although OLS provides a consistent estimator of  $\phi$ , since  $v_t$  is orthogonal to the regressor  $\Delta^d y_{t-1} = \varepsilon_{t-1}$ , it is not the most efficient one. By contrast, the regression model used in the EFDF test does not suffer from this problem since, by construction, it yields an *i.i.d.* error term. In order to distinguish this test from the one proposed in the next subsection for  $H_0 : d = 0$ , we denote it in the sequel as the EFDF(1) test. Finally, note that application of L' Hôpital rule to  $z_{t-1}(d^*)$  in the limit case as  $d^* \rightarrow 1$  leads to a regressor equal to  $-\ln(1-L)\Delta y_t = \sum_{j=1}^{\infty} j^{-1} \Delta y_{t-j}$ , which is the one used in Robinson's LM test.

Theorem 1 in LV (2007), which we reproduce below for completeness, establishes the asymptotic properties of  $t_\varphi$ .

---

<sup>6</sup>A similar model was first proposed by Granger (1986) in the more general context of testing for cointegration with multivariate series, a modification of which has been recently considered by Johansen (2005).

**Theorem 1** Under the assumption that the DGP is given by  $y_t = \Delta^{-d}\varepsilon_t 1_{\{t>0\}}$ , where  $\varepsilon_t$  is i.i.d. with finite fourth moment, the asymptotic properties of the  $t$ -statistic,  $t_\varphi$ , for testing  $\varphi = 0$  in (3), where the input of  $z_{t-1}(\widehat{d}_T)$  is a  $T^\kappa$ -consistent estimator of  $d^*$ , for some  $d^* > 0.5$  with  $\kappa > 0$ , are given by:

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

$$t_\varphi(\widehat{d}_T) \xrightarrow{w} N(0, 1).$$

b) Under local alternatives, ( $d = 1 - \gamma/\sqrt{T}$ ),

$$t_\varphi(\widehat{d}_T) \xrightarrow{w} N(-\gamma h(d^*), 1),$$

where  $h(d^*) = \sum_{j=1}^{\infty} j^{-1} \pi_j (d^* - 1) / \sqrt{\sum_{j=1}^{\infty} \pi_j (d^* - 1)^2}$ ,  $d^* > 0.5$ ,  $d^* \neq 1$ .

c) Under fixed alternatives  $d \in (0, 1)$ , the test based on  $t_\varphi(\widehat{d}_T)$  is consistent.

LV (2007) have shown that the function  $h(d^*)$  achieves a global maximum at 1 where  $h(1) = \sqrt{\pi^2/6}$ , and that  $h(1)$  equals the noncentrality parameter of the locally optimal LM test (see subsection 2.2 below).<sup>7</sup> Thus, insofar as a  $T^\kappa$ -consistent estimator of  $d$ , with  $\kappa > 0$ , is used as an input of  $z_{t-1}(d^*)$ , the EFDF test is locally asymptotically equivalent to Robinson's LM test.

In practice, the obtained estimate of  $d$  could be smaller than 0.5. In these cases, the input value can be chosen according to the following rule:  $\widetilde{d}_{1T} = \max\{\widehat{d}_T, 0.501\}$ , with which the test can be easily proved to diverge under  $H_1$ .

A power-rate consistent estimate of  $d$  can be easily obtained by applying some available semi-parametric estimators. Among them, the estimators proposed by Abadir et al. (2005), Shimotsu (2006a) and Velasco (1999) provide appropriate choices since they also cover the case where deterministic components are present, as we do below.

## 2.2 I(0) vs. I(d)

Although the EFDF(1) test was originally derived for testing  $I(1)$  vs.  $I(d)$  processes, it can be easily extended to cover the case of  $I(0)$  vs.  $I(d)$ , with  $d \in (0, 1]$ . This new test is labelled hereafter as the EFDF(0) test. In effect, as before, the maintained hypothesis is taken to be (1),

---

<sup>7</sup>DGM (2002, Theorem 3) in turn obtained that the corresponding distribution under local alternatives of the FDF test in (2) is  $N(-\gamma, 1)$ . Hence, the asymptotic efficiency of the FDF test relative to the EFDF(1) test is 0.78 ( $\simeq \sqrt{6}/\pi$ ).

but now with the null being  $H_0 : d = 0$ , and the alternative  $H_1 : 0 < d \leq 1$ .<sup>8</sup> We first focus on the simple case where  $\mu(t) \equiv 0$ . Adding and subtracting  $y_t$  to both sides of (1) and solving for  $y_t$ , yields

$$y_t = \psi s_{t-1}(d) + \varepsilon_t, \quad (4)$$

where

$$s_{t-1}(d) = \frac{1 - \Delta^d}{d} y_t,$$

such that  $\psi = d$ . Like in (3),  $s_{t-1}(d)$  does not contain the current value of  $y_t$  since  $(1 - \Delta^d) = (dL + \frac{1}{2}d(d-1)L^2 - \dots)$ . Under  $H_0$ ,  $\psi = 0$ , while, under  $H_1$ ,  $0 < \psi \leq 1$ . When  $\psi = 0$ , the model becomes  $y_t = \varepsilon_t$  whereas it becomes  $\Delta^d y_t = \varepsilon_t$  for  $\psi = d \in (0, 1]$ . As in the  $I(1)$  vs.  $I(d)$  case, equation (4) motivates a test of the null hypothesis  $\psi = 0$  based on the t-statistic of  $\hat{\psi}$ ,  $t_\psi$ , computed in a regression of  $y_t$  on  $s_{t-1}(d^*)$ , where  $d^*$  is an input value needed to make the test feasible. Thus, the null is tested by means an upper-side test based on  $t_\psi$ . As with the EFDF(1) test, the limit case as  $d^* \rightarrow 0$  implies that  $s_{t-1}(d) \rightarrow -\ln(1 - L)y_t = \sum_{j=1}^{\infty} j^{-1} y_{t-j}$ , that is, it yields the regressor used in the LM test.

In this case, the following theory holds

**Theorem 2** *Under the assumption that the DGP is given by  $y_t = \Delta^{-d} \varepsilon_t 1_{\{t>0\}}$ , where  $\varepsilon_t$  is i.i.d. with finite fourth moment, the asymptotic properties of the t-statistic,  $t_\psi$ , for testing  $\psi = 0$  in (4) where the input of the regressor  $s_{t-1}(\hat{d}_T)$  is a  $T^{\kappa}$ -consistent estimator of  $d^*$ , for some  $d^* < 0.5$  with  $\kappa > 0$ , are given by:*

a) *Under the null hypothesis ( $d = 0$ ),*

$$t_\psi(\hat{d}_T) \xrightarrow{w} N(0, 1).$$

b) *Under local alternatives, ( $d = \gamma/\sqrt{T}$ ),*

$$t_\psi(\hat{d}_T) \xrightarrow{w} N(\gamma g(d^*), 1),$$

where  $g(d^*) = \sum_{j=1}^{\infty} j^{-1} \pi_j(d^*) / \sqrt{\sum_{j=1}^{\infty} \pi_j(d^*)^2}$ ,  $d^* < 0.5$ ,  $d^* \neq 0$ .

c) *Under fixed alternatives ( $d \in (0, 1)$ ), the test based on  $t_\psi(\hat{d}_T)$  is consistent.*

---

<sup>8</sup>Note that if we were to take a null of  $I(d_0)$ ,  $d_0 \in (0, 1]$ , and an alternative of  $I(0)$ , the EFDF regression model would be  $\Delta^{d_0} y_t = \rho[\Delta_0^{-1}(\Delta^{-d_0} - 1)]\Delta^{d_0} y_t + \varepsilon_t$ , with  $\rho = -d_0$ . In this case, under  $H_0$ ,  $\rho \neq 0$ , whereas, under  $H_1$ ,  $\rho = 0$ .



It is easy to show that the function  $g(\cdot)$  achieves an absolute maximum at 0, in which case  $g(0)$  equals the noncentrality parameter of the locally optimal Robinson's LM test. Therefore, if the input of  $s_{t-1}(\cdot)$ ,  $\hat{d}_T$ , is a  $T^\kappa$ -consistent estimator of  $d$  with  $\kappa > 0$ , the test based on  $t_\psi(\hat{d}_T)$  is locally optimal. In practice, to perform regression (4) the input value  $\tilde{d}_{0T} = \min\{\hat{d}_T, 0.499\}$  can be employed, so that this value is always strictly smaller than 0.5.

### 2.3 Power comparisons under fixed alternatives

As discussed in the Introduction, the closer competitor to the EFDF test is the LM test proposed by Robinson (1991, 1994) in the frequency domain, subsequently extended by Tanaka (1999) to the time domain. In this section we discuss the power properties of the two competing tests under the case of fixed alternatives.<sup>9</sup>

We start with the LM test, henceforth denoted as  $LM_T$ , which considers  $H_0 : \theta = 0$  against  $H_1 : \theta \neq 0$  for the DGP  $\Delta^{d_0+\theta}y_t = \varepsilon_t$ . In line with the hypotheses considered in this paper, we focus on the particular cases where  $d_0 = 1$  and  $-1 < \theta \leq 0$ , and  $d_0 = 0$  and  $0 < \theta \leq 1$ . Assuming that  $\varepsilon_t \sim n.i.d.$   $(0, \sigma^2)$ , the score-LM test is computed as

$$LM_T = \sqrt{\frac{6}{\pi^2}} T^{1/2} \sum_{j=1}^{T-1} j^{-1} \hat{\rho}_j \xrightarrow{w} N(0, 1), \quad (5)$$

where  $\hat{\rho}_j = \sum_{t=j+1}^T \Delta^{d_0} y_t \Delta^{d_0} y_{t-j} / \sum_{t=1}^T (\Delta^{d_0} y_{t-j})^2$  (see Robinson, 1991 and Tanaka, 1999). Breitung and Hassler (2002) have shown that an alternative way to compute the LM test is as the t-ratio ( $t_\lambda$ ) in the regression

$$\Delta^{d_0} y_t = \lambda x_{t-1}^* + e_t, \quad (6)$$

where  $x_{t-1}^* = \sum_{j=1}^{t-1} j^{-1} \Delta^{d_0} y_{t-j}$ .

Under a sequence of local alternatives of the type  $\theta = 1 - T^{-1/2}\gamma$  with  $\gamma > 0$  for  $H_0 : d_0 = 1$ , the  $LM_T$  (or  $t_\lambda$ ) test is the UMPI test. However, as discussed earlier, the EFDF(1) is asymptotically equivalent to the UMPI test whenever a  $T^\kappa$ -consistent estimator of  $d$ ,  $\hat{d}_T$ , is used to compute the test since the limit case as  $\hat{d}_T \rightarrow 1$  in the filter  $(\Delta^{\hat{d}_T-1} - 1) / (1 - \hat{d}_T)$  yields the linear filter used in the LM test, namely  $-\ln(1 - L) = \sum_{j=1}^{\infty} j^{-1} L^j$ . Similar arguments hold for the EFDF(0) test, where  $H_0 : d_0 = 0$ , and  $\theta = T^{-1/2} \gamma$ .

---

<sup>9</sup>The available results in the literature establish the consistency of the Wald and LM tests and derive their (identical) speed of divergence under fixed alternatives. However, they do not derive the non-centrality parameters as we do below which can be useful to characterize power differences for a *given* sample size.

In the rest of this section, we analyze the case with fixed alternatives where, to our knowledge, results are new. In particular, we first derive the non-centrality parameters of two above-mentioned tests under an  $I(d)$  alternative where the DGP is assumed to be  $\Delta^d y_t = \varepsilon_t$ . The permissible ranges of  $d$  in this analysis are  $d \in [0, 1)$  for the EFDF(1) test, and  $d \in (0, 0.5)$  for the EFDF(0) test.<sup>10</sup> In the case of the EFDF(1) test,  $H_0 : d = 1$  and, hence,  $\Delta y_t = \Delta^{-b} \varepsilon_t$  where  $b = d - 1 < 0$ . Then, the following result holds.

**Theorem 3** *If  $\Delta^d y_t = \varepsilon_t$  with  $d \in [0, 1)$ , the  $t$ -statistic,  $t_\varphi$ , associated to the EFDF(1) test satisfies*

$$T^{-1/2} t_\varphi \xrightarrow{p} - \left( \frac{\Gamma(3-2d)}{\Gamma^2(2-d)} - 1 \right)^{1/2} := c_{1,EFDF}(d),$$

while, under the same DGP, the LM test defined in (5) satisfies

$$T^{-1/2} LM_T \xrightarrow{p} - \sqrt{\frac{6}{\pi^2}} \frac{\Gamma(2-d)}{(1-d)\Gamma(d-2)} \sum_{j=1}^{\infty} \frac{\Gamma(j+d-1)}{j\Gamma(j+2-d)} := c_{1,LM}(d), \quad (7)$$

where  $c_{1,EFDF}(d)$  and  $c_{1,LM}(d)$  denote the non-centrality parameters under the fixed alternative  $H_1 : d \in [0, 1)$  of the EFDF(1) and LM tests, respectively.

Secondly, for the EFDF(0) test,  $H_0 : d = 0$ , whereby now  $y_t = \Delta^{-b} \varepsilon_t$  with  $b = d$ . Then

**Theorem 4** *If  $\Delta^d y_t = \varepsilon_t$  with  $d \in (0, 0.5)$ , the  $t$ -statistic,  $t_\psi$ , associated to the EFDF(0) test satisfies,*

$$T^{-1/2} t_\psi \xrightarrow{p} \left( \frac{\Gamma(1-2d)}{\Gamma^2(1-d)} - 1 \right)^{1/2} := c_{0,EFDF}(d),$$

while, under the same DGP, the LM test defined in (5) satisfies

$$T^{-1/2} LM_T \xrightarrow{p} \sqrt{\frac{6}{\pi^2}} \frac{\Gamma(1+d)}{\Gamma(-d)} \sum_{j=1}^{\infty} \frac{\Gamma(j-d)}{j\Gamma(j+d+1)} := c_{0,LM}(d),$$

where  $c_{0,EFDF}(d)$ , and  $c_{0,LM}(d)$  denote the non-centrality parameters under the fixed alternative  $H_1 : d \in (0, 0.5)$  of the EFDF(0) and LM tests, respectively.

---

<sup>10</sup>The intuition for why the two cases differ is that, under a fixed  $I(d)$  alternative, the EFDF(1) test proceeds to first-difference the series, so that  $\Delta y_t \sim I(d-1)$ , and then, all the variables in regression (2) are stationary under the alternative hypothesis of  $d < 1$ . The EFDF(0) treats the series in levels so that  $y_t \sim I(d)$ , and then for values of  $d > 0.5$ , regression (3) includes both stationary and non-stationary variables. As a result, the LLN can be applied on the EFDF(1) test for all  $d < 1$  but only for values of  $d < 0.5$  in the EFDF(0) case. If  $d > 0.5$ , the non-centrality parameter will converge to a random variable.

Figures 1 and 2 display the two non-centrality parameters derived in Theorems 3 and 4. As expected, the EFDF and the LM tests behave similarly for values of  $d$  very close to the corresponding null hypotheses. However, despite being devised to be the UIMP test for local alternatives, for finite sample sizes, the LM test performs significantly worse than the EFDF tests when the alternative is not local:  $c_{1,EFDF}(d)$  ( $c_{0,EFDF}(d)$ ) is much more negative (positive) than  $c_{1,LM}(d)$  ( $c_{0,LM}(d)$ ) when  $d$  departs from the respective nulls. The intuition for this worse power performance of the LM test is that there does not exist any value for  $\lambda$  in (6) that makes  $e_t$  both *i.i.d.* and independent of the regressor for fixed alternatives, implying that  $x_{t-1}^*$  does not maximize the correlation with  $\Delta^{d_0}y_t$ .

As regards the power of semiparametric estimators, whose confidence intervals could be directly used for inference purposes, both the Fully Extended Local Whittle (FELW, see Abadir et al., 2005) and the Exact Local Whittle estimators (ELW, see Shimotsu and Phillips, 2005) verify the asymptotic property:  $\sqrt{m}(\hat{d}_T - d) \xrightarrow{w} N(0, \frac{1}{4})$  for  $m = o(T^{\frac{4}{5}})$ . For example, test statistics for a unit root are based on  $\tau_d = 2\sqrt{m}(\hat{d}_T - 1) \xrightarrow{w} N(0, 1)$ . Therefore, their rate of divergence under  $H_1 : d < 1$  is the nonparametric rate  $O_p(\sqrt{m})$  which is smaller than the  $O_p(\sqrt{T})$  parametric rate achieved by the Wald test. This loss of power is just the counterpart of the higher robustness against misspecification achieved by semiparametric tests.

[Figures 1 and 2 about here]

### 3. DETERMINISTIC COMPONENTS WITHOUT BREAKS

In the case where  $\mu(t) \neq 0$ , DGM (2007) have derived the properties of the EFDF(1) test when the time series is generated by (1) and  $\mu(t)$  verifies the following condition.

*Condition ET (Evolving trend):*  $\mu(t)$  is a polynomial in  $t$  of known order.

Under Condition *ET*, the DGP is allowed to contain trending regressors in the form of polynomials (of known order) of  $t$ . Hence, when the coefficients of  $\mu(t)$  are unknown, the test described above are unfeasible. Nevertheless, it is still possible to obtain a feasible test with the same asymptotic properties as in Theorem 1 if a consistent estimate of  $\mu(t)$  is removed from the original process. Indeed, under  $H_0$ , the relevant coefficients of  $\mu(t)$  can be consistently estimated by OLS in a regression of  $\Delta y_t$  on  $\Delta \mu(t)$ . For instance, consider the case where the DGP contains a linear time trend, that is,

$$y_t = \alpha + \beta t + \Delta^{-d} \varepsilon_t \mathbf{1}_{\{t>0\}}, \quad (8)$$

which, under  $H_0 : d = 1$ , leads to the popular case of a random walk with drift. Taking first differences, it follows that  $\Delta y_t = \beta + \Delta^{1-d} \varepsilon_t 1_{\{t>0\}}$ . Then, the OLS estimate of  $\beta$ ,  $\hat{\beta}$ , (i.e., the sample mean of  $\Delta y_t$ ) is consistent under both  $H_0$  and  $H_1$ . In effect, under  $H_0$ ,  $\hat{\beta}$  is a  $T^{1/2}$ -consistent estimator of  $\beta$  whereas, under  $H_1$ , it is  $T^{3/2-d}$ -consistent with  $3/2 - d > 0.5$  (see Hosking 1996, Theorem 8). Hence, if one uses the regression model

$$\widetilde{\Delta y}_t = \varphi \widetilde{z}_{t-1}(\hat{d}_T) + e_t, \quad (9)$$

where the input of  $\widetilde{z}_{t-1}(\hat{d}_T)$  is a  $T^\kappa$ -consistent estimator of  $d$  with  $\kappa > 0$ ,  $\widetilde{\Delta y}_t = \Delta y_t - \Delta \hat{\mu}(t)$ ,  $\widetilde{z}_{t-1}(\hat{d}_T) = \frac{(\Delta^{\hat{d}_T-1} - 1)}{(1 - \hat{d}_T)}(\Delta y_t - \Delta \hat{\mu}(t))$ , and the coefficients of  $\Delta \hat{\mu}(t)$  are estimated by an OLS regression of  $\Delta y_t$  on  $\Delta \mu(t)$ , then the asymptotic properties of the EFDF(1) test for testing  $\varphi = 0$  in (9) are identical to those stated in Theorem 1.

Whereas in the previous test, the choice of a  $T^\kappa$ -consistent estimator of  $d$ , with  $\kappa > 0$ , suffices to maintain the asymptotic distribution of the feasible test using residuals, a similar result holds for the EFDF(0) test but this time using a  $T^\kappa$ -consistent estimator of  $d$ , with  $\kappa > 0$ . In this alternative setup of  $I(0)$  vs.  $I(d)$ , the OLS estimators  $\hat{\alpha}$  and  $\hat{\beta}$  in the regression in levels of  $y_t$  on  $\mu(t)$  are  $T^{d-1/2}$  and  $T^{3/2-d}$ -consistent estimators of  $\alpha$  and  $\beta$ , respectively. Consequently, the estimator of the trend slope,  $\hat{\beta}$ , is always consistent for  $d \in (0, 1)$  whereas the estimator of the intercept,  $\hat{\alpha}$ , is only consistent for  $d \in (0, 0.5)$ , implying that the residuals from the OLS detrending procedure in levels are only valid for  $d^* < 0.5$ . Under fixed alternatives, since the true value of  $d$  could well exceed 0.5, one possibility to obtain consistent detrended series is to use Shimotsu's (2006a) detrending approach for  $I(d)$  processes. This author has noticed that if one chooses the initial value of the series,  $y_1$ , as an estimator of  $\alpha$ , then it holds that the deviations  $y_1 - \alpha (= \Delta^{-d} \varepsilon_1 1_{t>0})$  are  $O_p(1)$ , implying that its variance is dominated by the exploding variance of  $y_t$  when  $d \in (0.5, 1)$ . Thus, he advises to use the previously discussed FELW estimation procedure to the detrended series in levels  $\widetilde{y}_t = y_t - \hat{\alpha} - \hat{\beta}t$  where  $\hat{y}_t = y_t - \hat{\alpha} - \hat{\beta}t$  are the OLS residuals and  $\hat{\alpha}(d) = \omega(d)T^{-1}\sum \hat{y}_t + [1 - \omega(d)]y_1$ . Notice that  $\hat{\alpha}(d)$  is a weighted average of the two alternative estimators of  $\alpha$  mentioned above where  $\omega(d)$  is a smooth (twice continuously differentiable) weight function such that  $\omega(d) = 1$  for  $d \in (0, 0.5)$ .<sup>11</sup> By means of this alternative detrending procedure, the difference between  $\Delta^d \widetilde{y}_t$  and  $\varepsilon_t$  becomes negligible for any value of  $d \in (0, 1]$ . Therefore, if one considers the regression model

$$\widetilde{\Delta y}_t = \psi \widetilde{s}_{t-1}(\hat{d}_T) + e_t, \quad (10)$$

---

<sup>11</sup>An example of  $\omega(d)$  for  $d \in (0.5, 1)$  is  $(1/2)[1 - \cos \pi d]$ .

where the input of  $\widetilde{s}_{t-1}(\widehat{d}_T)$  is a  $T^\kappa$ -consistent estimator of  $d^*$ , for some  $d^* < 0.5$  with  $\kappa > 0$ , having used as residuals  $(\widetilde{y}_t)$  the ones obtained from an OLS regression of  $y_t$  on  $\mu(t)$ , the asymptotic properties of the EFDF(0) test for testing  $\psi = 0$  in (10) are identical to those stated in Theorem 2. Likewise, under fixed alternatives, a similar result holds for cases where  $d \in (0.5, 1)$ , this time using Shimotsu's (2006a) residuals and  $\widetilde{d}_{0T}$ , as an alternative estimator of the corresponding input value of the regressor  $\widetilde{s}_{t-1}(\cdot)$ .

#### 4. DETERMINISTIC COMPONENTS WITH BREAKS

In this section, we extend the EFDF test to cover the case where the deterministic component,  $\mu(t)$ , of the time series  $y_t$  in (1) is possibly subject to structural breaks, denoted hereafter as  $\mu_B(t)$ . One possibility is to consider breaks both under the null and the alternative hypotheses discussed in sections 2.1 and 2.2. In this case, similar two-stage procedures to those described in section 3 could be applied.<sup>12</sup> However, it is well known in the statistical literature that some features of long-range dependence (LRD) can be generated by either the process being  $I(d)$  with smooth deterministic components or by an  $I(0)$  process subject to breaks; see, e.g., Bhattacharya *et al.* (1983), Mikosch and Starica (2004), and Berkes *et al.* (2006). Indeed, these studies show that conventional statistics designed to detect long range dependence behave similarly under weak dependence with change-points.<sup>13</sup>

For this reason we focus here on the pure distinction between these two *alternative* models that can account for the observed strong persistence of  $y_t$ : (i)  $u_t$  is an  $I(d)$  process, with  $d \in (0, 1)$  and  $\mu(t)$  is smooth, and (ii)  $u_t$  is a short-memory  $I(0)$  process and  $\mu(t)$  is subject to breaks. The EFDF approach, where one of the hypothesis encompasses the other, cannot directly accommodate these two types of models. This is so since the  $I(d)$  hypothesis clearly nests the  $I(0)$  one, but then the  $\mu(t)$  component cannot nest  $\mu_B(t)$  component at the same time. We therefore follow

---

<sup>12</sup>For tests of  $I(1)$  vs.  $I(0)$  with breaks under both the null and the alternative, see Banerjee and Urga (2005), and Kim and Perron (2006). Extensions of these tests to a fractional setup can be found in DGM (2005), Mayoral (2006), and Shimotsu (2006b).

<sup>13</sup>More recently, a similar issue has re-emerged in the econometric literature dealing with financial data. For example, Ding and Granger (1996), and Mikosch and Starica (2004) claim that the stochastic components of both the absolute and the squared returns of financial prices (bonds, exchange rates, options, etc.) are  $I(0)$  and explain the evidence about LRD found in the data as spuriously induced by structural breaks in the parameters of the deterministic components over different subsamples due to significant events, such as the Great Depression of 1929 or the oil-price shocks in the 1970s. On the contrary, Lobato and Savin (1998) conclude that the LRD evidence found in the squared returns is genuine and, thus, not a spurious feature of structural breaks.

a comprehensive model approach, whereby non-nested models are tested within an artificially constructed general model that includes them as special cases. This approach was advocated by Atkinson (1970) and later taken up under a different guise by Davidson and MacKinnon (1981) in developing their J-test. In effect, let us think of two alternative models for  $y_t$ , denoted as  $M1$  and  $M2$ , respectively, which are defined as follows

$$M1 : y_t = \mu_B^i(t) + \varepsilon_t, \quad (11)$$

and

$$M2 : y_t = \mu(t) + \Delta^{-d} \varepsilon_t 1_{t>0}, \text{ with } d \in (0, 1), \quad (12)$$

where  $\mu_B^i(t)$  is a linear deterministic trend function that may contain breaks at known or unknown dates (in principle, just a single break at date  $T_B$  would be considered) while  $\mu(t)$  does not contain breaks. In line with Perron (1989), three definitions of  $\mu_B^i(t)$ ,  $i \in \{A, B, C\}$  will be considered,

$$\text{Case A : } \mu_B^A(t) = \mu_0 + (\mu_1 - \mu_0)DU_t(\omega_B), \quad (13)$$

$$\text{Case B : } \mu_B^B(t) = \mu_0 + \beta_0 t + (\beta_1 - \beta_0)DT_t^*(\omega_B), \quad (14)$$

$$\text{Case C : } \mu_B^C(t) = \mu_0 + \beta_0 t + (\mu_1 - \mu_0)DU_t(\omega_B) + (\beta_1 - \beta_0)DT_t(\omega_B). \quad (15)$$

Case A corresponds to the *crash* hypothesis, case B to the *changing growth* hypothesis and case C to a combination of both. The dummy variables are defined as follows:  $DU_t(\omega_B) = 1_{(T_B+1 \leq t \leq T)}$ ,  $DT_t^*(\omega_B) = (t - T_B)1_{(T_B+1 \leq t \leq T)}$  and  $DT_t(\omega_B) = t1_{(T_B+1 \leq t \leq T)}$  where  $\omega_B = T_B/T$  is a fixed value belonging to the subset of the interval  $(0, 1)$  that describes the relative location of the break in the sample.

Then, noticing that  $M2$  can be rewritten as

$$M2 : y_t = y_t - \Delta^d [y_t - \mu(t)] + \varepsilon_t, \quad (16)$$

one could follow Davidson and MacKinnon (1981) in considering the following linear combinations of  $M1$  and  $M2$

$$y_t = (1 - \zeta)\mu_B^i(t) + \zeta\{y_t - \Delta^d [y_t - \mu(t)]\} + \varepsilon_t, \quad (17)$$

or

$$y_t = (1 - \zeta)\{y_t - \Delta^d [y_t - \mu(t)]\} + \zeta\mu_B^i(t) + \varepsilon_t, \quad (18)$$

so that two  $J$ - tests can be applied, depending on whether  $M1$  or  $M2$  is considered to be the null hypothesis. In the case where  $M1$  is taken to be  $H_0$  and  $M2$  to be  $H_1$ , the unknown parameters in  $\{y_t - \Delta^d[y_t - \mu(t)]\}$  are not identified under  $H_0$  since  $\zeta = 0$ . A solution of this problem is to replace the term  $\{y_t - \Delta^d[y_t - \mu(t)]\}$  in (17) by  $\{y_t - \Delta^{\hat{d}_T}[y_t - \hat{\mu}(t)]\}$ , where  $\hat{d}_T$  and  $\hat{\mu}(t)$  are consistent under  $H_1$ , e.g., using the, Shimotsu's (2006a) estimation procedure described in section 3. Hence, the following regression can be estimated

$$y_t = \mu_B^{*i}(t) + \zeta \widetilde{\nu}_{t-1} + \varepsilon_t, \quad (19)$$

where  $\widetilde{\nu}_{t-1} = \{y_t - \Delta^{\hat{d}_T}[y_t - \hat{\mu}(t)]\}$  and  $\mu_B^{*i}(t) = (1 - \zeta)\mu_B^i(t)$ . Under  $H_0$ , it follows that  $\zeta = 0$ , and this hypothesis can be tested using a  $t$ -test on the coefficient of  $\widetilde{\nu}_{t-1}$ ,  $t_\zeta$ . We will denote this test as the EFDF(B) test.

Conversely, if one chooses  $M2$  to be  $H_0$  and  $M1$  to be  $H_1$ , the corresponding regression model becomes

$$\Delta^d y_t = \Delta^d \mu^*(t) - \zeta \widetilde{\nu}_{t-1}(d) + \varepsilon_t, \quad (20)$$

where now  $\widetilde{\nu}_{t-1}(d) = \{y_t - \Delta^d y_t - \hat{\mu}_B^i(t)\}$ ,  $\mu^*(t) = (1 - \zeta)\mu(t)$  where  $d$  is taken to be known under the null, and  $\hat{\mu}_B^i(t)$  is estimated in a preliminary regression under the alternative of  $I(0)$  cum breaks. Again, under  $H_0$ , we have that  $\zeta = 0$ , and a  $t$ -test,  $t_\zeta$ , could be used to test for this hypothesis.

For simplicity, we have operated above as if the break dates were known in regressions (17) and (18). The more realistic case of unknown breaks when  $y_t$  is  $I(0)$ , are under current investigation following Bai's (1997) or Bai and Perron's (1998) procedures.

Finally, notice that, because non-nested hypothesis tests are designed as specification tests, rather than as procedures for choosing among competing models, it is not at all surprising that sometimes they do not lead us to choose one model over the other. If we would simply wish to choose the *best* model between  $M1$  and  $M2$ , one could use some information criteria that help to discriminate between them. This approach is also under current investigation by the authors.

## 5. BREAKS IN THE LONG-MEMORY PARAMETER

Granger and Ding (1996) were the first to analyze the consequences of having a variable memory parameter  $d$ .<sup>14</sup> They consider two possible scenarios: (i)  $d_t$  is a stochastic process, e.g., an  $AR(1)$

---

<sup>14</sup>The problem of detecting a change in the persistence of a process is usually tackled in the context of AR processes within the  $I(0)/I(1)$  framework (see, eg., Buseti and Taylor, 2004). Nevertheless, as argued above, this

process with mean  $\bar{d}$ , and (ii)  $d_t$  switches between two regimes, e.g.,  $y_t = \lambda_t x_{1t} + (1 - \lambda_t)x_{2t}$ , with  $x_{1t} \sim I(d_1)$ ,  $x_{2t} \sim I(d_2)$  and  $\lambda_t$  following a 0-1 Markov switching process. Since this paper is focused on testing, we consider a different setup. The memory parameter  $d$  can take two values,  $d_1$  in a first given proportion of the sample and  $d_2$  in the remaining proportion.

Both stationary and non-stationary fractional roots are considered. Although it is not difficult to generalize the analysis to allow for breaks in the deterministic components as well as short-term correlation in the disturbance terms, for simplicity we will focus in the sequel only on the case where the error terms are *i.i.d.* and no deterministic terms are present. More specifically, we assume that  $y_t$  is generated as

$$(1 - L)^{d_0 + \theta D_t(\omega_B)} y_t = \varepsilon_t 1_{t > 0}, \quad (21)$$

so that  $y_t$  is a zero-mean integrated process (with an integer or fractional integration order), that can be either stationary or nonstationary. The order of integration of  $y_t$  is allowed to change along the sample at time  $T_B$ , with the dummy variable  $D_t(\omega_B)$  taking a value equal to 1 if  $\omega_B T < t$  and zero otherwise. Then, the process  $y_t$  is  $I(d + \theta)$  until  $T_B$  and  $I(d)$  after  $T_B$ , where  $\theta$  can be either larger or smaller than zero.

Under  $H_0$ , no change in persistence occurs and therefore  $H_0 : \theta = 0$ , whereas, under  $H_1$ , a change in persistence occurs at time  $T_B$ , that is  $H_1 : \theta < 0$  or  $H'_1 : \theta > 0$ , where the first (second) case corresponds to an increase (decrease) in persistence after  $T_B$ .

Since, to our knowledge, the LM tests have not been used so far to test this type of hypothesis, we start by deriving such a test in the present case. Under gaussianity, recall that Tanaka's (1999) time-domain version of the LM statistic for testing  $H_0 : d = d_0$  vs.  $H_1 : d \neq d_0$  uses the the log-likelihood

$$L(\theta, \sigma^2, \omega_B) = -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^T \left\{ (1 - L)^{d_0 + \theta D_t(\omega_B)} y_t \right\}^2 \quad (22)$$

Thus, an LM test for  $H_0 : \theta = 0$  vs.  $H_1 : \theta \neq 0$  rejects  $H_0$  for large values of,

$$LM_T = \frac{\partial L(\theta, \sigma^2, \omega_B)}{\partial \theta} \Bigg|_{H_0: \theta=0, \sigma^2=\hat{\sigma}^2} = -\frac{1}{\hat{\sigma}^2} \sum_{t=1}^T \left( \left\{ \log(1 - L) \times D_t(\omega_B) \Delta^{d_0} y_t \right\} \Delta^{d_0} y_t, \right.$$

where the estimated variance is  $\hat{\sigma}^2 = T^{-1} \sum (\Delta^{d_0} y_t)^2$ .

framework can be too narrow in many empirical applications.



Since  $\log(1 - L) = -(L + L^2/2 + L^3/3 + \dots)$ , and  $D_t(\omega_B) = 0$  for  $t > T_B$ , then

$$LM_T = \frac{1}{\hat{\sigma}^2} \sum_{t=2}^{T_B} \left( \sum_{k=1}^{t-1} \frac{1}{k} \hat{\varepsilon}_{t-k} \right) \hat{\varepsilon}_t = T \sum_{t=2}^{T_B} \frac{\left( \sum_{k=1}^{t-1} \frac{1}{k} \hat{\varepsilon}_{t-k} \hat{\varepsilon}_t \right)}{\sum_{t=1}^T \hat{\varepsilon}_t^2} = T \sum_{k=1}^{T_B-1} \frac{1}{k} \hat{\rho}_k^*(\hat{\varepsilon}_t), \quad (23)$$

where  $\hat{\rho}_k^*(\hat{\varepsilon}_t) = \sum_{t=k+1}^{T_B} \hat{\varepsilon}_{t-k} \hat{\varepsilon}_t / \sum_{t=1}^T \hat{\varepsilon}_t^2$ . Notice that in finite samples  $\hat{\rho}_k^*(\hat{\varepsilon}_t)$  is not identical to the  $k$ -th autocorrelation of residuals since in order to compute the numerator, only observations previous to the break are considered whereas all observations are employed to compute the denominator. This difference vanishes asymptotically. The following theorem describes the asymptotic properties of the test under local alternatives when the break date is known.

**Theorem 5** *Under the hypothesis of  $\theta = \delta/\sqrt{T}$ , for a known value of  $T_B$  and a fixed  $\delta$  it holds that, as  $T \rightarrow \infty$ ,*

$$T^{-1/2} S_T(\omega_B) = \sqrt{T} \sum_{k=1}^{T_B-1} \frac{1}{k} \hat{\rho}_k^*(\hat{\varepsilon}_t) \xrightarrow{w} N\left(\frac{\pi^2}{6} \delta, \frac{\pi^2}{6} \omega_B\right). \quad (24)$$

Note that, since  $\omega_B < 1$ , the variance of this distribution is smaller than the variance in the case where no break occurs. This reflects the fact that only a fraction of the data is employed but the data is divided by  $\sqrt{T}$ . Along the lines of Tanaka (1999), it can also be shown that the test statistic proposed in (23) is locally optimal.

An EFDF test, denoted as EFDF(Bd), can also be constructed for this case. Following the derivations in section 2, one could consider the following maintained hypothesis

$$\Delta^{d_0} y_t = [1 - \Delta^{\theta D_t(\omega_B)}] \Delta^{d_0} y_t + \varepsilon_t, \quad (25)$$

which can be expressed in EFDF format as

$$\Delta^{d_0} y_t = \vartheta \left[ \frac{1 - \Delta^\theta}{\theta} \right] \Delta^{d_0} y_t D_t(\omega_B) + \varepsilon_t, \quad (26)$$

where  $\vartheta = \theta$ . Thus, conditional upon the choice of  $\omega_B$ , the EFDF(Bd) test would test  $H_0 : \theta = 0$  against  $H_1 : \theta \neq 0$  by means of a two-sided test based on the t-ratio,  $t_\vartheta$ , which is estimated with observations  $1, \dots, T\omega_B$ , and whose asymptotic distribution, under the null, would be  $N(0, 1)$  and, under local alternatives, satisfies Theorem 5. To construct the regressor in (26), the first step is to apply the "deadstart" filter  $\theta^{-1}[\Delta^{d_0} - \Delta^{d_0+\theta}]$  to  $\{y_t\}_{t=1}^T$ ; next, the resulting filtered series is truncated to the first subsample by means of the dummy variable,  $D_t(\omega_B)$ . If  $\theta$  is taken to be unknown, one could use a  $T^\kappa$ -consistent estimator of  $d$  from the first subsample and subtract it from  $d_0$  using any of the estimation procedures discussed above.

This way of testing breaks in the long-memory parameter opens the possibility of testing a wide set of other alternative explanations for time varying long-memory behaviour. For instance, inspired by Granger and Ding (1996), the changes in  $d$  could be triggered by a variable  $W_t$  that characterizes different regimes of the economy. More concretely, we are interested on testing  $H_0 : y_t \sim I(d)$  versus  $H_A : y_t \sim I(d)$  when  $W_{t-1} \leq r$  and  $I(d + \theta)$  when  $W_{t-1} > r$ . Substituting the structural break dummy  $D_t(\omega_B)$  by the threshold dummy  $I(W_{t-1} > r)$  in (26) and running the regression

$$\Delta^{d_0} y_t = \vartheta \left[ \frac{1 - \Delta^\theta}{\theta} \right] \Delta^{d_0} y_t I(W_{t-1} > r) + \varepsilon_t, \quad (27)$$

where  $\vartheta = \theta$ , the corresponding EFDF test for threshold long memory, denoted by EFDF(Td), is a simple two-sided test based on the t-ratio,  $t_\vartheta$ , whose asymptotic distribution, under the null, would be  $N(0, 1)$  assuming  $r$  is known (e.g.,  $r = 0$ ). Further issues stemming from an unknown  $r$  are beyond the scope of this paper and being under current investigation by the authors.

## 6. ALLOWING FOR SERIAL CORRELATION

Lastly, we generalize the DGPs considered in Section 2 to the case where  $u_t$  follows an stationary linear  $AR(p)$  process, namely,  $\Phi_p(L)u_t = \varepsilon_t 1_{t>0}$  with  $\Phi_p(L) = 1 - \phi_1 L - \dots - \phi_p L^p$  and  $\Phi_p(z) \neq 0$  for  $|z| \leq 1$ . This motivates the following nonlinear regression model

$$\Delta^{d_0} y_t = \varphi[\Phi_p(L)x_{t-1}(d)] + \sum_{j=1}^p \phi_j \Delta^{d_0} y_{t-j} + \varepsilon_t, \quad (28)$$

where  $x(\cdot) = z(\cdot)$ ,  $s(\cdot)$ , depending on whether we consider the EFDF(1) or EFDF(0) tests. The new model is similar to (3) and (4), except for the inclusion of the lags of  $\Delta^{d_0} y_t$  and for the filter  $\Phi_p(L)$  in the regressor  $x_{t-1}(d)$ . Estimation of this model is cumbersome due to the nonlinearity in the parameters  $\varphi$  and  $\Phi = (\phi_1, \dots, \phi_p)$ . Compared with the *i.i.d.* case, LV (2007) claim that a practical problem arises because the vector  $\Phi$  is unknown and therefore the regressor  $[\Phi_p(L)x_{t-1}(d)]$  is unfeasible. For this reason, they recommend to apply the following two-step procedure that allows one to obtain efficient tests also with autocorrelated errors.

Assuming, for simplicity, that  $\mu(t) \equiv 0$  (or known),<sup>15</sup> in the first step, the coefficients of  $\Phi_p(L)$  are estimated (under  $H_1$ ) by OLS in the equation  $\Delta^{\hat{d}_T} y_t = \sum_{j=1}^p \phi_j \Delta^{\hat{d}_T} y_{t-j} + a_t$ , where  $\hat{d}_T$  satisfies the conditions stated in Theorems 1 and 2. The estimator of  $\Phi_p(L)$  is consistent with a

---

<sup>15</sup>For the case where the coefficients of  $\mu(t)$  are considered to be unknown, a similar procedure as that described in section 2.1 can be implemented and efficient tests will still be obtained.

convergence rate which depends on the rate  $\kappa$ . The second step consists of estimating by OLS the equation  $\Delta^{d_0}y_t = \varphi[\widehat{\Phi}_p(L)x_{t-1}(\widehat{d}_T)] + \sum_{j=1}^p \phi_j \Delta^{d_0}y_{t-j} + v_t$ , where  $\widehat{\Phi}_p(L)$  is the estimator from the first step, and  $\widehat{d}_T$  denotes the same estimated input used in that step as well. As LV (2007, Theorem 2) have shown, for the  $I(1)$  vs.  $I(d)$  case, the  $t_\varphi$  statistic in this augmented regression is still both normally distributed and locally optimal, but a similar argument applies to the  $I(0)$  vs.  $I(d)$  case. The tests will be denoted as AEFDF( $i$ ),  $i = 1, 0$ , (augmented EFDF) tests in the sequel.

However, in DGM (2007) we claim that a feasible single-step procedure in the case of the AEFDF(1) test can also be applied with the same properties. In effect, under  $H_1$ , the process would be  $\Phi_p(L)\Delta^d y_t = \varepsilon_t$ , so that adding and subtracting the process under  $H_0$ ,  $\Phi_p(L)\Delta y_t$ , it becomes

$$\Delta y_t = \varphi[\Phi_p(L)z_{t-1}(d)] + [1 - \Phi_p(L)]\Delta y_t + \varepsilon_t. \quad (29)$$

The one-step method we propose is based on the following decomposition of the lag polynomial  $\Phi_p(L)$

$$\Phi_p(L) = \Phi_p(1) + \frac{1}{\Delta^{d-1} - 1} \Phi_p^*(L), \quad (30)$$

where the polynomial  $\Phi_p^*(L)$  is defined by equating (30) to the standard polynomial decomposition

$$\Phi_p(L) = \Phi_p(1) + \Delta \widetilde{\Phi}_p(L). \quad (31)$$

Hence,

$$\Phi_p^*(L) = (\Delta^d - \Delta) \widetilde{\Phi}_p(L) = \Delta^d \widetilde{\Phi}_p(L) - [\Phi_p(L) - \Phi_p(1)]. \quad (32)$$

Substitution of (32) into (29), using (30) and noticing that  $\varphi = d - 1$  and  $z_{t-1}(d) = \frac{\Delta^{d-1}-1}{1-d} \Delta y_t$ , yields after some simple algebra

$$\Delta y_t = \varphi \frac{\Phi_p(1)}{C} z_{t-1}(d) - \frac{\widetilde{\Phi}_p(L)}{C} [\Delta^d - 1] \Delta y_t - \frac{\widetilde{\Phi}_p(L) - \widetilde{\Phi}_p(0)}{C} \Delta y_t + \frac{1}{C} \varepsilon_t. \quad (33)$$

where  $C = \Phi_p(1) + \widetilde{\Phi}_p(0)$ . Notice that the second and third regressors are predetermined since  $(\Delta^d - 1)$  and  $(\widetilde{\Phi}_p(L) - \widetilde{\Phi}_p(0))$  do not include contemporaneous values of  $\Delta y_t$ . Hence, in this model the AEFDF(1) test can be done in a single step, using a t-test on the coefficient of  $z_{t-1}(\widehat{d}_T)$ . For example, in the case of an AR(1) disturbance, i.e.,  $\Phi_1(L) = 1 - \phi L$ , we have that  $\Phi_1(1) = 1 - \phi$  and  $\widetilde{\Phi}_1(L) = \widetilde{\Phi}_1(0) = \phi$ , so that (33) becomes

$$\Delta y_t = \varphi(1 - \phi) z_{t-1}(d) - \phi[\Delta^d - 1] \Delta y_t + \varepsilon_t. \quad (34)$$

A similar procedure can be used to obtain a single-step testing procedure in the case of the AEFDF(0) test. In effect, adding and subtracting the process under  $H_0$ ,  $\Phi_p(L)y_t$ , to the process under  $H_1$ , yields

$$y_t = \psi[\Phi_p(L)s_{t-1}(d)] + [1 - \Phi_p(L)]y_t + \varepsilon_t. \quad (35)$$

Then, using the decompositions

$$\Phi_p(L) = \Phi_p(0) + \frac{1}{\Delta^d - 1} \Phi_p^*(L), \quad (36)$$

$$\Phi_p(L) = \Phi_p(0) + L\Phi_{p-1}(L), \quad (37)$$

and operating, yields

$$y_t = \psi s_{t-1}(d) - \frac{\Phi_{p-1}(L)}{\Phi_p(0)} \Delta^d y_{t-1} + \frac{1}{\Phi_p(0)} \varepsilon_t, \quad (38)$$

which in the case of an AR(1) disturbance, i.e.,  $\Phi_1(L) = 1 - \phi L$ , becomes

$$y_t = \psi s_{t-1}(d) + \phi \Delta^d y_{t-1} + \varepsilon_t. \quad (39)$$

Following LV (2007) one can show that the asymptotic properties of both single-step AEFDF(i=1,2)) tests are identical to those stated in Theorems 1 and 2 above, except that, under local alternatives ( $d = 1 - \gamma/\sqrt{T}$  for AEFDF(1) and  $d = \gamma/\sqrt{T}$  for AEFDF(0), with  $\gamma > 0$ ), we have that  $t_\varphi(\hat{d}_T) \xrightarrow{w} N(-\gamma\omega, 1)$  and  $t_\psi(\hat{d}_T) \xrightarrow{w} N(\gamma\omega, 1)$  where

$$\omega^2 = \frac{\pi^2}{6} - \varkappa' \Psi^{-1} \varkappa, \quad (40)$$

such that  $\varkappa = (\varkappa_1, \dots, \varkappa_p)'$  with  $\varkappa_k = \sum_{j=k}^{\infty} j^{-1} c_{j-k}$ ,  $k = 1, \dots, p$ ,  $c_j$ 's are the coefficients of  $L^j$  in the expansion of  $1/\Phi(L)$ , and  $\Psi = [\Psi_{k,j}]$ ,  $\Psi_{k,j} = \sum_{t=0}^{\infty} c_t c_{t+|k-j|}$ ,  $k, j = 1, \dots, p$ , denotes the Fisher information matrix for  $\Phi(L)$  under Gaussianity. Note that  $\omega^2$  is identical to the drift of the limiting distribution of the LM test under local alternatives (see Tanaka, 1999). Notice also that the use of semiparametric estimators for  $d$  is very convenient here, since one can be agnostic about a parametric specification of the autocorrelation in the error terms when estimating the input value of  $d$ . Although it has not been proved yet, we conjecture that the single step procedure can be generalized to deal with ARMA processes, rather than AR ones by increasing the number of regressors in (33) or (38) at a certain rate, along the lines of DGM (2002, Theorem 7).

## 7. CONCLUDING REMARKS

Long-memory processes have become a very attractive research topic in econometrics during the last few years, due both to their flexibility and realistic microfoundations. Indeed, they received a lot of attention from the theoretical viewpoint but, in our modest opinion, so far this has not been sufficiently reflected in empirical work. There must be several reasons for this disconnection. We believe that one of them is that empirical researchers have found difficulties in implementing many of those theoretical results. Thus, our main goal in this paper has been to frame the long-memory testing procedures in a setup somewhat equivalent to the nowadays familiar unit roots testing approach (à la Dickey-Fuller): t-statistics in simple time-domain regressions, with known conventional asymptotic distributions and easy to implement using standard econometrics softwares. Although our illustrations have focused on univariate processes, extensions to fully-fledged multivariate models should not be hard to derive. For example, a first try at applying the Wald test principles to the reduced-rank analysis in a system of  $I(1)$  processes with fractional cointegrating relationships of order  $(1 - b)$ ,  $b \in [0, 0.5)$ , can be found in Avarucci and Velasco (2007).

## REFERENCES

- Akonom, J. and C. Gouriéroux (1987), “A functional limit theorem for fractional processes,” CEPREMAP, mimeo.
- Abadir, K., Distaso, W. and L. Giraitis (2005), “Semiparametric estimation for trending I(d) and related processes,” Imperial College, mimeo.
- Atkinson, A. (1970), “A Method for discriminating between models (with discussion),” *Journal of the Royal Statistical Society*, B, B32, 323-353.
- Avarucci, M. and C. Velasco (2007), “A Wald test for the cointegration rank in nonstationary fractional systems,” Universidad Carlos III, mimeo.
- Bai, J. (1997), “Estimation of a change point in a multiple regression models,” *Review of Economic and Statistics*, 79, 551-563.
- Bai, J. and P. Perron (1998), “Estimating and testing linear models with multiple structural changes,” *Econometrica*, 66, 47-78.
- Banerjee, A. and G. Urga (2005), “Modelling structural breaks, long memory and stock market volatility: an overview,” *Journal of Econometrics*, 129, 1-34.
- Berkes, I., Hovarth, L., Kokoszka, P., and Q-M. Shao (2006), “On discriminating between long-range dependence and changes in mean,” *Annals of Statistics*, 34, 1140-1165.
- Bhattacharya, R.N, Gupta, V.K. and E. Waymire (1983), “The Hurst effect under trends,” *Journal of Applied Probability*, 20, 649-662.
- Breitung, J. and U. Hassler (2002), “Inference on the cointegrated rank of in fractionally integrated processes,” *Journal of Econometrics*, 110, 167-185.
- Busetti, F. and A.M.R. Taylor (2004), “Tests of stationarity against a change in persistence,” *Journal of Econometrics*, 123, 33-66.
- Davidson, R., and J.G. Mackinnon (1981), “Several tests for model specification in the presence of alternative hypothesis,” *Econometrica*, 49, 781-793.
- Diebold, F. and A.Inoue (2001), “Long memory and regime switching,” *Journal of Econometrics*, 105, 131-159.
- Ding, Z. and C.W.J. Granger (1996), “Varieties of long memory models,” *Journal of Econometrics*, 73, 61-77.
- Dolado, J., Gonzalo, J. and L. Mayoral (2002), “A fractional Dickey-Fuller test for unit roots,” *Econometrica*, 70, 1963-2006.

Dolado, J., Gonzalo, J. and L. Mayoral (2005), "Structural breaks vs. long memory: What is what?" Universidad Carlos III, mimeo.

Dolado, J., Gonzalo, J. and L. Mayoral (2007), "Wald tests of  $I(1)$  against  $I(d)$  alternatives: Some new properties and an extension to processes with trending components," Universidad Carlos III, mimeo.

Engle, R.F. (1984) "Wald, likelihood ratio and Lagrange multiplier tests in econometrics," in Z. Griliches and R. Intriligator (eds), *Handbook of Econometrics*, vol II, 75-826, Amsterdam: North Holland.

Gil-Alaña, L. A. and P. Robinson (1997), "Testing unit roots and other statistical hypothesis in macroeconomic time series," *Journal of Econometrics*, 80, 241-268.

Gonzalo, J. and T. Lee (1998), "Pitfalls in testing for long-run relationships," *Journal of Econometrics*, 86, 129-154.

Granger, C.W.J. (1980), "Long memory relationships and the aggregation of dynamic models," *Journal of Econometrics*, 14, 227-238.

Granger, C. W. J. (1986), "Developments in the study of cointegrated economic variables," *Oxford Bulletin of Economics and Statistics*, 48, 213-228.

Granger, C.W.J. and Z. Ding (1996), "Varieties of long memory models," *Journal of Econometrics* 73, 61-77.

Hosking, J.R.M., (1996), "Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series," *Journal of Econometrics*, Vol. 73, (1), 261-284.

Johansen, S. (2005), "A representation theory for a class of vector autoregressive models for fractional processes," University of Copenhagen, mimeo.

Kim, D. and P. Perron (2006), "Unit roots tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses," Boston University, mimeo.

Lo, A.W. and J.G. Haubrich (2001), "The sources and nature of long-term dependence in the business cycle," *Economic Review*, 37, 15-30.

Lobato, I. and G. Savin (1998), "Real and spurious long memory properties of stock market data," *Journal of Business and Economic Statistics*, 16, 261-268.

Lobato, I. and C. Velasco (2006), "Optimal fractional Dickey-Fuller tests for unit roots," *Econometrics Journal*, 9, 492-510.

Lobato, I. and C. Velasco (2007), "Efficient Wald tests for fractional unit roots," *Econometrica*, 75, 575-589.

Mayoral, L. (2006), "Testing for fractional integration versus short memory with trends and structural breaks," Universitat Pompeu Fabra, mimeo.

Michelacci, C. and P. Zaffaroni, P.(2000), "Fractional beta convergence," *Journal of Monetary Economics*, 45, 129-153.

Mikosch, T. and C. Starica (2004), "Nonstationarities in financial time series, long range dependence and the IGARCH model," *Review of Economics and Statistics*, 86, 378-390.

Perron, P. (1989), "The great crash, the oil price shock and the unit root hypothesis," *Econometrica*, 57, 1361-1401.

Parke, W. R. (1999), "What is fractional integration?" *Review of Economics and Statistics*, 81, 632-638.

Robinson, P.M. (1978), "Statistical inference for a random coefficient autoregressive model," *Scandinavian Journal of Statistics*, 5, 163-168.

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, 1420-1437.

Robinson, P.M. and D. Marinucci (2001), "Narrow-band analysis of nonstationary processes," *Annals of Statistics*, 29, 947-976.

Shimotsu, K. and P. C. B. Phillips (2005), " Exact local Whittle estimation of fractional integration, " *Annals of Statistics*, 33, 1890-1933.

Shimotsu, K. (2006a), "Exact local Whittle estimation of fractional integration with unknown mean and time trend," Queen's Economics Dept. Working Paper 1061.

Shimotsu, K. (2006b), "Simple (but effective) tests of long memory versus structural breaks," Queen's Economics Dept. Working Paper no 1101.

Tanaka, K. (1999), "The nonstationary fractional unit root," *Econometric Theory*, 15, 249-264.

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



## FIGURES

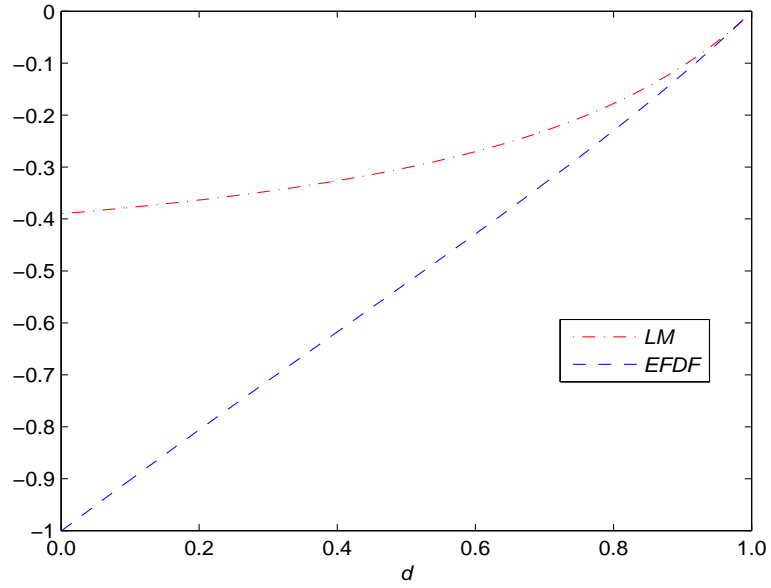


FIG. 1: Non-centrality parameters of EFDF(1) and LM tests

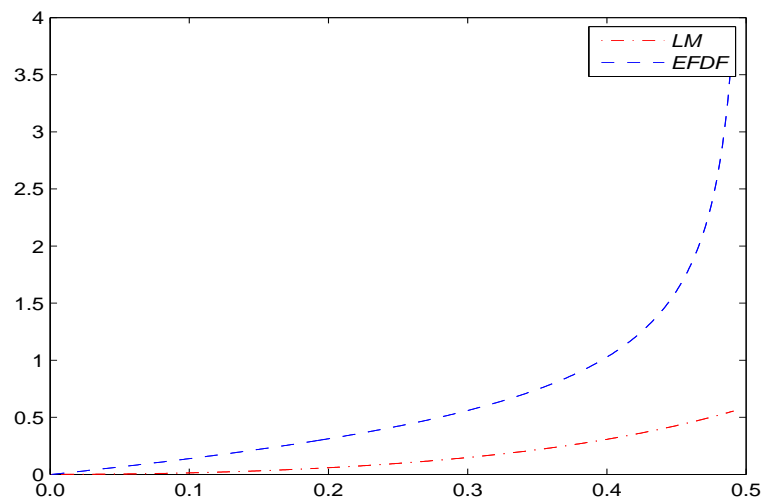


FIG. 2: Non-centrality parameters of EFDF(0) and LM tests