LM tests for joint breaks in the dynamics and level of a long-memory time series *

Juan J. Dolado

European University Institute

Heiko Rachinger

Universitat de les Illes Balears

Carlos Velasco†

Universidad Carlos III de Madrid

This version: January 25, 2018.

Abstract.

We consider two easily implementable versions of the Lagrange Multiplier (LM) test for joint breaks in the long memory parameter, the short-run dynamics and the level of a fractionally integrated time-series process. These are: (i) a conventional LM test with parameters estimated under the null of parameter stability, and (ii) a regression-based test, labeled as LMW-type test, which also uses some information under the alternative (in the spirit of a Wald test). We derive their asymptotic distributions for known and unknown breakpoints, both under the null and a local break hypothesis. Monte Carlo simulations show that the proposed tests behave satisfactorily in finite samples. We finally provide an empirical application of the proposed tests to characterize shifts in the long-memory properties of forward discount rates for several G7 countries, detecting the source(s) of these breaks.

JEL Classification: C13, C22

MSC 2010: Primary: 62M10; *Secondary:* 60G22.

Keywords: Structural Breaks, Fractional Integration, Long Memory, LM Test

---

*We are very grateful to the Editor and two anonymous referees for many useful comments that helped improve the paper a lot, and to Eric Zivot for making his exchange-rates dataset available to us. We also thank Juan-Carlos Escanciano, Robinson Kruse, Jesús Gonzalo, Uwe Hassler, Laura Mayoral, Peter Robinson, Abderrahim Taamouti and participants in several seminars and workshops for helpful suggestions on earlier drafts of the paper. Financial support from the Spanish Ministerio de Economía y Competitividad (grants ECO2016-78652 and ECO2017-86009-P, ECO2014-57007p and MDM 2014-0431), and Comunidad de Madrid, MadEco-CM (S2015/HUM-3444) is gratefully acknowledged. The usual disclaimer applies.

†Corresponding author: Carlos Velasco, Department of Economics, Universidad Carlos III de Madrid, Getafe, 28903 Madrid, Spain; E-Mail: carlos.velasco@uc3m.es

1 Introduction

The confoundness issues raised in Diebold and Inoue (2001) and Granger and Hyung (2004) have spurred controversy on the origin of long-memory features in some time series. On the one hand, there is the issue of whether these features are truly driven by fractionally integrated processes of order d, FI(d), or are spuriously generated by level shifts in short-memory time series (see, e.g., Lobato and Savin (1998), Mikosch and Starica (2004), and Perron and Qu (2010)). On the other hand, it has also been pointed out that stochastic processes with breaks in d could be misleadingly interpreted as having breaks in the level, μ (see, e.g., McCloskey (2010) and Shao (2011)).

This debate has led to two strands in the literature on this topic (see Aue and Horváth (2011) for a nice overview). The first one has focused on testing for breaks in d. Motivated by the popular rationalization of FI(d) processes in terms of aggregation arguments (Robinson (1978) and Granger (1980)), it has been argued that changes in the distribution of the persistence parameters of the disaggregated components of many macro and financial variables may be due to regime shifts in monetary policy, financial regulation or industrial and labour policies. As a result, the long-memory properties of relevant aggregates (e.g., inflation, unemployment, GDP, squared financial returns, etc.) are likely to have experienced shifts over relevant subsamples (see, e.g. Gadea and Mayoral, 2005, for empirical evidence on this issue). Accordingly, several tests in both the time and frequency domains have been proposed to test the null of a constant value of the memory parameter d against the alternative of a break at known or unknown dates; see, *inter alia*, Beran and Terrin (1996) and (1999), Sibbertsen and Kruse (2009), Hassler and Scheithauer (2011) and Yamaguchi (2011). Forerunners of this line of research are the approaches proposed by Kim et al. (2002), Busetti and Taylor (2004), and Harvey et al. (2009) to test for changes in time series from being $I(0)$ to being $I(1)$ or viceversa. Multiple changes are tackled in Leybourne et al. (2007) and Kejriwal et al. (2013). In parallel, there has been another strand of the literature which has focused on developing tests for breaks in the level or other deterministic components of a time series with stationary long-memory disturbances, but without allowing for breaks in d; see, e.g., Hidalgo and Robinson (1996), Kuan and Hsu (1998), Lavielle and Moulines (2000), Shao (2011), and Iacone et al. (2013). Finally, there is also research devoted to the design of robust estimation procedures of the memory parameter in the presence of level shifts or other deterministic trends, as e.g., in McCloskey and Perron (2013).

Nevertheless, seemingly less attention has been paid to considering potential *joint* breaks in both d and μ (and possibly in the short-run dynamics). Whenever a break is detected, a joint test would help identify whether it originates in only one or in all parameters simultaneously.[1] Among

[1]Dolado *et al.* (2005) argue that it is important to distinguish between long memory, breaks in d and breaks in the level for at least three reasons. First, because it can improve forecasting. In particular, the

the scant literature addressing this issue, the following works are the most closely related to ours. Gil-Alaña (2008) is the first one to propose a single-step testing procedure based on an F-test whose limiting distribution is conjectured to correspond to the one derived by Bai and Perron (1998) for parameter breaks in regressions involving I(0) series. However, no formal proof of this claim is provided. Next, Hassler and Meller (2014) have extended Robinson (1994) and Breitung and Hassler's (2002) LM test of I(1) *vs.* FI(d) to deal with breaks in d while allowing also for level shifts. This test is conducted in a two-step sequential way. Initially, the location of the mean break is detected using Hsu's (2005) semiparametric testing approach; next, the corresponding broken intercept is removed from the time series to test for a break in d. How the two-step procedure affects the asymptotic properties of the test on d is not formally investigated and, in some cases, this could be problematic. For example, at the demeaning stage, the level could be very imprecisely estimated when d is close to 0.5, due to the $T^{1/2-d}$ rate of convergence of the sample mean. Thus, if results hinge on correct demeaning, there is additional uncertainty which is not properly taken into account in their testing procedure. Some of these shortcomings have been recently addressed in Rachinger (2017), who proposes a unified testing procedure for modeling parameter breaks jointly, rather than sequentially. Following Gil-Alaña (2008), his approach relies on extending Bai and Perron's (1998) test from I(0) to FI(d) processes. Specifically, when $d \in [0, 0.5)$, a Likelihood Ratio (LR) version of the well–known Chow test for parameter stability of d and μ is derived. Consistency results, T-rate convergence of the break fraction estimator and the limiting distributions of the estimated parameters under different sources of break are provided.

In line with Hassler and Meller (2014), our goal in this paper is to propose LM alternatives to the LR test for joint breaks which are simpler to compute since parameter estimation is only required under the null. However, we differ from these authors in two respects. First, we derive a single-step testing procedure rather than a sequential one with the caveat pointed out above. Second, given that Wald tests often exhibit higher power than LM tests, but at the cost of being more difficult to implement, we propose another test statistic which combines the computational simplicity of the LM test with the power gains of the Wald test. Inspired by Wooldridge (1990), these are LM regression-based tests which can also be interpreted as Wald tests since the relevant coefficients to be tested in the estimated regression are linearly related to the parameter of interest. For this reason, they are labeled as "LMW-type" tests in the sequel.

larger d is, the more observations are required to produce good forecasts. Further, forecasting requires some knowledge on the stability of the series. Secondly, because it can help to identify shocks. For economic modeling it matters whether the underlying shocks are persistent or transitory. Take, for example, the characteristics of the inflation rate as a measure of the credibility of the central bank. The less persistent the shocks are, the more credible is the central bank. Finally, in order to model two series as fractionally cointegrated, both series should share the same memory. Thus, if the memory is estimated too high due to instabilities, fractional cointegration could be a spurious outcome.

LMW-type tests have been proposed by Dolado et al. (2002, 2009), and Lobato and Velasco (2007) to test the nulls of I(1)/I(0) *vs.* the alternative of FI(d) processes, with $d \in (0, 1)$. We generalize their approach to testing for joint breaks in d and μ when the null is an FI(d) process with stable parameters. Moreover, both LM and LMW-type tests can deal with shifts in $d \in (-0.5, 0.5)$ under the alternative, which covers a wider range of values of d than those considered in the derivation of Rachinger´s (2017) LR tests, where it is assumed that $d \in [0, 0.5)$. This is so because the only requirement for implementing our tests is adequate performance of the constrained estimators under the null while LR tests require good performance of the constrained and unconstrained estimators of d under the null. Lastly, an additional advantage of our proposed tests is that, under a parametric setup, they provide a simple way of dealing with shifts in short-memory parameters, in addition to breaks in memory and level.[2]

Overall, this paper contributes to the relevant literature on the source of breaks in persistent time-series processes by deriving single-step LM and LMW-type tests (and their asymptotic distribution under the null and local alternatives) to test for the presence of a break either due to non-stable dynamics and/or level parameter. Both tests are easy to compute under the joint null of parameter constancy and have similar asymptotic behaviour under the null and local alternatives. However, we illustrate in finite-sample simulations that LMW-type tests could lead to power gains under fixed alternatives, especially when they involve a break in d.[3] Second, we briefly discuss how to extend the previous tests when breaks in different parameters may not be coincidental in time, as well as how they can be modified to test for multiple breaks.

The rest of the paper is structured as follows. In Section 2, we lay out the data generating processes (DGP). In Sections 3 and 4, we derive the asymptotic properties of the LM and LMW-type tests, respectively, both under the null and under local alternatives. We distinguish between two different settings: known and unknown break dates. In Section 5, we provide simulation results regarding the finite-sample performance of the tests. In Section 6 we apply the proposed methodology to the empirical analysis of structural changes in the forward discount of exchange rates. Finally, in Section 7 we draw some conclusions and briefly sketch how the tests could be generalized to allow for multiple breaks, therefore relaxing the previous simplifying assumption of coincident breaks in time. All the proofs and additional simulation results are collected in an online Appendix.

[2] Although a semiparametric approach would help us abstract from short-term dynamics when estimating d, we opt here for a parametric approach due to our interest in identifying further potential breaks in short-term dynamics.

[3] Notice that, in spite of the nonlinear nature of our proposed tests, this result someone echoes the well-known ranking in terms of power of Wald and LM tests in linear regression setups; see Engle (1984).

2 Data generation process

For simplicity, we start with the case of a *single* breakpoint (at a *known* or *unknown* date) which changes in the asymptotics as a fraction λ_0 of the sample size which lies in the interval $\Lambda = [\epsilon, 1-\epsilon]$, where $\epsilon > 0$ is assumed to be known. In particular, it is assumed that the time-series process is autoregressive FI(d_0) with $d_0 \in D$, where $D \subset (-0.5, 0.5)$ for $t = 1, \ldots, [T\lambda_0]$, while it becomes FI($d_1$) with $d_1 \in D$ for $t = [T\lambda_0]+1, \ldots, T$.[4] The level of the series is denoted as μ_0 in the first subsample and as μ_1 in the second subsample, with $\mu_0, \mu_1 \in M$, where M is a compact set. The following transition model is considered in the sequel as the DGP

$$\alpha_t(L)\,\Delta_t^{d_t}(y_t - \mu_t) = \varepsilon_t, \quad t = 1, 2, \ldots, \tag{1}$$

where ε_t is i.i.d.$(0, \sigma^2)$, so that

$$\begin{aligned}
\alpha_t(L)\,\Delta_t^{d_t} &= 1(t \le [T\lambda_0])\,\alpha_0(L)\,\Delta_t^{d_0} + 1(t > [T\lambda_0])\,\alpha_1(L)\,\Delta_t^{d_1}, \\
\mu_t &= 1(t \le [T\lambda_0])\,\mu_0 + 1(t > [T\lambda_0])\,\mu_1,
\end{aligned}$$

where $1(\cdot)$ is an indicator function of the relevant subsample, and $\alpha_i(L) = 1 - \alpha_{1,i}L \cdots - \alpha_{p,i}L^p$ are stable AR lag polynomials of order p with all roots outside the unit circle and a vector of unknown coefficients $\boldsymbol{\alpha}_i = (\alpha_{1,i}, \ldots, \alpha_{p,i})'$, $i = 0, 1$.[5] Moreover, at $[\lambda_0 T]$, there could be a shift in the parameters of DGP (1), so that $d_1 = d_0 + \theta_0$, $\mu_1 = \mu_0 + \nu_0$, and $\alpha_1(L) = \alpha_0(L) + \beta(L)$, where $\beta(L)$ is another lag polynomial with $\beta(0) = 0$ and coefficients $\boldsymbol{\beta} = \left(\beta_1, \ldots, \beta_p\right)'$. Finally, $\Delta_t^b := \sum_{j=0}^{t-1} \pi_j(b) L^j$, with $\pi_j(b) := \frac{\Gamma(j-b)}{\Gamma(-b)\Gamma(j+1)}$, $j = 0, 1, \ldots$, denotes the (truncated or "Type II") fractional-differencing filter for $b \in D$.

Remark 1. Notice that the previous definition of $\Delta_t^{d_t}$ implies that the filter applied to $(y_t - \mu_t)$ is $\sum_{j=0}^{t-1} \pi_j^*(d_0, \boldsymbol{\alpha}_0)$ when $t < [T\lambda_0]$ and $\sum_{j=0}^{t-1} \pi_j^*(d_1, \boldsymbol{\alpha}_1)$ when $t > [T\lambda_0]$, where $\alpha_i(L)\,\Delta_t^{d_i} := \sum_{j=0}^{t-1} \pi_j^*(d_i, \boldsymbol{\alpha}_i)\,L^j$. We prefer to use a truncated "Type II" filter, rather than an non-truncated "Type-I filter, because it is easier to accommodate nonstationary series with $d > 1/2$ after first differencing (see Remark 3 below).

[4] Our choice of the invertible range $D \subset (-0.5, 0.5)$ is dictated in part by the result in Hualde and Nielsen (2017) showing that consistent estimation of the level in an ARFIMA (p, d, q) process with a constant term (γ_0 =1, in their notation) and d lying in an arbitrarily large finite interval requires $d < 0.5$. However, when $d > 0.5$, the estimates of the other parameters governing the dynamics of the process are consistent and asymptotically normal, as in Hualde and Robinson (2011). Remark 3 below includes further discussion about the implementation of our tests when $d_0, d_1 > 0.5$.

[5] Formally the previous expression for the filter $\alpha_t(L)\,\Delta_t^{d_t}$ should be multiplied by $1(t > 0)$ since nesting the AR(p) lag polynomial $\alpha_t(L)$ with the truncated fractional filter $\Delta_t^{d_t}$ would require using pre-sample observations (negative lags). For simplicity we omit this more precise notation in the sequel.

Remark 2. Notice also that, by rewriting the DGP as $y_t = \mu_0 + \left(1 - \alpha_0(L)\Delta_t^{d_0}\right)(y_t - \mu_0) + \varepsilon_t$, and using recursively the truncated filters $\pi_j^*(d_i, \boldsymbol{\alpha}_i)$, $i = 0, 1$, it follows that

$$\begin{aligned} y_t &= \mu_0 - \sum_{j=1}^{t-1} \pi_j^*(d_0, \boldsymbol{\alpha}_0)\{y_{t-j} - \mu_0\}, \quad \text{for } t \leq [T\lambda_0] \\ y_t &= \mu_1 - \sum_{j=1}^{t-[T\lambda_0]-1} \pi_j^*(d_1, \boldsymbol{\alpha}_1)\{y_{t-j} - \mu_1\} - \sum_{j=t-[T\lambda_0]}^{t-1} \pi_j^*(d_1, \boldsymbol{\alpha}_1)\{y_{t-j} - \mu_0\} + \varepsilon_t, \quad \text{for } t > [T\lambda_0], \end{aligned}$$

so that the chosen filtering guarantees that the lags of y_t in the autoregression are centered around the appropriate value of μ_t.

Remark 3. Our approach can also deal with a non-stationary process with both $d_0, d_1 > 0.5$, and a potentially breaking linear trend, such that

$$\alpha_t(L)\Delta_t^{d_t}(y_t - \mu_t - \xi_t t) = \varepsilon_t,$$

$\xi_t = \xi_0 1(t \leq [T\lambda_0]) + \xi_1 1(t > [T\lambda_0])$, by applying our testing procedure to the first-differenced data to test for breaks in the intercept ξ_t and in the memory $d_t - 1$ of the increments Δy_t. This yields a consistent estimator of d_0 since Hualde and Nielsen's (2017) procedure works for a process with a single deterministic component and the memory lying in an arbitrarily large compact interval.

In sum, using the previous notation for potential shifts in the memory parameter (θ_0), the AR stable component ($\boldsymbol{\beta}$) and in the level (ν_0), and relabeling the indicator of the first and second regime as $R_t^{(1)}(\lambda) = 1(t \leq [T\lambda])$ and $R_t^{(2)}(\lambda) = 1(t > T\lambda)$, the models to be considered in Sections 3 and 4 are as follows

$$\left(\alpha_1(L) - R_t^{(1)}(\lambda_0)\beta(L)\right)\Delta_t^{d_1 - \theta_0 R_t^{(1)}(\lambda_0)}\left(y_t - \mu_1 + \nu_0 R_t^{(1)}(\lambda_0)\right) = \varepsilon_t, \ \ t = 1, \ldots, T, \quad \text{(Regime 1)}$$

$$\left(\alpha_0(L) + R_t^{(2)}(\lambda_0)\beta(L)\right)\Delta_t^{d_0 + \theta_0 R_t^{(2)}(\lambda_0)}\left(y_t - \mu_0 - \nu_0 R_t^{(2)}(\lambda_0)\right) = \varepsilon_t, \ \ t = 1, \ldots, T, \quad \text{(Regime 2)}$$

where, for brevity, we focus on the case where the indicator variable in (2) is $R_t^{(2)}$ (implementation of the test in *Regime 2*) to derive the properties of the tests. The consequences of using $R_t^{(1)} = 1 - R_t^{(2)}$ (*Regime 1*) instead of $R_t^{(2)}$ will be briefly discussed once these properties are established and a proposal for a symmetric test using both regimes is made.

One further issue which is worth discussing before presenting our proposed tests is the assumption of known lag length p for the AR process capturing short memory. In practice, one can fix a sufficiently large finite value of p such that the residuals of the model finally chosen are i.i.d. under the null of parameter stability (if not rejected) or under the alternative (if the null of no break is rejected). Whatever is the conclusion of the test, if the residuals in the finally chosen model do not pass the typical diagnostics, then p should be increased to clean the residuals from autocorrelation, and the whole testing procedure should be repeated. If this does not work, then this is an indication that the time series does not follow our proposed model.[6]

3 LM tests

According to the LM principle, we test the null hypothesis:

$$H_0 : (\theta_0, \boldsymbol{\beta}_0', \nu_0) = \mathbf{0}. \tag{H0}$$

As for the alternative, we start by analyzing the case where all parameters shift at an unknown fraction λ_0 of the sample size, and later deal with the simpler case of known λ_0,

$$H_1(\lambda_0) : (\theta_0, \boldsymbol{\beta}_0', \nu_0) \neq \mathbf{0}. \tag{H1}$$

For the chosen specification in *Regime 2*, the following Gaussian pseudo-log-likelihood function is used

$$\mathcal{L}_T(\psi, \lambda) = -\frac{T}{2} \log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \varepsilon_t(\psi, \lambda)^2, \tag{2}$$

for every possible breakpoint λ, and $\psi = \left(\theta, \boldsymbol{\beta}', \nu, d_0, \boldsymbol{\alpha}_0', \mu_0, \sigma^2\right)'$ where, using the *Regime 2* indicator $R_t^{(2)}$, the residuals are defined as follows

$$\varepsilon_t(\psi, \lambda) = \left(\alpha_0(L) + R_t^{(2)}(\lambda)\beta(L)\right)\left(\Delta_t^{d_0+\theta R_t^{(2)}(\lambda)}(y_t - \mu_0) - \nu \Delta_t^{d_0+\theta R_t^{(2)}(\lambda)} R_t^{(2)}(\lambda)\right).$$

For each λ, the LM test is based on the derivatives of $\mathcal{L}_T(\psi, \lambda)$ in the direction of ψ evaluated at the restricted estimates $\tilde{\psi}_T = (0, \mathbf{0}', 0, \tilde{d}_{0T}, \tilde{\boldsymbol{\alpha}}_{0T}', \tilde{\mu}_{0T}, \tilde{\sigma}_T^2)'$, where the last four elements of $\tilde{\psi}_T$ denote estimates under the null of no break of parameters $d_0, \boldsymbol{\alpha}_0'$, μ_0 and σ^2, respectively, using the whole sample of observations, $t = 1, \dots, T$ (see below for further details on the estimation procedure),

$$\widetilde{LM}_{2,T}(\lambda) = \left.\frac{\partial \mathcal{L}_T(\psi, \lambda)}{\partial \psi'}\right|_{\psi=\tilde{\psi}_T} \left(-\left.\frac{\partial^2 \mathcal{L}_T(\psi, \lambda)}{\partial \psi \partial \psi'}\right|_{\psi=\tilde{\psi}_T}\right)^{-1} \left.\frac{\partial \mathcal{L}_T(\psi, \lambda)}{\partial \psi}\right|_{\psi=\tilde{\psi}_T}, \tag{3}$$

where the subscript 2 in $\widetilde{LM}_{2,T}(\lambda)$ indicates that LM test is implemented in *Regime 2*.

[6] Ideally the number and location of breaks, and the order of the model (and of the trend) should all be chosen simultaneously, but this is well beyond the scope of the paper. Thus, it is preferable to shape our approach for break testing as being robust to the choice of p by taking it large enough to provide a good fit, but not tending to infinite with T (since this would involve a completely different theory).

In particular, the score in the directions of θ, $\boldsymbol{\beta}$ and ν can be expressed as,

$$\begin{aligned}
\tilde{\mathcal{L}}_{\theta,T}(\lambda) &= \left.\frac{\partial \mathcal{L}_T(\psi,\lambda)}{\partial \theta}\right|_{\psi=\tilde{\psi}_T} = -\frac{1}{\tilde{\sigma}_T^2}\sum_{t=[\lambda T]+1}^{T}\left(\log \Delta_t \tilde{\varepsilon}_t\right)\tilde{\varepsilon}_t \\
\tilde{\mathcal{L}}_{\boldsymbol{\beta},T}(\lambda) &= \left.\frac{\partial \mathcal{L}_T(\psi,\lambda)}{\partial \boldsymbol{\beta}}\right|_{\psi=\tilde{\psi}_T} = \frac{1}{\tilde{\sigma}_T^2}\sum_{t=[\lambda T]+1}^{T}\begin{pmatrix}\tilde{\alpha}_T^{-1}(L)\,\tilde{\varepsilon}_{t-1}\\ \dots \\ \tilde{\alpha}_T^{-1}(L)\,\tilde{\varepsilon}_{t-p}\end{pmatrix}\tilde{\varepsilon}_t \\
\tilde{\mathcal{L}}_{\nu,T}(\lambda) &= \left.\frac{\partial \mathcal{L}_T(\psi,\lambda)}{\partial \nu}\right|_{\psi=\tilde{\psi}_T} = \frac{1}{\tilde{\sigma}_T^2}\sum_{t=[\lambda T]+1}^{T}(\Delta_{t-[\lambda T]}^{\tilde{d}_{0T}}1)\tilde{\varepsilon}_t,
\end{aligned}$$

where $\log \Delta_t \tilde{\varepsilon}_t = -\sum_{j=1}^{t-1} j^{-1}\tilde{\varepsilon}_{t-j}$ depends on the restricted residuals $\tilde{\varepsilon}_t$ defined as,

$$\tilde{\varepsilon}_t = \varepsilon_t\left(\tilde{\psi}_T\right) = \tilde{\alpha}_{0T}(L)\,\Delta_t^{\tilde{d}_{0T}}\left(y_t - \tilde{\mu}_{0T}\right), \quad t = 1, 2, \dots, T. \tag{4}$$

The restricted estimates of the parameters using the whole sample, denoted as $(\tilde{d}_{0T}, \tilde{\boldsymbol{\alpha}}_{0T}, \tilde{\mu}_{0T})$, which are used to compute $\tilde{\varepsilon}_t$ and $\tilde{\sigma}_T^2$, result from minimizing the conditional sum of squares (CSS),

$$(\tilde{d}_{0T}, \tilde{\boldsymbol{\alpha}}_{0T}, \tilde{\mu}_{0T}) = \arg\min_{d\in D, \boldsymbol{\alpha}\in A, \mu\in M,}\sum_{t=1}^{T}\left(\alpha(L)\,\Delta_t^{d}(y_t-\mu)\right)^2, \tag{5}$$

while the corresponding variance estimator is

$$\tilde{\sigma}_T^2 = \frac{1}{T}\sum_{t=1}^{T}\tilde{\varepsilon}_t^2. \tag{6}$$

The properties of $\tilde{d}_{0T}$ have been discussed, *inter alia*, in Chung and Baillie (1993), Robinson (2006) and Hualde and Robinson (2011) in models without drift ($\mu = 0$), while Hualde and Nielsen (2017) extend the results in the last paper to cover joint estimation of memory and level. Using the results in Hualde and Nielsen (2017), the estimators $\tilde{d}_{0T}$ and $\tilde{\boldsymbol{\alpha}}_{0T}$ are $T^{1/2}$-consistent and asymptotically normal for $d_0 \in Int(D)$ and $\boldsymbol{\alpha}_0 \in Int(A)$, while $\tilde{\mu}_{0T}$ is $T^{1/2-d_0}$-consistent.

The relevant block of the inverse Hessian matrix concerning the subset of parameters $(\theta, \boldsymbol{\beta}', \nu)$ of ψ can be approximated by

$$\left[\left(\left.\frac{\partial^2 \mathcal{L}_T(\psi,\lambda)}{\partial\psi\partial\psi'}\right|_{\psi=\tilde{\psi}_T}\right)^{-1}\right]_{[1:(2+p),1:(2+p)]} = \widetilde{P}_2^{-1/2}\left(\left.\frac{\partial^2 \mathcal{L}_T(\psi,\lambda)}{\partial\left(\theta,\boldsymbol{\beta}',\nu\right)'\partial\left(\theta,\boldsymbol{\beta}',\nu\right)}\right|_{\psi=\tilde{\psi}_T}\right)^{-1}\widetilde{P}_2^{-1/2}\left(1+o_p(1)\right),$$

where the scaling matrix corresponding to testing in *Regime 2,* $\widetilde{P}_2$, defined as

$$\widetilde{P}_2 = P_2\left(\lambda; \tilde{d}_{0T}\right) = \begin{pmatrix}\lambda\cdot\mathbf{I}_{p+1} & 0\\ 0 & \frac{L_T\left(\tilde{d}_{0T};\lambda,\lambda\right)}{L_T\left(\tilde{d}_{0T};\lambda,\lambda\right)-L_T^2\left(\tilde{d}_{0T};0,\lambda\right)}\end{pmatrix}, \tag{7}$$

captures the effect of replacing the unknown values of d_0, $\boldsymbol{\alpha}_0$ and μ_0 by their (restricted) estimates, with $L_T(d;a,b) = T^{2d-1}(1-2d)\,\Gamma^2(1-d)\sum_{t=[\max(a,b)T]+1}^{T}(\Delta_{t-[aT]}^{d}1)(\Delta_{t-[bT]}^{d})$ and $\mathbf{I}_{p+1}$ the $p+1$ dimensional identity matrix.

Then our LM test statistic for unknown λ_0 becomes

$$\sup_{\lambda\in\Lambda}\widetilde{LM}_{2,T}(\lambda) = \widetilde{LM}_{2,T}\left(\tilde{\lambda}_T\right)$$

with $\Lambda = [\epsilon, 1-\epsilon]$, $0 < \epsilon < 1/2$, with ϵ denoting the sample trimming cutoff, and where $\tilde{\lambda}_T = \arg\max_\lambda \widetilde{LM}_{2,T}(\lambda)$.

Remark 4. When the LM test is implemented in *Regime 1*, denoted as $\widetilde{LM}_{1,T}(\lambda)$, the derivations are similar to those above, except that the sums in the score in the directions of $\theta, \boldsymbol{\beta}'$ and ν go from 1 to $[\lambda T]$ and the scaling factor $\widetilde{P}_2$ is replaced by

$$\widetilde{P}_1 = P_1(\lambda, \tilde{d}_{0T}) = \begin{pmatrix} (1-\lambda)\cdot\mathbf{I}_{p+1} & 0 \\ 0 & \frac{1+L_T\left(\tilde{d}_{0T};\lambda,\lambda\right)-2L_T\left(\tilde{d}_{0T};0,\lambda\right)}{L_T\left(\tilde{d}_{0T};\lambda,\lambda\right)-L_T^2\left(\tilde{d}_{0T};0,\lambda\right)} \end{pmatrix}.$$

3.1 Asymptotic theory of LM tests under local alternatives

We next derive the asymptotic distributions of the proposed $\widetilde{LM}$ tests under the following set of assumptions:

Assumption 1. $\varepsilon_t \sim iid\left(0, \sigma^2\right)$ with q moments such that $q > \max\{8, \frac{2}{1-2d_0}\}$.

Assumption 2. $d_0 \in Int(D)$, $D = \left[\underline{d}, \bar{d}\right]$, $-0.5 < \underline{d} < \bar{d} < 0.5$, $\mu_0 \in Int(M)$, $\boldsymbol{\alpha}_0 \in Int(\mathbf{A})$ and the compact set $\mathbf{A}$ excludes roots of $\alpha(L)$ on or inside the unit circle, and $\lambda_0 \in \Lambda$.

These assumptions correspond to those in Hualde and Nielsen (2017) for the particular case where $d_0 \in Int(D)$, whereas these authors consider that d_0 lies in compacts sets which can be arbitrarily large and we require at least eight moments to prove tightness for weak convergence. To assess the asymptotic null distribution and local power of the LM test, we analyze its properties under the following local-break alternatives,

$$H_{1,T}^{d,\alpha,\mu}(\lambda_0) : \left(\theta, \boldsymbol{\beta}', \nu\right) = \left(\delta/T^{1/2}, \boldsymbol{\gamma}'/T^{1/2}, \eta/T^{1/2-d_0}\right), \tag{8}$$

for some $\lambda_0 \in \Lambda$, where $\boldsymbol{\gamma} = \left(\gamma_1, \ldots, \gamma_p\right)'$, and the null is recovered by setting $(\delta, \boldsymbol{\gamma}', \eta) = 0$ while leaving λ_0 unspecified.

We next derive the asymptotic distribution of the $\widetilde{LM}_{2,T}$ test in (3) in the case of an unknown break fraction λ. The limiting distribution is a function of both standard Brownian Motion (BM) and a variant of fractional BM (fBM).

Let $\kappa = (\kappa_1, \ldots, \kappa_p)'$ with $\kappa_k = \sum_{j=k}^{\infty} j^{-1}c_{j-k}$, $k = 1, \ldots, p$, where the c_j are the coefficients of L^j in the expansion of $1/\alpha_0(L)$, and $\Phi = [\Phi_{k,j}]$, $\Phi_{k,j} = \sum_{t=0}^{\infty} c_t c_{t+|k-j|}$, $k, j = 1, \ldots, p$, denotes the Fisher information matrix for $\boldsymbol{\alpha}$ under Gaussianity. Further, let

$$\varpi_p(\lambda,\delta,\boldsymbol{\gamma}) = \Xi^{1/2}\left\{B_{p+1}(\lambda) - \lambda B_{p+1}(1)\right\} + \Xi\begin{pmatrix}\delta\\ \boldsymbol{\gamma}\end{pmatrix}\left(\lambda(1-\lambda_0) - (\lambda-\lambda_0)_+\right),\ \Xi = \begin{pmatrix}\pi^2/6 & \kappa'\\ \kappa & \Phi\end{pmatrix},$$

with B_{p+1} being a $(p+1)$-dimensional standardised Brownian motion.

Define,

$$\mathcal{A}^0_{d,p}(\lambda,\delta,\boldsymbol{\gamma}) = \frac{1}{\lambda(1-\lambda)}\varpi_p(\lambda,\delta,\boldsymbol{\gamma})'\,\Xi^{-1}\varpi_p(\lambda,\delta,\boldsymbol{\gamma})$$

and

$$\mathcal{A}^0_\mu(d_0,\lambda,a) = \frac{\left(\tilde{\mathcal{W}}_{d_0}(\lambda,1) - L(d_0;0,\lambda)\tilde{\mathcal{W}}_{d_0}(0,1) + a\left(L(d_0;\lambda,\lambda_0) - L(d_0;0,\lambda_0)L(d_0;0,\lambda)\right)\right)^2}{L(d_0;\lambda,\lambda) - L^2(d_0;0,\lambda)},$$

where $L(d;a,b) = (1-2d)\int_{\max(a,b)}^1 (s-a)^{-d}(s-b)^{-d}ds$ (so that $L(d;a,a) \equiv (1-a)^{1-2d}$ and $L(0;a,b) = 1-\max(a,b)$) and where $\tilde{\mathcal{W}}_d(a,b) = (1-2d)^{1/2}\int_a^b(s-a)^{-d}dB'(s)$ is a variant of fBM, $B(s)$ and $B'(s)$ are two independent BM. Notice that $\tilde{\mathcal{W}}_d(a,b)$ differs from the standard fBM, $\mathcal{W}_d(\lambda) = (1-2d)^{1/2}\int_0^\lambda(\lambda-s)^{-d}dB'(s)$, by its particular covariance structure, given by $Cov\left(\tilde{\mathcal{W}}_d(s,1), \tilde{\mathcal{W}}_d(t,1)\right) = L(d;s,t)$.

Then, the following result holds.

Theorem 1 *With an unknown break fraction λ_0, under Assumptions 1 and 2 and $H_{1T}(\lambda_0)$,*

$$\sup_{\lambda\in\Lambda}\widetilde{LM}_{2,T}(\lambda) \xrightarrow{d} \sup_{\lambda\in\Lambda}\left\{\mathcal{A}^0_{d,p}(\lambda,\delta,\boldsymbol{\gamma}) + \mathcal{A}^0_\mu\left(d_0,\lambda,\frac{\eta/\sigma}{\sqrt{1-2d_0}\Gamma(1-d_0)}\right)\right\}.$$

Remark 5. The distribution of the sup-$\widetilde{LM}_{2,T}(\lambda)$ test under H_0 is then given by

$$\sup_{\lambda\in\Lambda}\left\{\mathcal{A}^0_{d,p}(\lambda,0,\mathbf{0}) + \mathcal{A}^0_\mu(d_0,\lambda,0)\right\}.$$

Notice that, due to the lack of identification of the break fraction under the null of no breaks in any of the parameters, the distribution only depends on d_0, but not on λ_0, $\boldsymbol{\alpha}_0$ or μ_0. Critical values of such limiting distribution are shown in Table 1, for a grid of values of d_0 and ϵ, which have been generated from Theorem 1 using 2,000 grid points for the break fraction, and 20,000 simulations. To compute the critical values for an unknown d_0, we interpolate between these values and replace d_0 by $\tilde{d}_{0T}$ as in (5) (see Giraitis et al., 2006, for a similar solution).

[Table 1 about here]

Remark 6. Under local alternatives, the two components $A^0_{d,p}$ and A^0_{μ} in the asymptotic distribution of the sup- $\widetilde{LM}_{2,T}(\lambda)$ test capture the contributions of the local shifts of the dynamics parameters and of the level, respectively.[7] It is noteworthy that, while $\mathcal{A}^0_{d,p}(\lambda, \delta, \boldsymbol{\gamma})$ is symmetric around the break fraction $\lambda_0 = 0.5$, $\mathcal{A}^0_{\mu}\left(d_0, \lambda, \eta/(\sigma_0\sqrt{1-2d_0}\Gamma(1-d_0))\right)$ turns out to be positively (negatively) skewed if $d_0 > 0$ ($d_0 < 0$). Hence, if there is only a break in $(d, \boldsymbol{\alpha}')$, the local power of the sup-$\widetilde{LM}_{2,T}(\lambda)$ test is maximized for $\lambda_0 = 0.5$. Yet, if there were either only breaks in μ or in both $(d, \boldsymbol{\alpha}')$ and μ, then local power would be highest for some $\lambda_0 < 0.5$ (resp. $\lambda_0 > 0.5$) if $d_0 > 0$ (resp. $d_0 < 0$).

Theorem 1 also nests the special cases where one tests exclusively for a break in a subset of the parameter space., e.g., for d and $\boldsymbol{\alpha}$ (so that A^0_{μ} drops) or only in μ (so that $A^0_{d,p}$ drops) under H_0, reflecting that these two tests are asymptotically independent. Notice that if it is assumed that *only* a subset of the parameters breaks, a testing procedure which does not allow for a break in the other parameters could lead to better power properties in finite samples. However, estimation of the model under this null could lead to misleading conclusions if the other parameters are the ones that actually shift, while the tested parameter happens to be constant.

Lastly, Corollary 1 below provides the asymptotic distribution of the $\widetilde{LM}_{2,T}$ test for the more restrictive case when the break fraction λ_0 is taken to be known.

Corollary 1 *With known break fraction λ_0, under Assumptions 1 and 2, and hypothesis $H_{1T}(\lambda_0)$,*

$$\widetilde{LM}_{2,T}(\lambda_0) \xrightarrow{d} \chi^2_{2+p}(c(\lambda_0)),$$

with non-centrality parameter

$$c(\lambda_0) = \omega^2_p(\delta, \boldsymbol{\gamma})\lambda_0(1-\lambda_0) + \frac{\eta^2}{\sigma^2}\frac{L(d_0; \lambda_0, \lambda_0) - L^2(d_0; 0, \lambda_0)}{(1-2d_0)\Gamma^2(1-d_0)} \equiv c_{d,\boldsymbol{\alpha}}(\lambda_0) + c_{\mu}(\lambda_0),$$

where $\omega^2_p(\delta, \boldsymbol{\gamma}) = (\delta \quad \boldsymbol{\gamma}')\,\Xi\,(\delta \quad \boldsymbol{\gamma}')'$.

As expected, when λ_0 is known, the asymptotic distribution becomes a chi-square with $2+p$ degrees of freedom and with a non-centrality parameter $c(\lambda_0)$ which depends on the two drifts under local alternatives, namely $c_{d,\boldsymbol{\alpha}}(\lambda_0)$ and $c_{\mu}(\lambda_0)$. Moreover, as in the case of unknown λ_0, Corollary 1 nests the cases of testing for a break in a subset of the parameters: (i) if we test for

[7]Note that the limit term $A^0_{d,p}$ is similar to that obtained by Horváth and Shao (1999) in their test for a break only in d using an LR test from Whittle estimation. Likewise A^0_{μ} is similar to the limit term derived by Iacone et al. (2013) for a break only in μ in the first-differenced version of their model to test for a break only in the linear trend in a FI(d) process under any memory, using Abadir et al.' s (2007) Extended Local Whittle estimation. Thus, our result differs from these authors' results in that it allows for joint breaks in both parameters, besides in the short-run dynamics.

a break only in $(d, \boldsymbol{\alpha})$ the limiting distribution becomes $\chi^2_{1+p}\left(c_{d,\boldsymbol{\alpha}}\left(\lambda_0\right)\right)$, where c_μ drops even if $\eta \neq 0$, and (ii) if we test for a break only in μ, the limiting distribution becomes $\chi^2_1\left(c_\mu\left(\lambda_0\right)\right)$, where $c_{d,\boldsymbol{\alpha}}$ drops even when $(\delta \quad \boldsymbol{\gamma}') \neq 0$.

Remark 7. Martins and Rodriguez (2014) and Hassler and Meller (2014) have recently proposed similar LM test statistics for a break in d, but under the assumption that the first-regime memory parameter (d_0) is known. In such a case, the variance of the test would be smaller than under unknown d_0, resulting in a higher local power.[8] However, since the assumption of known d_0 could be quite restrictive in practice, they suggest some estimators of the memory parameter. Martins and Rodriguez (2014) plug in a parametric estimator of d to derive the asymptotic distribution of the corresponding LM test statistic. Yet, their approximation may not be accurate enough since it ignores the covariance between the test statistic and the estimator under the null. As already discussed, Hassler and Meller (2014) plug in a semiparametric estimator for d but they do not derive the limiting distribution of their LM test despite acknowledging that it would be altered due to the lower rate of convergence of their proposed estimator.

3.2 Consistency of the LM test

In this section we prove the consistency of the LM tests for breaks in either a subset or all of the parameters. In particular, as regards the $\widetilde{LM}_{2,T}$ tests for the null $H_0 : \left(\theta_0, \boldsymbol{\beta}_0', \nu_0\right) = \mathbf{0}$, we consider the following alternative fixed hypotheses:

$$\begin{aligned} H_1^{d,\alpha}\left(\lambda_0\right) &: \quad \theta_0 \neq 0,\ \boldsymbol{\beta}_0 \neq \mathbf{0} \text{ and } \nu_0 = 0, \\ H_1^{\mu}\left(\lambda_0\right) &: \quad \theta_0 = 0,\ \boldsymbol{\beta}_0 = \mathbf{0} \text{ and } \nu_0 \neq 0, \\ H_1^{d,\alpha,\mu}\left(\lambda_0\right) &: \quad \theta_0 \neq 0,\ \boldsymbol{\beta}_0 \neq \mathbf{0} \text{ and } \nu_0 \neq 0. \end{aligned}$$

where $H_1^{d,\alpha}\left(\lambda_0\right)$, $H_1^{\mu}\left(\lambda_0\right)$ and $H_1^{d,\alpha,\mu}\left(\lambda_0\right)$ entail respectively: (i) only a break in $(d, \boldsymbol{\alpha})$, (ii) only a break in μ, and (iii) joint breaks in $(d, \boldsymbol{\alpha})$ and μ.

Under the corresponding alternative hypotheses, the following result holds.

Proposition 1 *Under Assumptions 1 and 2, then:*
The LM test statistics for a break in all parameters, $\widetilde{LM}_{2,T}\left(\lambda_0\right)$ and $\sup_\lambda \widetilde{LM}_{2,T}\left(\lambda\right)$, diverge: (i) at rate T under either $H_1^{d,\alpha,\mu}\left(\lambda_0\right)$ or $H_1^{d,\alpha}\left(\lambda_0\right)$, and (ii) at rate T^{1-2d_0} (resp. T) under $H_1^{\mu}\left(\lambda_0\right)$ with $d_0 \geq 0$ (resp. $d_0 < 0$). The same results hold for $\widetilde{LM}_{1,T}(\lambda_0)$ and $\sup_\lambda \widetilde{LM}_{1,T}\left(\lambda\right)$.

Remark 8. As anticipated above, it is important to remark that the use of individual $\widetilde{LM}$ tests for breaks in a single parameter – say, in d or μ– may lead to spurious rejections when the

[8] Notice that this result can also be easily extended to the case of a known first-regime level (μ_0).

non-tested parameter happens to be the only one breaking. In the rest of this subsection, we discuss this issue in further detail.

Whenever the joint $\widetilde{LM}_{2,T}$ tests reject the null of parameter stability, one may also be then interested in identifying the source of the break under any of the aforementioned fixed alternatives. For ease in the following exposition, let us consider the simple case where λ_0 is known, $\alpha_0(L) = 1$, and $\boldsymbol{\beta}_0 = \mathbf{0}$. To achieve break-source identification in this case, it is convenient to derive individual LM tests under the following two simple null hypotheses,

$$H_0^d(\lambda_0) : \theta_0 = 0,$$
$$H_0^\mu(\lambda_0) : \nu_0 = 0,$$

which, unlike the versions of the $\widetilde{LM}_{2,T}$ tests which assume that the other parameters are constant, abstain from specifying whether the other (non-tested) parameter is breaking or not. Then, a sequential procedure can be designed to test first for the presence of breaks in d and μ using the joint $\widetilde{LM}_{2,T}$ tests and, in case of rejection of the joint null, then a test of the individual null $H_0^d(\lambda_0)$ (resp. $H_0^\mu(\lambda_0)$) could be applied at a second stage to confirm if d (resp. μ) is actually breaking, irrespectively of whether the other parameter shift or not. To robustify these individual tests against misleading inference, it is preferable to remain agnostic about how the non-nested parameter behaves. For example, to implement a robust test of the null $H_0^d(\lambda_0)$ against the alternative $H_1^d(\lambda_0) : \theta_0 = 0$, rather than using a $\widetilde{LM}_{2,T}$ test based on the score in the direction of θ with H_0-restricted estimates $(\tilde{d}_{0T}, \tilde{\mu}_{0T})$ as in (6), the following $H_0^d(\lambda_0)$-restricted estimates should be considered

$$(\bar{d}_{0T}, \bar{\mu}_{0T}, \bar{\nu}_{0T}) = \arg \min_{d \in D, \mu, \mu+\nu \in M} \sum_{t=1}^{T} \left(\Delta_t^d \left(y_t - \mu - \nu R_t^{(2)}(\lambda_0) \right) \right)^2, \tag{9}$$

where different levels are allowed in each of the two regimes. Then, the robust individual version of the LM test for $H_0^d(\lambda_0)$, denoted as $\overline{LM}_{2,T}^d(\lambda_0)$, is given by

$$\overline{LM}_{2,T}^d(\lambda_0) = \left.\frac{\partial \mathcal{L}_T(\psi, \lambda_0)}{\partial \psi}\right|_{\psi=\bar{\psi}_T} \left(- \left.\frac{\partial^2 \mathcal{L}_T(\psi, \lambda_0)}{\partial \psi \partial \psi'}\right|_{\psi=\bar{\psi}_T} \right)^{-1} \left.\frac{\partial \mathcal{L}_T(\psi, \lambda_0)}{\partial \psi}\right|_{\psi=\bar{\psi}_T},$$

where $\bar{\psi}_T = (0, \bar{\nu}_{0T}, \bar{d}_{0T}, \bar{\mu}_{0T}, \bar{\sigma}_T^2)'$ and $\bar{\sigma}_T^2 = T^{-1} \sum_{t=1}^T \bar{\varepsilon}_t^2$ uses the $H_0^d(\lambda_0)$-restricted residuals $\bar{\varepsilon}_t = \varepsilon_t(\bar{\psi}_T) = \Delta_t^{\bar{d}_{0T}} \left(y_t - \bar{\mu}_{0T} - \bar{\nu}_{0T} R_t^{(2)}(\lambda_0) \right)$. Likewise, we can define $\overline{LM}_{2,T}^\mu(\lambda_0)$ to test H_0^μ based on the corresponding $H_0^\mu(\lambda_0)$-restricted estimation, where this time $\bar{\psi}_T = (\bar{\theta}_{0T}, 0, \bar{d}_{0T}, \bar{\mu}_{0T}, \bar{\sigma}_T^2)'$. Similar robustified test statistics can be derived under the first regime, namely, $\overline{LM}_{1,T}^\mu(\lambda_0)$ and $\overline{LM}_{1,T}^\mu(\lambda_0)$.

The following Proposition discusses the asymptotic behaviour of the robustified LM tests.

Proposition 2 *Under Assumptions 1 and 2:*

(a) *The robustified test statistic* $\overline{LM}^{d}_{2,T}(\lambda_0)$ *for a break only in the memory, diverges at rate* T *under either* $H_1^{d,\mu}(\lambda_0)$ *or* $H_1^{d}(\lambda_0)$*. By contrast, it converges to a* χ^2_1 *distribution under* $H_1^{\mu}(\lambda_0)$*, namely, when only* μ *breaks. The same results hold for* $\overline{LM}^{d}_{1,T}(\lambda_0)$*.*

(b) *The robustified test statistic* $\overline{LM}^{\mu}_{2,T}(\lambda_0)$ *for a break only in the level, diverges: (i) at rate* T^{1-2d_0} *(resp.* T*) for* $0 < d < 0.5$ *(resp.* $-0.5 < d < 0$*) under* $H_1^{\mu}(\lambda_0)$*; (ii) at rate* T^{1-2d_1} *(resp.* T*) for* $0 \leq d_1 < 0.5$ *(resp.* $-0.5 < d_1 < 0$*) and* $H_1^{d,\mu}(\lambda_0)$*. By contrast, it converges to a* χ^2_1 *distribution under* $H_1^{d}(\lambda_0)$*, namely, when only* d *breaks. The same results hold for* $\overline{LM}^{\mu}_{1,T}(\lambda_0)$*, except that the divergence rate in (ii) is* T^{1-2d_0} *for* $0 \leq d_0 < 0.5$ *and* $H_1^{d,\mu}(\lambda_0)$*.*

Proposition 2 illustrates why, upon rejection in the first stage, the robust individual tests $\overline{LM}^{d}_{2,T}(\lambda_0)$ and $\overline{LM}^{\mu}_{2,T}(\lambda_0)$ help identify which parameter or parameters actually break. The reason is that the individual test of $H_0^{d}(\lambda_0)$ (resp. $H_0^{\mu}(\lambda_0)$) will reject asymptotically this null under $H_1^{d}(\lambda_0)$ (resp. $H_1^{\mu}(\lambda_0)$) but will have only trivial power under $H_1^{\mu}(\lambda_0)$ (resp. $H_1^{d}(\lambda_0)$). Notice also that, from Propositions 1 and 2, the rates of divergence of the test statistics $\widetilde{LM}^{d,\mu}_{2,T}$ and $\overline{LM}^{\mu}_{2,T}(\lambda_0)$ under $H_1^{\mu}(\lambda_0)$ depend on the value of the memory parameter during the second regime: d_0 if memory is constant, or d_1 if it breaks. We conjecture that, when λ_0 is unknown, then it can be replaced by the estimate of the break date obtained from the first step, namely, $\widetilde{\lambda}_T = \arg\max_{\lambda} \widetilde{LM}_{2,T}(\lambda)$, following the arguments by Rachinger (Theorems 1 and 2, 2017) on how $\widetilde{\lambda}_T$ provides a T-consistent estimator of λ_0. Although Rachinger's (2017) consistency proof relies on conditional sums of squares (CSS) rather than on scores, LM tests based on the latter can be re-formulated in terms of CSS, as in Breitung and Hassler (2002) and Hassler and Meller (2014).

4 Regression-based LMW-type tests

As an alternative to the LM test based on the restricted ML estimates, an LMW-type test based on an auxiliary regression can be derived along the lines of Lobato and Velasco (2007; LV henceforth). Building upon previous results by Dolado et al. (2002), LV derive an Efficient Fractional Dickey Fuller (EFDF) test for the null hypothesis of $d = 1$ against the alternative of $d < 1$. Later on, Dolado et al. (2009) have generalized this testing approach by allowing the null to be any memory $d = d_0$ against the alternative $d \neq d_0$.[9] Moreover, they argue that, while remaining asymptotically equivalent under local alternatives, the LMW-type EFDF test could achieve higher power than the LM test under fixed alternatives. Thus, relying upon this approach, a similar test for joint breaks in $d, \boldsymbol{\alpha}$ and μ is proposed here, focusing on the null hypothesis (H0) in *Regime 2*.

[9] They also consider the estimation of a deterministic component and show that its pre-estimation does not affect the asymptotic distribution of the test.

For simplicity, we start by considering the case where the parameters in *Regime 1*, $d_0, \boldsymbol{\alpha}_0$ and μ_0 are assumed to be known, so the data in *Regime 2*, $t = [T\lambda_0] + 1, \ldots, T$, satisfies

$$\begin{aligned}
\alpha_0(L)\Delta_t^{d_0}(y_t - \mu_0) &= \alpha_0(L)\Delta_t^{d_0}(y_t - \mu_0) - \alpha_1(L)\Delta_t^{d_1}(y_t - \mu_t) + \varepsilon_t \\
&= \alpha_0(L)\left[1 - \Delta_t^{\theta_0}\right]\Delta_t^{d_0}(y_t - \mu_0) + \beta(L)\Delta_t^{d_1}\left(y_t - \mu_0 - \nu_0 R_t^{(2)}(\lambda_0)\right) \\
&\quad + \nu_0\alpha_0(L)\Delta_t^{d_1}R_t^{(2)}(\lambda_0) + \varepsilon_t,
\end{aligned}$$

where recall that $d_1 = d_0 + \theta_0$, $\alpha_1(L) = \alpha_0(L) + \beta(L)$ and $\mu_1 = \mu_0 + \nu_0$ with $\mu_t = \mu_0 + \nu_0 R_t^{(2)}(\lambda_0)$ and $\Delta_t^d R_t^{(2)}(\lambda) = \sum_{j=0}^{t-1} 1(j < t - [T\lambda])\pi_j(d) = \sum_{j=0}^{t-[T\lambda]-1}\pi_j(d) = \pi_{t-[T\lambda]-1}(d-1)$. Then, a test for the joint null of $\left(\theta_0, \boldsymbol{\beta}', \nu_0\right) = 0$ can be constructed by means of a joint test of

$$H_0 : \vartheta_1 = \vartheta_2 = \cdots = \vartheta_{2+p} = 0$$

in the following regression model,[10]

$$\begin{aligned}
\alpha_0(L)\Delta_t^{d_0}(y_t - \mu_0) &= \vartheta_1\alpha_0(L)\left[\frac{1 - \Delta_t^{\theta_0 R_t^{(2)}(\lambda)}}{\theta_0}\right]\Delta_t^{d_0}(y_t - \mu_0) \qquad (10)\\
&\quad + \sum_{j=1}^{p}\vartheta_{j+1}R_{t-j}^{(2)}(\lambda)\Delta_{t-j}^{d_1}\left(y_{t-j} - \mu_0 - \nu_0 R_{t-j}^{(2)}(\lambda)\right) \\
&\quad + \vartheta_{p+2}\alpha_0(L)\Delta_t^{d_1}R_t^{(2)}(\lambda) + \varepsilon_t,
\end{aligned}$$

for $t = 1, \ldots, T$, and each λ. Defining $\Theta = (\vartheta_1, \vartheta_2 ..., \vartheta_{2+p})'$, $Y_t^0 = \alpha_0(L)\Delta_t^{d_0}(y_t - \mu_0)$ and for each $(\lambda, \theta, \nu, d, \boldsymbol{\alpha}', \mu)$

$$X_t(\lambda) = \left(\alpha(L)\left[\frac{1 - \Delta_t^{\theta R_t^{(2)}(\lambda)}}{\theta}\right]\Delta_t^d(y_t - \mu), \left\{R_t^{(2)}(\lambda)\Delta_{t-j}^{d+\theta}\left(y_{t-j} - \mu - \nu R_{t-j}^{(2)}(\lambda)\right)\right\}_{j=1}^{p}, \ \alpha(L)\Delta_t^{d+\theta}R_t^{(2)}(\lambda)\right)$$

the regression (10) can be rewritten in a more compact way as

$$Y_t^0 = \Theta' X_t^0(\lambda) + \varepsilon_t, \qquad (11)$$

with $X_t^0(\lambda) = X_t(\theta_0, \upsilon_0, d_0, \boldsymbol{\alpha}_0, \mu_0)$.

Under the more realistic assumption of unknown $d_0, \boldsymbol{\alpha}_0$ and μ_0, running regression (11) requires the estimation of those parameters, on top of θ and ν. For $d_0, \boldsymbol{\alpha}_0$ and μ_0, one can use the restricted estimates $\tilde{d}_{0T}$, $\tilde{\boldsymbol{\alpha}}_{0T}$ and $\tilde{\mu}_{0T}$ obtained under the null using the whole sample, (5). As for θ and ν, one can set $\hat{\theta}_T(\lambda) = \hat{d}_{1T}(\lambda) - \tilde{d}_{0T}$ and $\hat{\nu}_T(\lambda) = \hat{\mu}_{1T}(\lambda) - \tilde{\mu}_{0T}$ where $\hat{d}_{1T}(\lambda)$ and $\hat{\mu}_{1T}(\lambda)$ are the CSS estimates obtained from the second subsample, defined by a given λ. Our main justification for the estimation of $(d_0, \boldsymbol{\alpha}_0, \mu_0)$ based on the minimization of (5) with observations for the whole

[10] As pointed out in LV (2007) notice that, for $\theta \to 0$, the filter $\left[\frac{1-\Delta_t^{\theta}}{\theta}\right]$ becomes $-\log\Delta_t$ when $\theta \to 0$, which corresponds to the well-known lag filter $\sum_{k=1}^{t-1} k^{-1}L^k$ used in the regression-based LM test.

sample, rather than just for *Regime 1*, is that it facilitates comparison of the LM and LMW-type tests, since both use the same parameter estimates under the null.[11] This leads to the following feasible regression model

$$\tilde{Y}_t = \Theta' \tilde{X}_t(\lambda) + e_t \tag{12}$$

with $\tilde{Y}_t = \tilde{\alpha}_{0T}(L)\,\Delta_t^{\tilde{d}_{0T}}(y_t - \tilde{\mu}_{0T})$ on $\tilde{X}_t(\lambda) = X_t\left(\lambda, \hat{\theta}_T(\lambda), \hat{\nu}_T(\lambda), \tilde{d}_{0T}, \tilde{\boldsymbol{\alpha}}_{0T}, \tilde{\mu}_{0T}\right)$.

Testing for breaks in all parameters corresponds to the joint null hypothesis of $\vartheta_1 = \vartheta_2 = ... = \vartheta_{2+p} = 0$ in (12), while testing for example for a break only in the dynamics (resp. only in μ) corresponds to the null hypothesis of $\vartheta_1 = ... = \vartheta_{1+p} = 0$ (resp. $\vartheta_{2+p} = 0$).[12] Then, the LMW-type test statistic (implemented in *Regime 2*) from regression (12) for the joint hypothesis $H_0 : \Theta = 0$ is defined as

$$\widetilde{LMW}_{2,T}(\lambda) = \tilde{\Theta}_T(\lambda)'\,\tilde{V}_T^{-1}(\lambda)\,\tilde{\Theta}_T(\lambda), \tag{13}$$

where $\tilde{\Theta}_T(\lambda) = \left(\tilde{\vartheta}_{1T}(\lambda), \tilde{\vartheta}_{2T}(\lambda), ..., \tilde{\vartheta}_{2+pT}(\lambda)\right)'$ denotes the LS estimate of Θ. We, further, set

$$\tilde{V}_T(\lambda) = \hat{\sigma}_T^2(\lambda)\,\widetilde{P}_2^{1/2}\left(\sum_{t=1}^{T}\tilde{X}_t(\lambda)\,\tilde{X}_t(\lambda)'\right)^{-1}\widetilde{P}_2^{1/2},$$

where

$$\hat{\sigma}_T^2(\lambda) = \frac{1}{T}\sum_{t=1}^{[T\lambda]}\left(\hat{\alpha}_{0T}(L)\,\Delta_t^{\hat{d}_{0T}}(y_t - \hat{\mu}_{0T})\right)^2 + \frac{1}{T}\sum_{t=[T\lambda]+1}^{T}\left(\hat{\alpha}_{1T}(L)\,\Delta_{t-[T\lambda]}^{\hat{d}_{1T}}(y_t - \hat{\mu}_{1T})\right)^2 \tag{14}$$

with $\left(\hat{d}_{0T}, \hat{\boldsymbol{\alpha}}_{0T}, \hat{\mu}_{0T}\right)$ and $\left(\hat{d}_{1T}, \hat{\boldsymbol{\alpha}}_{1T}, \hat{\mu}_{1T}\right)$ being the CSS estimators for *Regimes 1* and *2*, respectively, defined by λ, and $\widetilde{P}_2$ as in (7). Notice that in the construction of the LMW-type test, we use the variance estimate under the alternative to improve its power properties.

From the discussion in Wooldridge (2002) and LV (2007), it follows that, when $(d_0, \boldsymbol{\alpha}_0, \mu_0)$ are taken as known, the estimation of (θ, ν) by $\left(\hat{\theta}_T(\lambda), \hat{\nu}_T(\lambda)\right)$ does not affect the null asymptotic distribution of the Wald-type test derived from (11). However, this is no longer true when $(d_0, \boldsymbol{\alpha}_0, \mu_0)$ need to be estimated since these parameters affect the left-hand-side variable in regression (12), and, as with the LM test, this estimation of the parameters $(d_0, \boldsymbol{\alpha}_0, \mu_0)$ increases the variance of the LMW-type test statistic. This is reflected in the need to pre- and post-multiply by $\widetilde{P}_2^{1/2}$ in the definition of $\tilde{V}_T(\lambda)$ compared to the usual LS expression. Then, our test statistic becomes

$$\sup_{\lambda\in\Lambda}\widetilde{LMW}_{2,T}(\lambda) = \widetilde{LMW}_{2,T}(\tilde{\lambda}_T),$$

[11] In addition, as found in our simulation study, the size in finite samples of the test becomes closer to the nominal size when the longer sample is used.

[12] As before, note that if it assumed that only a subset of the parameters breaks, a test not allowing for a break in the non-tested parameter again should enjoy better finite sample properties (e.g. set $\mu_0 = \mu_1$ or $\nu = 0$ in (10) when testing for a break in the dynamics, that is $H_0 : \vartheta_1 = \vartheta_2 = ... = \vartheta_{1+p} = 0$).

where $\tilde{\lambda}_T = \arg\max_{\lambda\in\Lambda} \widetilde{LMW}_{2,T}(\lambda)$.

Remark 9. As in the case of the LM test, the LMW-type test can also be implemented in the *first* regime, being denoted as $\widetilde{LMW}_{1,T}(\lambda)$, by means of the $R_t^{(1)}$ indicator. The previous regression model then becomes,

$$\begin{aligned}
\tilde{\alpha}_{1T}(L)\,\Delta_t^{\tilde{d}_{1T}}(y_t - \tilde{\mu}_{1T}) \quad &= \quad \vartheta_1 \left[\frac{1-\Delta_t^{-\hat{\theta}_T R_t^{(1)}(\lambda)}}{\hat{\theta}_T}\right] \tilde{\alpha}_{1T}(L)\,\Delta_t^{\tilde{d}_{1T}}(y_t - \tilde{\mu}_{1T}) \\
&\quad + \sum_{j=1}^{p} \vartheta_{j+1} R_{t-j}^{(1)}(\lambda)\,\Delta_{t-j}^{\hat{d}_{0T}}\left(y_{t-j} - \tilde{\mu}_{1T} + \hat{\nu}_T R_t^{(1)}(\lambda)\right) \\
&\quad + \vartheta_{p+2}\left[R_t^{(1)}(\lambda)\,\tilde{\alpha}_{1T}(L)\,\Delta_t^{\hat{d}_{0T}} + \left(1 - R_t^{(1)}(\lambda)\right)\hat{\alpha}_{0T}(L)\left(\Delta_t^{\hat{d}_{0T}} - \Delta_{t-[\lambda T]}^{\hat{d}_{0T}}\right)\right] + \tilde{\varepsilon}_t.
\end{aligned}$$

4.1 Asymptotic properties of LMW-type tests

Using estimates $(\tilde{d}_{0T}, \tilde{\boldsymbol{\alpha}}_{0T}, \tilde{\mu}_{0T})$ in place of $(d_0, \boldsymbol{\alpha}_0, \mu_0)$, we next show that the asymptotic distributions of the LMW-type tests are identical to those of the equivalent LM tests under local alternatives. The insight for this result is that the LMW-type tests just become regression versions of the usual LM test statistics when these restricted estimates are used to construct the dependent variable in the regression model above.

Theorem 2 *Under Assumptions 1 and 2 and under the local hypothesis H_{1T}, for unknown parameters $d_0, \boldsymbol{\alpha}_0$ and μ_0 and for*

(a) an unknown break fraction λ, the asymptotic behaviour of the LMW-type test $\sup_\lambda \widetilde{LMW}_{2,T}(\lambda)$ corresponds to the one derived for the $\sup_\lambda \widetilde{LM}_{2,T}(\lambda)$ test in Theorem 1, and idem for $\sup_\lambda \widetilde{LMW}_{1,T}(\lambda)$.

(b) a known break fraction λ_0, the asymptotic behaviour of the LMW-type test $\widetilde{LMW}_{2,T}(\lambda_0)$ corresponds to the one derived for the $\widetilde{LM}_{2,T}(\lambda_0)$ test in Corollary 1, and idem for $\widetilde{LMW}_{1,T}(\lambda_0)$.

In addition, we discuss the consistency of the LMW-type test for breaks in the dynamics and/or μ under fixed alternatives, where the following result holds.

Proposition 3 *The LMW-type tests for a break in all parameters, $\widetilde{LMW}_{2,T}^{d,\alpha,\mu}(\lambda_0)$ and $\sup_\lambda \widetilde{LMW}_{2,T}^{d,\alpha,\mu}(\lambda)$, behave as the joint $\widetilde{LM}$ tests in Proposition 1, and idem for $\widetilde{LMW}_{1,T}^{d,\alpha,\mu}(\lambda_0)$ and $\sup_\lambda \widetilde{LMW}_{1,T}^{d,\alpha,\mu}(\lambda)$.*

The main difference between the LM test and the LMW-type test is that while the former uses the filter $-\log\Delta_t$, the latter uses $(1-\Delta_t^{\theta})/\theta$, which converges to $-\log\Delta_t$ when $\theta \uparrow 0$ under local alternatives but it can be very different under fixed alternatives, when θ does not converge to zero.

Indeed, under fixed alternatives, the (rescaled) LM and LMW-type tests have different drift terms whose comparison can shed light on their relative asymptotic power despite not affecting the rates of divergence of the two tests. As explained in Dolado et al. (2017), these drifts are larger for the LMW-type test than for the LM test. For example, considering the alternative of only a break in d for known λ_0, with $\alpha_0(L) = 1$ and $\boldsymbol{\beta} = \mathbf{0}$, we have that

$$\begin{aligned} p\lim_{T\to\infty} \frac{\widetilde{LM}_{2,T}(\lambda)}{T} &= \frac{1-\lambda_0}{\lambda_0} C_{LM}(d_1, d_A); \\ p\lim_{T\to\infty} \frac{\widetilde{LMW}_{2,T}(\lambda)}{T} &= \frac{1-\lambda_0}{\lambda_0} C_{LMW}(d_1, d_A), \end{aligned}$$

such that the drift terms are given by,

$$\begin{aligned} C_{LM}(d_1, d_A) &= \frac{\left(\sum_{j=1}^{\infty} \left(\sum_{k=1}^{j} \frac{\Gamma(j-k+d_1-d_A)}{k\Gamma(j-k+1)} \right) \frac{\Gamma(j+d_1-d_A)}{\Gamma(j+1)} \right)^2}{\frac{\bar{\sigma}^2_{d,LM}}{\sigma^2} \sum_{j=1}^{\infty} \left(\sum_{k=1}^{j} \frac{\Gamma(j-k+d_1-d_A)}{k\Gamma(j-k+1)} \right)^2}, \\ C_{LMW}(d_1, d_A) &= \frac{\Gamma(1 + 2(d_A - d_1))}{\Gamma^2(1 + (d_A - d_1))} - 1, \end{aligned}$$

where d_A and $\bar{\sigma}^2_{d,LM}$ are the probability limits of the restricted estimate $\widetilde{d}_{0T}$ obtained from (5) and the estimated variance in the $\widetilde{LM}$ test, respectively, under the alternative $H_1^d(\lambda_0)$, which are equal to,

$$d_A = \arg\min_{d\in D_i} \left\{ \lambda_0 \frac{\Gamma(1 - 2(d_0 - d))}{\Gamma^2(d - d_0 + 1)} + (1-\lambda_0) \frac{\Gamma(1 - 2(d_1 - d))}{\Gamma^2(d - d_1 + 1)} \right\}, \tag{15}$$

$$\bar{\sigma}^2_{d,LM} = \sigma_0^2 \left(\lambda_0 \frac{\Gamma(1 + 2(d_A - d_0))}{\Gamma^2(1 + (d_A - d_0))} + (1-\lambda_0) \frac{\Gamma(1 + 2(d_A - d_1))}{\Gamma^2(1 + (d_A - d_1))} \right).$$

As illustrated in Figure 1 (panel a), where $\lambda_0 = 0.5$, $d_0 = 0.5$, and $\sigma_0 = 1$, the drift terms satisfy $C_{LMW}(d_1, d_A) > C_{LM}(d_1, d_A)$, being steeper when d decreases ($\theta < 0$). Likewise, the previous inequality between the drift terms also holds should one compute both tests under *Regime 1*, but in this case the drifts are steeper when d increases ($\theta > 0$). Therefore, under $H_1^d(\lambda_0)$, the LMW-type tests tend to slightly dominate the LM tests in terms of asymptotic power due their larger drift terms.

[Figure 1 about here]

4.2 Symmetric tests

Given that the power of the LM and LMW-type test depend on the direction in which the memory parameter d breaks, one possible suggestion to ameliorate such a dependence would be to take a simple average of the tests implemented in the first and second regimes. In what follows, these average tests will be denoted as $\widetilde{LM}_{1+2,T}$ and $\widetilde{LMW}_{1+2,T}$, respectively. Notice that, although other possibilities for pooling information exist (such as taking the maximum of both tests), averaging has the advantage of leading to symmetric versions of the $\widetilde{LM}$ and $\widetilde{LMW}$ test statistics. For this reason, in the sequel they will be labeled in short as *symmetric* tests. For example, in the case of an unknown breaking point, these tests are defined as follows

$$\sup_{\lambda\in\Lambda} \widetilde{LM}_{1+2,T}(\lambda) = \sup_{\lambda\in\Lambda} \frac{1}{2}\left(\widetilde{LM}_{1,T}(\lambda) + \widetilde{LM}_{2,T}(\lambda)\right)$$

$$\sup_{\lambda\in\Lambda} \widetilde{LMW}_{1+2,T}(\lambda) = \sup_{\lambda\in\Lambda} \frac{1}{2}\left(\widetilde{LMW}_{1,T}(\lambda) + \widetilde{LMW}_{2,T}(\lambda)\right),$$

whereas in the case where the breaking point λ_0 is assumed to be known, they become

$$\widetilde{LM}_{1+2,T}(\lambda_0) = \frac{1}{2}\left(\widetilde{LM}_{1,T}(\lambda_0) + \widetilde{LM}_{2,T}(\lambda_0)\right)$$

$$\widetilde{LMW}_{1+2,T}(\lambda_0) = \frac{1}{2}\left(\widetilde{LMW}_{1,T}(\lambda_0) + \widetilde{LMW}_{2,T}(\lambda_0)\right),$$

It is worth noticing that, under the null and the alternative, the asymptotic properties of the symmetric tests mimic those of the corresponding tests for the null that only d shifts in Theorems 1 and 2, so that they can be implemented using the same critical values provided in Table 1. The insight is that, since both tests converge to the same stochastic limit (depending on the same underlying BM and fBM when the break point is unknown or on the same chi-square variate when it is known), their average converges to the same limit under the null, while the drift equals the average drift of the tests under the fixed alternative.

5 Finite sample evidence

In this section, we report some simulation results regarding size and power of LM and LMW-type test in their symmetric versions, namely $\widetilde{LM}_{1+2,T}$ and $\widetilde{LMW}_{1+2,T}$.[13] First, we consider the case of a known break fraction of $\lambda_0 = 0.5$ and abstract from short-run dynamics, i.e. we set $\alpha_0(L) = 1$ and $\beta' = 0$. The significance level is 0.05 and the sample sizes are $T = 200, 500$ and $1,000$ when considering size, and $T = 200$ when considering power. We assume an error variance

[13] Results on simulated size and power of the individual LM and LMW-type tests implemented in *Regimes 1* and *2*, respectively, can be found in Tables A1 and A2 in the Online Appendix. A comparison of the results in these Tables with those displayed in Table 1 below for the symmetric versions of both tests shows that there are some advantages from using the latter. In particular, the size of the symmetric tests is more stable over the different values of d and that their power depends less on the direction of the break in this parameter.

$\sigma_0^2 = 1$ and take draws from a $N(0,1)$ distribution. To compute size, we allow d to take the values $\{-0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4\}$ and a non-breaking level of $\mu_0 = 0$. To compute power, we consider $d_0 \in \{-0.2, 0, 0.2\}$, $d_1 \in \{-0.4, -0.2, 0, 0.2, 0.4\}$, $\mu_0 = 0$ and $\mu_1 = \{0, 0.25, 0.5, 1\}$. The number of simulations is $10,000$.

Table 2 (panels a and b) displays the size of the above-mentioned tests for breaks in both d and μ, respectively at a known break fraction of $\lambda_0 = 0.5$. The main finding is that the $\widetilde{LM}_{1+2,T}$ test (though slightly undersized), and the $\widetilde{LMW}_{1+2,T}$ test (slightly oversized) control size fairly well and approach the nominal 5% significance level as the sample size grows. Table 2 (panels c and d) displays the power results of the two symmetric tests for a break in d and/or μ at $\lambda_0 = 0.5$.[14] Figures in bold characters correspond to size. Our simulation results confirm that there are some power gains from using the $\widetilde{LMW}_{1+2,T}$ tests in finite samples.[15] As can be inspected, power for both tests is increasing in the magnitude of the shifts in d and μ. For example, looking at the second block in panel d, for $\mu_0 = \mu_1 = 0$, a shift in d from 0 to 0.2 increases the power of the $\widetilde{LMW}_{1+2,T}$ test by 26.1 pp. (= 32.0 – 5.9), whereas looking at the second block, for $d_0 = d_1 = 0$, a shift in μ from 0 to 0.25 raises power by 30.8 pp. (= 36.7 – 5.9). The corresponding gains in power when d shifts to 0.4 (for $\mu_0 = \mu_1 = 0$) and when μ shifts to 0.5 (for $d_0 = d_1 = 0$) are 81.6 pp. and 81.3 pp., respectively. Finally, as expected, the power arising from breaks in μ is lower the higher d is. For instance, using the shift in μ from 0 to 0.25, this time with $d_0 = d_1 = 0.2$ instead of $d_0 = d_1 = 0$, only raises the power of the LMW-type test by 9.2 pp. (= 15.5 – 6.3).

[Table 2 about here]

Next, we report simulation results regarding size and power of the symmetric sup $\widetilde{LM}_{1+2,T}$ and sup $\widetilde{LMW}_{1+2,T}$ test**,** abstracting again from short-run dynamics**,** for the case of an unknown break fraction $\lambda \in \Lambda = [\epsilon, 1-\epsilon]$, with $\epsilon = 0.25$. Again, d_0 takes the values {–0.4,–0.3,–0.2,–0.1,0, 0.1,0.2,0.3,0.4}, while a non-breaking level of $\mu_0 = 0$ is considered for size, and $d_0 \in \{-0.2, 0, 0.2\}$, $d_1 \in$ {–0.4,–0.2,0, 0.2,0.4}, $\mu_0 = 0$ and $\mu_1 = \{0, 0.25, 0.5\}$ for power. Sample sizes are $T = 200, 500$ for size and $T = 200$ for power. Table 3 shows that both tests have satisfactory size properties for $T = 500$, though there are some small distortions for $T = 200$. In terms of power, again the $\widetilde{LMW}_{1+2,T}$ test often performs better, especially for larger values of d_0 and d_1.

[Table 3 about here]

[14] Notice that the dip in power in some of the entries is due to those cases where the alternative coincides with the null.

[15] We have checked whether the higher power of the $\widetilde{LMW}$-type tests relative to the $\widetilde{LM}$ test could be due to the differences in their effective sizes. We do so by computing size-corrected power and found that the power of the former test remains higher, though to a slightly lesser extent than when the nominal size is used.

Finally, we conduct a small Monte Carlo for the size and power properties of the two tests this time considering short-run dynamics, captured by an AR(1) process with parameter $\alpha_0 = 0$ under the null, which is potentially shifting at a known fraction $\lambda_0 = 0.5$. For the size we consider $T = 200, 500$ and 1000 and $\mu_0 = 0$, $d_0 \in \{-0.2, 0, 0.2\}$, and $\alpha_0 \in \{-0.5, 0, 0.5\}$ while for power we choose, $T = 200$ and $d_1 \in \{-0.4, -0.2, 0, 0.2, 0.4\}$, $\mu_1 \in \{0, 0.5\}$ and $\alpha_1 \in \{\alpha_0 - 0.3, \alpha_0, \alpha_0 + 0.3\}$. Table 4 (panel a) shows that size is relatively well controlled for the $\widetilde{LM}_{1+2,T}$ test. Table 5 (panel a) illustrates that the $LMW_{1+2,T}$ test is still oversized for $T = 200$ but, as T gets larger, the effective size becomes closer to the nominal size. Again panels b in both Tables illustrate that there can be some gains in power from using the $\widetilde{LMW}_{1+2,T}$ test, especially for positive autoregressive components.

[**Table 4 and 5 about here**]

6 Empirical Application: Analysis of structural changes in forward discount rates

In this section, we apply the proposed methodology to the analysis of the forward discount in exchange rate markets. As is well known, rational expectations and risk neutrality combined with covered and uncovered interest rate parity imply the forward exchange rate unbiasedness hypothesis (FRUH) whereby the forward rate, f_t, is an unbiased predictor of the future spot exchange rate, s_{t+1}, that is $E_t(s_{t+1}) = f_t$. In particular, FRUH corresponds to testing $H_0 : \delta_0 = 0$, $\delta_1 = 1$ in the following regression

$$\Delta s_{t+1} = \delta_0 + \delta_1 (f_t - s_t) + \varepsilon_{t+1}$$

where $(f_t - s_t)$ is the forward discount. This null has been often rejected and the typical finding is that the OLS point estimate of δ_1 is small or even negative (see, e.g., Engel, 1996 for an overview of this literature), leading to what is known as the *forward discount anomaly.* It has been argued that this result may be due to the unbalanced nature of the previous regression. In effect, while Δs_{t+1} is conventionally found to be $I(0)$, there is a large body of literature documenting that the forward discount $(f_t - s_t)$ follows a fractionally integrated $FI(d)$ process with d generally lying in the non-stationary range, $0.5 < d < 1$ (see, e.g., Baillie and Bollerslev, 1994, and Maynard and Phillips, 2001). However, Choi and Zivot (2007) have shown that previous estimates of d are likely to be upward biased when structural instabilities in the level of the forward discount series are not taken into account. Using the residuals of the forward discount monthly series for five G7 countries, these authors first adjust the level of these series for several structural breaks detected by means of Bai and Perron's (1998) methodology (BP hereafter). Next, they estimate d non parametrically (by means of Kim and Phillips's (2000) log-periodogram regression approach) from these break-adjusted series. Their main findings are that $(f_t - s_t)$ is subject to several breaks and that the resulting memory estimates turn out to be considerably lower than those previously estimated in the literature, with $0 < d < 0.5$.

Following this controversy, we provide here a short empirical application of our proposed tests using the forward discount data for five G7 countries examined by Choi and Zivot (2007). Their dataset includes monthly forward discount rates for the period 1976:1-1996:1 corresponding to the exchange rates in terms of US dollars for Germany, France, Italy, Canada and U.K., where f_t is defined as the log 30-day forward rate.

Figure 2 displays the five time series. Choi and Zivot (2007) find five breaks for Germany and U.K., four breaks for France and Italy and three breaks for Canada, which are displayed using dashed lines in Figure 2. However, when the testing procedure is reversed (i.e., first d is estimated from the time series of $(f_t - s_t)$ without allowing for level shifts, and then the BP procedure is applied to the filtered series $\Delta_t^d(f_t - s_t)$ to detect multiple breaks), the number of breaks is quite smaller (none for Germany and France, one for Italy, and three for Canada and U.K.). These contrasting results possibly reflect the shortcomings of using a two-step procedure rather than our proposed single-stage approach. Moreover, since we have argued that the level and the dynamics could shift simultaneously, so that both changes can be confused, it is possible that the single-step approach could yield more reliable estimates of the number of breaks and their origin.

To check this possibility, after fitting several alternative models, we consider an ARFIMA$(1, d, 0)$ model with a drift as the most appropriate for each of the five forward discount series, allowing for simultaneous breaks in all three parameters (d, α, μ) which are tested using our symmetric sup-$\widetilde{LM}_{1+2,T}$ and sup-$\widetilde{LMW}_{1+2,T}$ test statistics. To allow for multiple breaks (see the discussion in the next Section), we test sequentially 0 *vs.* 1 break and, upon rejection, 1 *vs.* 2, and so forth, to determine the number of breaks together with the break fractions.[16] Table 6 displays in Figure 2 the found breaks for each forward discount series for the LM (thick lines) and LMW-type (thin lines) tests. The results of the two tests are comparable, although the LMW-type test detects two more breaks than the LM test for Canada. In general, the breaking dates estimates gather around the early and mid- 1980s and early 1990s, possibly as a result of the creation of the European Exchange Rate Mechanism (ERM) in 1979 and the contractionary monetary stance in the U.S., and Canada in the early 1980, together with the collapse of the ERM in 1992. In comparison to Choi and Zivot (2007), we coincide in finding a break in the early 1980s for France, Italy and U.K. and early 1990s for U.K., Canada and Germany.

In order to provide further comparisons of our findings to Choi and Zivot (2007)'s, we briefly

[16] Notice that when we apply the sequential procedure to test 0 *vs.* 1 breaks, 1 *vs.* 2 breaks, etc., the critical values are the same as in Table 1. This is because, in principle, and after adjusting for sample sizes, it becomes only a problem of multiple testing.

apply the methodology in Subsection 3.2 to identify which parameters are shifting in each of the break points. Table 6 shows the results obtained from applying the robustified $\overline{LM}$ tests to detect the origin of the shifts in the five times series of forward discount rates. It becomes clear that while many breaks are indeed in all parameters - dynamics and level - some take place exclusively in the dynamics and only one exclusively in the level. Therefore, it seems that some of the breaks interpreted by Choi and Zivot (2007) as shifts in the level of forward discount rates, are instead due to shifts in their dynamics.

7 Conclusion and discussion

The starting point of this paper is to stress that the joint modeling of breaks in the dynamics and level of stochastic processes could be a relevant issue. By considering both breaks simultaneously, potential confounding problems about the sources of shifts in the persistence of a time-series process can be avoided. Our contribution here is to extend the well-known LM test for breaks only in the memory parameter to also account for breaks in the level as well as in the short-run dynamics. Furthermore, we propose a novel regression-based LMW-type test for $FI(d)$ processes with a drift that also accounts for these shifts. We also derive individual tests for constancy of a given parameter which are robust to the behaviour of the non-tested parameters. The proposed tests share several nice features. While LM tests are computationally attractive because they only require estimation under the null, LMW-type tests can exploit further information about the alternative, potentially leading to higher power without increasing computational complexity. Our Monte-Carlo simulations, based on some analytical results, show in particular that LMW-type tests for joint breaks can yield some power gains relative to LM tests in some instances. Finally, our empirical application on potential breaks in forward discount rates for several G7 countries provides new findings on the origin of these breaks (in dynamics and/or levels) in those time series which have been subject to considerable attention in the literature.

An additional advantage is that these tests can be easily extended to allow for the presence of multiple regimes, therefore allowing for breaks in d, α and μ at different periods of time. In this way, our maintained assumption that breaks are coincidental in time could be relaxed. We briefly sketch in the sequel how to implement the tests in this more general setup, where we consider for notational simplicity the case of a non-breaking $\alpha(L) = 1$.

Denoting the number of regimes by $i = 0, \ldots, m-1$, let us consider the following DGP

$$\Delta_t^{d_t}(y_t - \mu_t) = \varepsilon_t,\ t = [\lambda_i T] + 1, \ldots, [\lambda_{i+1} T],$$

with

$$\mu_t = \sum_{i=0}^{m-1} \mu_i R_t^{(i+1)}(\boldsymbol{\lambda})$$

where $R_t^{(i+1)}(\boldsymbol{\lambda}) = R_t^{(i+1)}(\lambda_i, \lambda_{i+1}) = 1\left([\lambda_i T] < t \leq [\lambda_{i+1} T]\right), \lambda_0 = 0, \lambda_m = 1, \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{m-1})'$. For example, in the case of testing for 0 *versus* 2 regimes (so that $m = 3$), the joint $\widetilde{LM}$ test is derived from the following likelihood function

$$\mathcal{L}_T(\psi, \boldsymbol{\lambda}) = -\frac{T}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t(\psi, \boldsymbol{\lambda})^2,$$

with $\psi = \left(\theta_1, \theta_2, \nu_1, \nu_2, d_0, \mu_0, \sigma^2\right)$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ and

$$\varepsilon_t(\psi, \boldsymbol{\lambda}) = \Delta_t^{d_0 + \theta_1 R_t^{(2)}(\boldsymbol{\lambda}) + \theta_2 R_t^{(3)}(\boldsymbol{\lambda})}\left(y_t - \mu_0 - \nu_1 R_t^{(2)}(\boldsymbol{\lambda}) - \nu_2 R_t^{(3)}(\boldsymbol{\lambda})\right)$$

where $R_t^{(2)}(\boldsymbol{\lambda}) = 1\left([\lambda_1 T] < t \leq [\lambda_2 T]\right)$, $R_t^{(3)}(\boldsymbol{\lambda}) = 1\left(t > [\lambda_2 T]\right)$. The LM test is constructed as in (3) and its asymptotic distribution turns out to be the sum of different terms related to the two breaks in memory and level, like in Theorem 1.

Likewise, to implement an LMW-type test when $m = 3$, the following regression model is used,

$$\begin{aligned}
\Delta_t^{d_0}(y_t - \mu_0) \quad = \quad & \left[\vartheta_1\left[\frac{1 - \Delta_t^{\theta_1}}{\theta_1}\right]\Delta_t^{d_0}(y_t - \mu_0) + \vartheta_2 \Delta_t^{d_0} 1\right] R_t^{(2)}(\boldsymbol{\lambda}) \\
& + \left[\vartheta_3\left[\frac{1 - \Delta_t^{\theta_2}}{\theta_2}\right]\Delta_t^{d_0}(y_t - \mu_0) + \vartheta_4 \Delta_t^{d_0} 1\right] R_t^{(3)}(\boldsymbol{\lambda}) + \varepsilon_t,
\end{aligned}$$

where a test of $H_0 : \vartheta_1 = \vartheta_2 = \vartheta_3 = \vartheta_4 = 0$ corresponds to testing for two breaks in both parameters, while testing $H_0' : \vartheta_1 = \vartheta_3 = 0$ (resp. $\vartheta_2 = \vartheta_4 = 0$) corresponds to testing for two breaks only in d (resp. μ).

References

[1] Abadir, K.M., Di Staso, W. and Giraitis, L. (2007). "Nonstationarity-extended local Whittle estimation", *Journal of Econometrics* 141, 1353-1384.

[2] Aue, A. and Horváth, L. (2013), "Structural breaks in time series". *Journal of Time Series Analysis*, 34, 1–16.

[3] Bai, J. and Perron, P. (1998) "Estimating and testing linear models with multiple structural changes". *Econometrica* 66, 47-78.

[4] Baillie, R.T and Bollerslev, T (1994) " The long memory of the forward premium". *Journal of International Money and Finance*, 13, 555-571.

[5] Beran, J. and Terrin, N. (1996)" Testing for a change of the long-memory parameter". *Biometrika* 83, 627-638.

[6] Breitung, J. and Hassler, U. (2002) "Inference on the cointegration rank in fractionally integrated processes," *Journal of Econometrics*, 110, 167-185.

[7] Busetti, F., and Taylor, A. M. R. (2004) "Tests of stationarity against a change in persistence". *Journal of Econometrics* 123, 33-66.

[8] Choi, K. and Zivot, E. (2007) " Long memory and structural changes in the forward discount: An empirical investigation," *Journal of Money an Finance*, 26, 342-363.

[9] Chung, C.F. and Baillie, R.T. (1993) "Small sample bias in conditional sum-of-squares estimators of fractionally integrated arma models". *Empirical Economics* 18, 791-806.

[10] Diebold, F. X. and Inoue, A. (2001) "Long memory and regime switching," *Journal of Econometrics* 105, 131-159

[11] Dolado, J. & Gonzalo, J. and L. Mayoral (2002) "A Fractional Dickey-Fuller test for unit roots," *Econometrica* 70, 1963-2006.

[12] Dolado, J.J., Gonzalo, J., and Mayoral, L. (2005) "What is what?: A simple time-domain test of long-memory vs. structural breaks". Economics Working Papers 954, Department of Economics and Business, Universitat Pompeu Fabra.

[13] Dolado, J.J., Gonzalo, J., and Mayoral, L. (2009) "Simple Wald tests of the fractional integration parameter: an overview of new results". In J. Castle & N. Shephard (eds.) *The Methodology and Practice of Econometrics*. A Festschift in Honour of David F. Hendry, 300-321, Oxford University Press.

[14] Dolado, J.J., Gonzalo, J., and Mayoral L. (2008) "Wald tests of I(1) against I(d) alternatives: some new properties and an extension to processes with trending components". *Studies in Nonlinear Dynamics & Econometrics* 12, 1-33.

[15] Dolado, J., Rachinger, H. and Velasco, C. (2017) "LM tests for joint breaks in the memory and level of a time series" shttps://sites.google.com/site/heikorachinger/DRV2017.

[16] Engel, E. (1996) "The forward discount anomaly and the risk premium: A survey of recent evidence". *Journal of Empirical Finance*, 3, 123-192.

[17] Engle, R. (1984). "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics". *Handbook of Econometrics*, vol II, ed. Griliches and Intrilligator (Amsterdam: North Holland, 1984), 775-826

[18] Gadea, M. D. and Mayoral, L. (2006). "The persistence of inflation in OECD countries: A fractionally integrated approach". *International Journal of Central Banking*, 2, 51-104.

[19] Gil-Alaña, L. A. (2008) "Fractional integration and structural breaks at unknown periods of time". *Journal of Time Series Analysis* 29, 163-185.

[20] Giraitis, L, Leipus, R. and Philippe, A. (2006) "A test for stationarity versus trends and unit roots for a wide class of dependent errors". *Econometric Theory*, 22.6, 989-1029.

[21] Granger, C. W. J. and Hyung, N. (2004) "Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns". *Journal of Empirical Finance* 11, 399-421.

[22] Granger, C. W. J. (1980) "Long memory relationships and the aggregation of dynamic models," *Journal of Econometrics.* 14(2), 227-238.

[23] Hassler, U. and Meller, B. (2014) "Detecting multiple breaks in long memory: the case of U.S. inflation". *Empirical Economics.* 46, 653-680.

[24] Hassler, U. and Scheithauer, J. (2011) "Detecting changes from short to long memory". *Statistical Papers* 52, 847-870.

[25] Horváth, L.and Shao, Q.-M. (1999) "Limit theorems for quadratic forms with applications to Whittle's estimate". The Annals of Applied Probability. 9, 146–187.

[26] Hsu, C.C. (2005) "Long memory or structural changes: An empirical examination on inflation rates". *Economics Letters* 88, 289-294

[27] Hualde, J. and Nielsen, M.Ø., (2017). "Truncated sum of squares estimation of fractional time series models with deterministic trends". Working Papers 1376, Queen's University, Department of Economics.

[28] Hualde, J. and Robinson, P.M. (2011) "Gaussian pseudo-maximum likelihood estimation of fractional time series models". *Annals of Statistics* 39, 3152-3181.

[29] Iacone, F., Leybourne, S. J., and Taylor, A.M.R. (2013) " Testing for a break in trend when the order of integration is unknown". *Journal of Econometrics*, 176, 30-45.

[30] Kim, J.-Y., Belaire-Franch, J., and Amador, R.B. (2002) "Corrigendum to detection of change in persistence of a linear time series". *Journal of Econometrics* 109, 389-392.

[31] Kuan, C.M. and Hsu, C.C. (1998) "Change-point estimation of fractionally integrated processes". *Journal of Time Series Analysis* 19, 693-708.

[32] Lavielle, M. and Moulines E. (2000) "Least squares estimation of an unknown number of shifts in a time series". *Journal of Time Series Analysis* 21, 33-59.

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

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

[35] Lobato, I. N. and Velasco, C. (2008) "Power comparison among tests for fractional unit roots". *Economics Letters* 99, 152-154.

[36] Maynard, A. and Phillips, P.C.B. (2001) " Rethinking an old puzzle: Econometric evidence on the forward discount anomaly" *Journal of Applied Econometrics*, 16, 671-708.

[37] Martins, L.F. and Rodrigues, P.M.M. (2014) "Testing for persistence change in fractionally integrated models: An application to world inflation rates". *Computational Statistics and Data Analysis*. 76(C), 502-522.

[38] McCloskey, A. (2010). "Semiparametric testing for changes in memory of otherwise stationary time series". Mimeo. January 2010.

[39] McCloskey, A. and Perron, P. (2013) "Memory parameter estimation in the presence of level shifts and deterministic trends" *Econometric Theory*, 29, 1196-1237.

[40] Mikosch, T. and Starica, C. (2004) "Changes of structure in financial time series and the GARCH model". Econometrics 0412003, EconWPA.

[41] Perron, P. and Qu, Z. (2010) "Long-memory and level shifts in the volatility of stock market return indices". *Journal of Business & Economic Statistics* 28, 275-290.

[42] Rachinger, H. (2017) "Multiple breaks in long memory time series". University of Vienna, mimeo. https://sites.google.com/site/heikorachinger/Rachinger2017.pdf

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

[44] Robinson, P.M. (1994) "Efficient tests of nonstationary hypotheses". *Journal of the American Statistical Association*, 89, 1420-1437

[45] Robinson, P.M. (2006) "Conditional-sum-of-squares estimation of models for stationary time series with long memory", *Time Series and Related Topics: In Memory of Ching-Zong Wei.* (H.-C. Ho, C.-K. Ing and T.L. Lai, eds.). IMS Lecture Notes - Monograph Series, 52, 130-137.

[46] Shao, X. (2011). "A simple test of changes in mean in the possible presence of long-range dependence". *Journal of Time Series Analysis* 32, 598-606.

[47] Sibbertsen, P. and Kruse, R. (2009) "Testing for a break in persistence under long-range dependencies". *Journal of Time Series Analysis* 30, 263-285.

[48] Wooldridge, J. M. (2002) *Econometric Analysis of Cross-Section and Panel Data.* Cambridge, MA. MIT Press.

[49] Yamaguchi, K. (2011) "Estimating a change point in the long memory parameter", *Journal of Time Series Analysis* 32, 304-314.

8 FIGURES AND TABLES

Table 1: **Critical Values of LM-tests for unknown λ for breaks in μ and d, and in μ, α and d.**

a) Critical Values of LM-tests for unknown λ for breaks in μ and d

$\epsilon \setminus d_0$	-0.49	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.49
0.15	10.9	10.9	11.1	11.2	11.3	11.8	12.4	13.4	15.0	17.2	19.1
0.2	10.5	10.5	10.6	10.7	11.0	11.3	11.9	12.9	14.6	16.7	18.6
0.25	10.0	10.1	10.1	10.3	10.5	10.8	11.5	12.4	14.1	16.3	18.1

b) Critical Values of LM-tests for unknown λ for breaks in d, α and μ

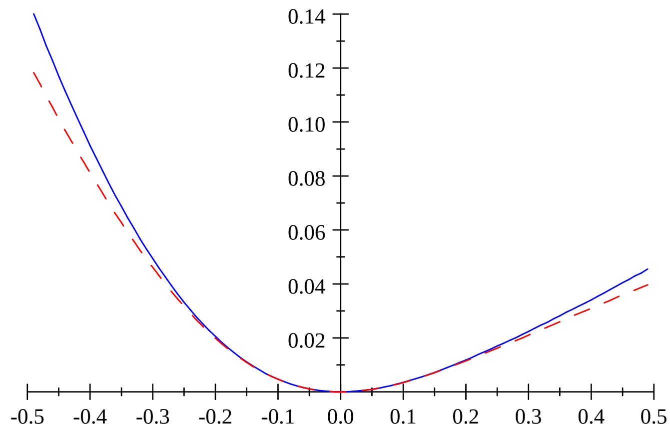
$\epsilon \setminus d_0$	-0.49	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.49
0.15	13.6	13.6	13.8	13.8	13.9	14.2	14.6	15.5	16.9	19.0	20.9
0.2	13.2	13.2	13.2	13.3	13.4	13.7	14.1	15.1	16.5	18.5	20.4
0.25	12.7	12.7	12.7	12.8	12.9	13.2	13.7	14.5	16.0	18.0	19.9

c) Critical Values of LM-tests for unknown λ for breaks in d, α_1,α_2 and μ

$\epsilon \setminus d_0$	-0.49	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.49
0.15	15.8	15.9	16.0	16.0	16.2	16.4	16.7	17.5	18.7	20.7	22.6
0.2	15.4	15.5	15.5	15.5	15.6	15.8	16.2	17.0	18.3	20.3	22.1
0.25	14.9	15.0	15.0	15.0	15.1	15.3	15.7	16.4	17.8	19.7	21.6

Figure 1: **Drift of the tests for a break in the memory**

a) Drift of the tests in the second regime as a function of the break magnitude $d_1 - d_0$. $\widetilde{LM}_2$ test (dashed line) and $\widetilde{LMW}_2$-type test (solid line) ($\lambda_0 = 0.5$, $d_0 = 0$).



b) Drift of the tests in the first regime as a function of the break magnitude $d_1 - d_0$. $\widetilde{LM}_1$ test (dashed line) and $\widetilde{LMW}_1$-type test (solid line) ($\lambda_0 = 0.5$, $d_0 = 0$).

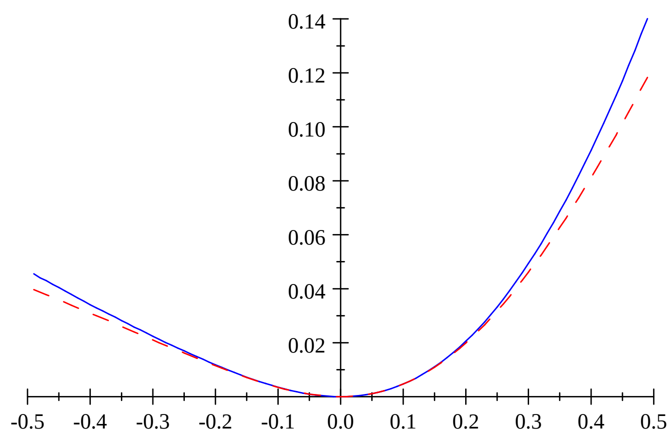


Table 2: **Simulated size and power of symmetric LM and LMW-type tests for a joint break in memory and level.**

a) Symmetric $\widetilde{LM}$ test: Size.

$T \ \backslash \ d_0:$	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4
200	4.9	4.7	4.2	4.4	4.6	4.0	4.5	4.8	6.3
500	5.1	4.8	4.7	4.8	4.7	4.3	4.2	5.1	5.8
1000	4.9	5.1	4.9	4.9	4.8	4.6	5.1	4.8	4.9

b) Symmetric $\widetilde{LMW}$-type test: Size.

$T \ \backslash \ d_0:$	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4
200	6.3	6.4	6.5	6.3	5.9	6.4	6.3	6.6	6.4
500	5.2	6.1	6.3	5.9	5.7	6.1	6.2	6.3	6.2
1000	5.1	5.2	5.3	5.2	5.3	5.2	5.3	5.4	5.4

Rejection probabilities of 5% test for joint break in d and μ at $\lambda_0 = 0.5,\ \mu_0 = 0,\ \sigma_0^2 = 1$.

c) Symmetric $\widetilde{LM}$ test: Power

d_0	-0.2				0				0.2			
$d_1 \ \backslash \ \mu_1$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
-0.4	36.3	95.9	100	100	86.4	92.2	97.7	100	99.6	98.4	95.9	97.0
-0.2	**4.2**	74.3	100	100	31.7	63.7	93.0	99.9	85.1	87.0	92.7	98.2
0	20.6	53.0	96.1	100	**4.6**	22.9	73.4	100	29.3	35.0	56.8	91.8
0.2	78.8	83.6	93.4	100	23.8	32.6	61.9	96.7	**4.5**	8.1	26.6	78.1
0.4	99.7	99.1	99.5	100	84.4	83.7	89.9	95.0	30.8	37.1	45.1	73.8

d) Symmetric $\widetilde{LMW}$-type test: Power

d_0	-0.2				0				0.2			
$d_1 \ \backslash \ \mu_1$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
-0.4	37.9	97.8	100	100	88.1	93.0	98.6	100	99.7	98.7	96.4	97.7
-0.2	**6.5**	87.8	100	100	36.9	71.8	96.9	100	88.7	89.8	95.6	99.5
0	30.8	74.4	98.9	100	**5.9**	36.7	87.2	100	36.3	46.5	72.6	98.7
0.2	86.5	90.1	97.7	100	32.0	47.0	75.2	99.2	**6.3**	15.5	40.7	89.1
0.4	99.7	99.3	99.9	100	87.5	86.5	92.8	96.5	34.9	41.6	48.3	76.6

Rejection probabilities of 5% test for joint break in d and μ at $\lambda_0 = 0.5$, $\mu_0 = 0$, $\sigma_0^2 = 1$, $T = 200$. Bold numbers correspond to size simulations.

Table 3: **Simulated size and power for symmetric LM and LMW-type tests for a joint break in memory and level for an unknown break fraction.**

a) Symmetric $\widetilde{LM}$ test: Size

$T \setminus d_0$	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4
200	7.0	7.1	6.5	6.4	4.8	4.9	5.4	5.3	6.0
500	6.2	5.6	5.6	5.1	4.9	4.7	4.8	4.8	5.2

b) Symmetric $\widetilde{LMW}$-type test: Size

$T \setminus d_0$	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4
200	8.3	7.4	7.1	7.3	6.9	7.0	6.2	5.8	5.7
500	6.5	6.0	6.5	5.9	6.3	6.1	6.0	5.4	5.1

Rejection probabilities of 5% test for joint break in d and μ, $\mu_0 = 0$, $\sigma_0^2 = 1$.

c) Symmetric $\widetilde{LM}$ test: Power

d_0	-0.2				0					0.2			
$\mu_1 \setminus d_1$	-0.4	-0.2	0	0.2	-0.4	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4
0	37.2	**6.5**	14.7	58.4	85.6	33.1	**4.8**	13.9	64.8	78.4	24.6	**5.4**	19.8
0.25	88.3	47.0	27.4	62.4	90.9	54.0	14.3	18.9	67.7	83.5	27.7	5.6	20.9
0.5	99.9	95.3	70.2	75.1	98.5	82.6	37.8	33.1	68.3	83.7	37.5	8.0	24.2

d) Symmetric $\widetilde{LMW}$-type test: Power

d_0	-0.2				0					0.2			
$\mu_1 \setminus d_1$	-0.4	-0.2	0	0.2	-0.4	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4
0	35.5	**7.1**	15.7	58.9	86.0	37.9	**6.9**	17.4	68.8	83.2	32.5	**6.2**	20.7
0.25	93.9	56.3	27.5	59.3	93.1	65.1	23.6	23.5	71.3	88.6	42.5	10.4	23.1
0.5	99.8	95.8	57.2	64.9	99.5	94.9	57.8	38.5	72.2	93.3	63.4	20.0	26.6

Rejection probabilities of 5% test for joint break in d and μ, $\epsilon = 0.25, \lambda_0 = 0.5$, $\mu_0 = 0$, $\sigma_0^2 = 1$, T=200. Bold numbers correspond to size simulations.

Table 4: **Simulated size and power of the symmetric LM test for a joint break in memory, level and autoregressive component.**

a) Symmetric $\widetilde{LM}$ test: Size

α_0	-0.5			0			0.5		
$T \setminus d_0$	-0.2	0	0.2	-0.2	0	0.2	-0.2	0	0.2
200	7.8	7.5	6.2	7.3	6.9	5.9	7.1	7.5	4.9
500	7.3	6.5	5.6	6.9	6.4	4.6	7.4	6.5	5.1
1000	7.4	6.0	6.1	7.0	6.0	5.7	7.3	6.4	5.0

Rejection probabilities of 5% test for joint break in d, α and μ at $\lambda_0 = 0.5,\ \mu_0 = 0$, $\sigma_0^2 = 1$.

b) Symmetric $\widetilde{LM}$ test: Power

		d_0	-0.2				0					0.2			
μ_1	α_0	$\alpha_1\backslash\ d_1$	-0.4	-0.2	0	0.2	-0.4	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4
		-0.8	96.6	71.5	38.3	49.6	99.9	95.1	70.5	39.5	54.7	99.9	95.2	69.8	43.2
0	-0.5	-0.5	37.4	**7.8**	34.4	85.1	87.8	36.6	**7.5**	37.6	89.4	85.7	31.5	**6.2**	41.8
		-0.2	36.5	60.0	93.5	99.9	61.1	33.2	61.1	94.1	99.8	55.5	30.0	60.9	95.1
		-0.3	92.4	48.5	10.2	22.3	99.0	90.8	46.4	10.7	28.9	98.9	90.5	43.7	11.5
0	0	0	36.0	**7.3**	30.5	81.8	77.7	34.9	**6.9**	32.1	86.5	79.0	29.9	**5.9**	37.1
		0.3	18.4	47.4	90.0	99.4	28.2	16.1	44.5	89.1	98.8	28.7	14.4	47.3	91.3
		0.2	77.8	36.8	8.5	10.6	89.6	79.0	37.5	6.6	14.2	90.2	78.1	34.9	8.6
0	0.5	0.5	16.6	**7.1**	15.2	67.1	63.2	18.4	**7.5**	18.9	73.8	60.3	14.9	**4.9**	26.3
		0.8	10.2	45.4	89.6	98.8	7.9	9.5	51.9	92.7	99.1	7.3	11.2	57.9	94.2
		-0.8	100	99.9	96.6	96.3	100	100	98.9	89.0	79.6	100	99.4	89.9	65.5
0.5	-0.5	-0.5	100	98.6	97.8	99.6	98.9	96.7	86.5	78.7	93.2	91.6	66.8	39.0	58.5
		-0.2	100	100	100	100	97.5	96.1	95.3	98.5	100	78.1	69.4	79.6	96.6
		-0.3	99.9	98.9	86.9	65.8	99.7	98.4	84.4	44.3	46.3	99.4	94.7	60.0	25.7
0.5	0	0	97.8	88.8	81.1	91.7	93.6	77.1	46.1	51.7	89.2	84.0	46.6	18.6	47.8
		0.3	92.8	91.9	97.0	99.6	70.6	61.5	73.4	93.0	98.9	44.7	30.1	60.8	91.9
		0.2	96.2	75.0	28.7	23.6	95.1	89.2	53.6	13.2	20.6	92.7	83.5	42.4	12.6
0.5	0.5	0.5	56.1	21.6	30.8	72.7	71.3	30.9	10.0	26.3	75.7	65.0	21.2	6.7	30.5
		0.8	23.1	56.9	90.0	98.6	15.4	15.1	52.1	92.8	98.8	12.4	15.1	57.4	93.9

Rejection probabilities of 5% test for joint break in d, α and μ at $\lambda_0 = 0.5,\ \mu_0 = 0$, $\sigma_0^2 = 1, T = 200$. Bold numbers correspond to size simulations.

Table 5: **Simulated size and power of the symmetric LMW-type test for a joint break in memory, level and autoregressive component.**

a) Symmetric $\widetilde{LMW}$-type test: Size

α_0	-0.5			0			0.5		
$T \setminus d_0$	-0.2	0	0.2	-0.2	0	0.2	-0.2	0	0.2
200	6.2	6.7	7.2	6.8	6.7	6.0	6.6	7.2	7.0
500	6.2	6.0	6.8	6.1	6.3	6.3	6.4	7.0	6.7
1000	5.9	5.8	6.0	5.6	6.1	6.2	6.0	6.1	5.9

Rejection probabilities of 5% test for joint break in d, α and μ at $\lambda_0 = 0.5,\ \mu_0 = 0, \sigma_0^2 = 1$.

b) Symmetric $\widetilde{LMW}$-type test: Power

		d_0	-0.2				0					0.2			
μ_1	α_0	$\alpha_1 \setminus d_1$	-0.4	-0.2	0	0.2	-0.4	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4
		-0.8	91.4	61.9	45.4	59.9	99.8	92.2	62.8	49.0	73.2	99.8	92.8	63.6	55.1
0	-0.5	-0.5	34.4	**6.2**	20.2	69.5	87.1	36.2	**6.7**	24.2	78.9	86.9	33.7	**7.2**	29.5
		-0.2	38.0	47.3	85.3	99.4	69.2	38.0	48.8	87.3	99.5	69.5	35.5	49.4	89.4
		-0.3	87.5	39.5	11.8	24.1	99.7	85.9	38.8	13.2	37.4	99.5	86.5	37.5	19.6
0	0	0	35.3	**6.8**	26.0	78.2	85.6	32.5	**6.7**	33.0	85.2	85.6	31.6	**6.0**	39.7
		0.3	27.2	50.4	91.3	99.8	48.6	26.6	52.2	92.3	99.8	47.1	24.6	51.6	91.8
		0.2	92.2	46.3	11.7	24.1	99.9	89.9	46.1	13.2	30.1	99.7	90.9	46.3	16.9
0	0.5	0.5	32.0	**6.6**	29.2	81.3	85.1	29.9	**7.2**	30.8	83.1	83.7	29.9	**7.0**	36.5
		0.8	20.5	62.2	97.4	100	17.1	18.4	65.2	97.4	100	17.3	20.2	69.9	97.5
		-0.8	100	100	96.6	67.6	100	100	98.8	86.3	81.8	100	99.7	92.4	75.2
0.5	-0.5	-0.5	100	100	93.6	85.3	99.6	99.6	92.6	71.6	86.8	96.8	88.2	52.0	53.5
		-0.2	100	100	99.5	99.9	99.4	99.3	96.4	97.1	99.7	94.3	84.8	78.3	92.4
		-0.3	99.9	99.2	81.2	59.3	99.4	98.8	85.4	43.9	52.0	99.5	95.1	61.5	36.0
0.5	0	0	97.0	91.7	85.1	91.4	92.4	84.5	61.1	57.0	88.9	89.5	63.5	33.1	51.2
		0.3	94.0	95.9	98.5	99.9	76.3	73.1	82.1	95.3	99.9	67.0	50.4	66.3	93.2
		0.2	95.5	81.3	54.4	51.4	98.8	94.5	62.8	27.8	37.4	99.8	93.2	54.3	22.4
0.5	0.5	0.5	69.7	55.7	62.2	88.2	82.2	53.7	26.8	41.9	84.6	87.2	41.6	15.2	40.0
		0.8	58.1	81.9	97.0	100	34.8	34.6	70.8	97.3	100	24.8	26.9	69.4	97.5

Rejection probabilities of 5% test for joint break in d, α and μ at $\lambda_0 = 0.5,\ \mu_0 = 0, \sigma_0^2 = 1, T = 200$.
Bold numbers correspond to size simulations.

Figure 2: **Forward discount series**

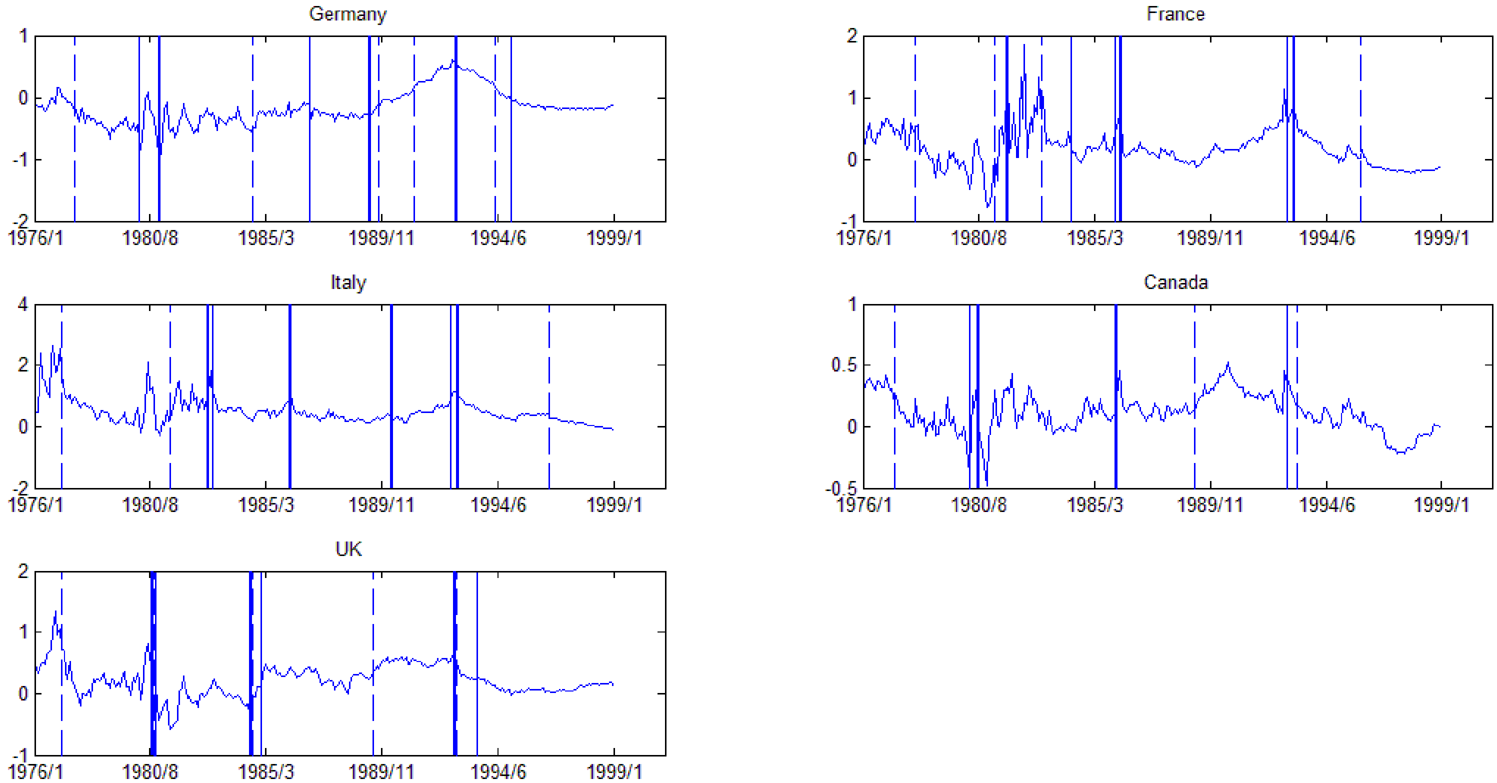


Table 6: **Breaks in the forward discount series**

a) LM test

Country	number of breaks	break dates
Canada	2	8/1980, 2/1986
France	3	10/1981, 4/1986, 3/1993
Germany	3	1/1981, 5/1989,10/1992
Italy	4	10/1980, 6/1984, 5/1988, 8/1992
U.K.	3	9/1980, 8/1984, 9/1992

b) LMW-type test

Country	number of breaks	break dates
Canada	4	4/1980, 2/1986, 6/1989, 11/1992
France	3	4/1983, 1/1986, 11/1992
Germany	4	3/1980, 12/1986, 6/1989, 12/1994
Italy	3	2/1983, 3/1986, 7/1992
U.K.	3	11/1980, 1/1985, 8/1993

Table 7: **Detection of breaks in the forward discount series**

Country	break dates			
Canada	8/1980: (d,α)	2/1986: (d,α)		
France	10/1981: (d,α)	4/1986: (d,α)	3/1993: $(d,\alpha)\,,\mu$	
Germany	1/1981: (d,α)	5/1989: $(d,\alpha)\,,\mu$	10/1992: $(d,\alpha)\,,\mu$	
Italy	12/1982: (d,α)	3/1986: μ	3/1990: $(d,\alpha)\,,\mu$	11/1992: $(d,\alpha)\,,\mu$
U.K.	9/1980: (d,α)	8/1984: $(d,\alpha)\,,\mu$	9/1992: (d,α)	