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

Juan J. Dolado

European University Institute

Heiko Rachinger

University of Vienna

Carlos Velasco†

Universidad Carlos III de Madrid

This version: December, 2016.

Abstract. We consider two easily implementable versions of the Lagrange Multiplier (LM) test for joint breaks in the memory and the level of a time-series process: (i) a conventional LM test with parameters estimated under the null of no breaks, and (ii) a regression-based test which also uses some information under the alternative (in the spirit of a Wald test), labeled as LMW-type tests. Together with tests of individual breaks robustified to allow for potential changes in the non-tested parameter, these tests contribute to solve a potential confounding problem regarding the sources of breaks. We derive their asymptotic distributions for known and unknown breakpoints, both under the null and a local break hypothesis, as well as compare their power properties under fixed alternatives. We find that a LMW-type test has better power, especially when the alternative involves a break in memory. In addition, we show that these tests can be easily modified to allow for potentially breaking short-run dynamics. Monte Carlo simulations confirm that the proposed tests behave satisfactorily in finite samples.

JEL Classification: C13, C22

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

*We are very grateful to Robinson Kruse, Laura Mayoral and Peter Robinson for very detailed comments, as well as to Juan-Carlos Escanciano, Jesús Gonzalo, Uwe Hassler, Abderrahim Taamouti and participants in several seminars and workshops for helpful suggestions on earlier drafts of the paper. 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 they are truly driven by fractionally integrated processes of order d, FI(d), or is long memory spuriously generated by breaks in the levels of the series.[1] 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 mean, μ (McCloskey, 2010).

This debate has led to two strands in the literature on this topic. The first one has focused on testing for breaks in d. Motivated by the popular rationalization of FI(d) processes on the basis 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 in industry composition of the economy. As a result, the long-memory properties of relevant aggregates (e.g., inflation, unemployment, GDP, squared financial returns, etc.) are likely to have experienced changes over relevant subsamples (see, e.g. Gadea and Mayoral, 2005). Accordingly, several tests have been proposed in semiparametric and parametric setups for the null of a constant value of d against the alternative of breaks at known or unknown dates.[2] In contrast, there is another strand of the literature which has developed tests for breaks in the level of a time series with stationary long-memory disturbances, but without allowing for breaks in d.[3]

Nevertheless, seemingly less attention has been paid to considering potential *joint* breaks in both d and μ. Whenever a break is detected, this joint test helps identify whether it originates in only one or in both parameters.[4] Among the scant literature which has considered joint breaks, the

[1] See, e.g., Lobato and Savin (1998), Mikosch and Starica (2004), and Perron and Qu (2010).

[2] See, e.g., Beran and Terrin (1996 and 1999), Hassler and Meller (2014), Hassler and Scheithauer (2011), Martins and Rodrigues (2014), McCloskey (2010), and Sibbertsen and Kruse (2009). Forerunners of these tests are the approaches proposed by Kim et al. (2002) and Busetti and Taylor (2004) to test for $I(0)$ series against alternatives where there is a change from $I(0)$ to $I(1)$. Likewise, Harvey et al. (2006) and Leybourne et al. (2007) develop tests for constant persistence (the series can be either $I(0)$ or $I(1)$ without breaks) against a change in an unknown direction (i.e. I(0) to I(1) or vice versa). Multiple changes are tackled in Leybourne et al. (2007) and Kejriwal et al. (2013).

[3] See, e.g., Kuan and Hsu (1998), Lavielle and Moulines (2000), and Shao (2011).

[4] 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 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

following works are the most closely related to ours. Gil-Alaña (2008) is the first attempt 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 sequential way. Initially, the location of the mean break is detected using Hsu's (2005) semiparametric testing approach; next, the original time series is demeaned (with a broken intercept) 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 may 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. More recently, some of these shortcomings have been addressed in Rachinger (2016), 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 distribution of the parameter estimates under different sources of break magnitudes 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 since parameter estimation is only required under the null. However, we differ from these authors in deriving a single-step procedure, rather than a sequential one. In addition, given that LM tests often lack power relative to Wald tests. Yet, since the latter are more difficult to implement, we propose another test statistic which combines the computational advantage 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 denoted in the sequel as "LMW-type" tests.

Such tests have been proposed by Lobato and Velasco (2007) and Dolado et al. (2009) 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 breaks 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

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.

of Rachinger´s (2016) 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 while LR tests require good performance of the constrained and unconstrained estimators of d under the null and the alternative. Lastly, an additional advantage of our proposed tests is that, under a parametric setup, they can also be easily amended to deal with breaks in short-memory parameters, on top of breaks in memory and level.[5]

Overall, this paper makes several contributions to the relevant literature on the origin of breaks in persistent time-series processes. First, we derive 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 a non-stable memory and/or level parameter. Both tests are easy to compute under the joint null and have similar asymptotic behaviour under the null and local alternatives. However, we illustrate both analytically and in finite-sample simulations that LMW-type tests could lead to power gains under fixed alternatives, especially when they involve a break in d.[6] Second, we propose robustified versions of the two proposed tests for the individual null of a break in only one of the parameters, irrespective of whether the non-tested parameter breaks or not. Third, under the alternative of breaks in either only one or both parameters, we propose a sequential testing procedure which consistently identifies the source of the break. This procedure relies on using both tests for joint breaks in d and μ, followed by the corresponding robustified individual tests for a single break in either d or μ. Fourth, we argue that an average of either LM or LMW-type tests computed under the first and second regimes (i.e., before and after the break point) can yield power gains relative to the individual tests. Fifth, we show that both testing procedures can be easily extended to deal with possible autoregressive dynamics whose parameters could break as well. Lastly, we briefly discuss how to extend the previous tests when breaks in d and μ may not be coincidental in time.

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 and fixed alternatives. We distinguish among different setups: known and unknown break date, change in only one of the parameters and simultaneous changes in both. In Section 5, we extend these tests to allow for potentially breaking short-run dynamics as well. In Section 6, we compare their non-centrality/drift parameters under fixed alternatives, showing that there is an asymptotic inequality among them. In Section 7, we provide simulation results regarding the finite-sample performance of the tests. Finally, in

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

[6] 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).

Section 8, 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 Data Generation Process

For simplicity, we start dealing with the case of a *single* breakpoint (at a *known* or *unknown* date) which changes in the asymptotics as a proportion λ_0 of the sample size. In particular, we start by assuming that the time series 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 $t = [T\lambda_0] + 1, \ldots, T$.[7] 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 parameter λ_0 denotes the true break fraction which lies in the interval $\Lambda = [\epsilon, 1-\epsilon]$, where $\epsilon > 0$ is assumed to be known. The following transition model is considered in the sequel,

$$\alpha_0 (L) \Delta_t^{d_t} (y_t - \mu_t) = \varepsilon_t, \quad t = 1, 2, \ldots, \tag{1}$$

such that

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

where $1(\cdot)$ is an indicator function of the relevant subsample, $\alpha_0 (L) = 1 - \alpha_{1,0} L \cdots - \alpha_{p,0} L^p$ is a stable autoregressive lag polynomial of order p with unknown coefficients, and $\Delta_t^b := \sum\limits_{k=0}^{t-1} \pi_k(b) L^k$, with $\pi_k (b) := \frac{\Gamma(k-b)}{\Gamma(-b)\Gamma(k+1)}$, $k = 0, 1, \ldots,$ denotes the (truncated or Type-II) fractional-differencing filter.

Remark 1: Notice that the previous definition of $\Delta_t^{d_t}$ implies that the filter applied to $(y_t - \mu_t)$ is $\sum_{j=0}^{t-1} \pi_j (d_0)$ when $t < [T\lambda_0]$ and $\sum_{j=0}^{t-1} \pi_j (d_1)$ when $t > [T\lambda_0]$. We prefer to use a truncated Type II filter, rather than an untruncated Type I filter, because the subsequent proofs become quite more involved under the latter. The reason why our procedure works is because it applies a truncated differencing filter to invert an equally truncated integration filter.[8]

Remark 2: Notice also that, by rewriting the DGP recursively in terms of the truncated filter

[7] Remark 3 below includes further discussion about the implementation of this test when $d_0, d_1 > 0.5$.

[8] At any rate, Hualde and Robinson (2011) have shown that both filters lead to asymptotically equivalent memory estimators, though the choice of filter may affect the estimator of the level.

defined as $\alpha_0(L)\Delta_t^d := \sum_{j=0}^{t-1}\pi_j^*(d)L^j$, it follows that

$$
\begin{aligned}
y_t &= \mu_0 + \left(1 - \alpha_0(L)\Delta_t^{d_0}\right)(y_t - \mu_0) + \varepsilon_t, \quad \text{for } t \leq [T\lambda_0] \\
y_t &= \mu_1 + \left(1 - \alpha_0(L)\Delta_t^{d_1}\right)(y_t - \mu_t) + \varepsilon_t \\
&= \mu_1 - \sum_{j=1}^{t-[T\lambda_0]-1}\pi_j^*(d_1)\{y_{t-j} - \mu_1\} - \sum_{j=t-[T\lambda_0]}^{t-1}\pi_j^*(d_1)\{y_{t-j} - \mu_0\} + \varepsilon_t, \quad \text{for } t > [T\lambda_0],
\end{aligned}
$$

so that the chosen filter guarantees that the lags of y_t in the autoregression are centered around the appropriate value of μ_t. As a result, our filtering procedure yields the correct level of the time series for every $t = 1, \ldots, T$.[9]

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

$$
\alpha_0(L)\Delta_t^{d_t}(y_t - \mu_t - \beta_t t) = \varepsilon_t,
$$

$\beta_t = \beta_0 1(t \leq [T\lambda_0]) + \beta_1 1(t > [T\lambda_0])$, by applying our testing procedure to the first-differenced data to test for breaks in the trend slope β_t and in the memory $d_t - 1$ of the increments.

Remark 4: Although our analysis allows for breaks in the short-memory parameters $\alpha_0(L)$, for the time being we will assume that they are stable and therefore only consider breaks in d and μ. For simplicity, we start with the simplest case where $\alpha_0(L) = 1$ and later consider the case of short memory with stable or breaking autoregressive parameters. We do this for ease of exposition and also because the confounding problem in practice involves d and μ. Yet, as shown further below later on in Section 5, the generalization to a break in $\alpha_0(L)$ is rather straightforward.

In sum, denoting the size of potential shifts in the memory by $\theta_0 = d_1 - d_0$ and in the level by $\nu_0 = \mu_1 - \mu_0$, respectively, and labeling 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 (with $\alpha_0(L) = 1$ unless otherwise indicated)

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

[9] Notice that, unlike the arguments in favour of re-initializing the process in the second regime when there is a shift from I(1) to I(0) in an AR(1) process (see Kejriwal and Perron, 2012), our setup is not amenable to do so. This is because FI(d) processes are AR(∞) rather than finite-lag. As a result, while it makes sense to allow for an abrupt change in the level, as is done here, allowing for such a change in persistence is not sensible. In other words, our DGP in (1) assumes that the memory parameter is the fractional difference which achieves an I(0) process, namely d_0 before the breakpoint and d_1 afterwards. It does so by applying the same filter $\Delta_t^{d_t}$ without resorting to different ad-hoc truncations. For example, if we were to define $y_t = \mu_1 + \Delta_{t-[T\lambda_0]}^{-d_1}\varepsilon_t$, $t = [T\lambda_0]+1, ..., T$, this filter would introduce a discontinuity even under the null with $d_0 = d_1$ and $\mu_0 = \mu_1$.

where, to derive the properties of the tests, we focus on the case where the indicator variable in (2) is $R_t^{(2)}$ (implementation of the test in the *second regime*), while the consequences of using $R_t^{(1)}$ (*first regime*) 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. Notice that, when testing in the *first regime*, d_1 is the persistence of the first regime and d_0 is the one of the second regime, leading to symmetric, but not equivalent tests.

3 LM Tests

According to the LM principle, we test the null

$$H_0 : \theta_0 = \nu_0 = 0. \tag{H0}$$

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

$$H_1(\lambda_0) : \theta_0 \neq 0 \text{ and/or } \nu_0 \neq 0. \tag{H1}$$

From (2), with $\alpha_0(L) = 1$, and under the conventional assumption that $\varepsilon_t \sim NID\left(0, \sigma^2\right)$, the log-likelihood function, $\mathcal{L}_T$, can then be written as

$$\mathcal{L}_T\left(\theta, \nu, d, \mu, \sigma^2, \lambda\right) = -\frac{T}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t\left(\psi, \lambda\right)^2, \tag{3}$$

for every possible breakpoint λ, and $\psi = \left(\theta, \nu, d, \mu, \sigma^2\right)'$. Using the second-regime indicator $R_t^{(2)}$, we define the residuals

$$\varepsilon_t\left(\psi, \lambda\right) = \Delta_t^{d+\theta R_t^{(2)}(\lambda)}\left(y_t - \mu\right) - \nu\Delta_t^{d+\theta R_t^{(2)}} R_t^{(2)}\left(\lambda\right).$$

For each λ, the LM test is based on the derivatives of $\mathcal{L}_T = \mathcal{L}_T\left(\psi, \lambda\right)$ in the direction of ψ evaluated at the restricted estimates $\tilde{\psi}_T = (0, 0, \tilde{d}_{0T}, \tilde{\mu}_{0T}, \tilde{\sigma}_T^2)'$,

$$\widetilde{LM}_{2,T}(\lambda) = \left.\frac{\partial\mathcal{L}_T\left(\psi, \lambda\right)}{\partial\psi'}\right|_{\psi=\tilde{\psi}_T}\left(-\left.\frac{\partial^2\mathcal{L}_T\left(\psi, \lambda\right)}{\partial\psi\partial\psi'}\right|_{\psi=\tilde{\psi}_T}\right)^{-1}\left.\frac{\partial\mathcal{L}_T\left(\psi, \lambda\right)}{\partial\psi}\right|_{\psi=\tilde{\psi}_T}. \tag{4}$$

where the subscript 2 in $\widetilde{LM}_{2,T}(\lambda)$ indicates implementation of the test in the *second regime*.

In particular, the score in the directions of θ and ν can be expressed as,

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

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

$$\tilde{\varepsilon}_t = \varepsilon_t \left(\tilde{\psi}_T \right) = \Delta_t^{\tilde{d}_{0T}} \left(y_t - \tilde{\mu}_{0T} \right), \quad t = 1, 2, \ldots, T, \tag{5}$$

while the corresponding variance estimator is

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

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

$$(\tilde{d}_{0T}, \tilde{\mu}_{0T}) = \arg \min_{d \in D, \mu \in M} \sum_{t=1}^{T} \left(\Delta_t^d \left(y_t - \mu \right) \right)^2. \tag{7}$$

The properties of $\tilde{d}_{0T}$ for known μ have been discussed, *inter alia*, in Chung and Baillie (1993), Robinson (2006) and Hualde and Robinson (2011), while Rachinger (2016) extends their analysis to cover simultaneous level estimation. The estimator $\tilde{d}_{0T}$ is $T^{1/2}$-consistent and asymptotically normal for $d_0 \in Int(D)$, for compact $D \subset (-0.5, 0.5)$ of range smaller than 0.5, while $\tilde{\mu}_{0T}$ is $T^{1/2-d_0}$-consistent.

The (relevant block of the) inverse of the Hessian matrix can be approximated by

$$\left[- \frac{\partial^2 \mathcal{L}_T(\psi, \lambda)}{\partial \psi \partial \psi'} \bigg|_{\psi = \tilde{\psi}_T} \right]_{[1:2,1:2]}^{-1} = \left(\tilde{\Lambda}_2^{1/2} \begin{pmatrix} \frac{1}{\tilde{\sigma}_T^2} \sum\limits_{t=[\lambda T]+1}^{T} (\log \Delta_t \tilde{\varepsilon}_t)^2 & 0 \\ 0 & \frac{1}{\tilde{\sigma}_T^2} \sum\limits_{t=[\lambda T]+1}^{T} (\Delta_{t-[\lambda T]}^{\tilde{d}_{0T}} 1)^2 \end{pmatrix} \tilde{\Lambda}_2^{1/2} \right)^{-1} (1 + o_p(1)),$$

where the matrix of scale factors

$$\tilde{\Lambda}_2 = \Lambda_2 \left(\lambda; \tilde{d}_{0T} \right) = \begin{pmatrix} \lambda & 0 \\ 0 & \frac{L_T\left(\tilde{d}_{0T}; \lambda, \lambda\right)}{L_T\left(\tilde{d}_{0T}; \lambda, \lambda\right) - L_T^2\left(\tilde{d}_{0T}; 0, \lambda\right)} \end{pmatrix}, \tag{8}$$

with $L_T(d; a, b) = T^{2d-1} (1 - 2d) \Gamma^2 (1 - d) \sum_{t=[\max(a,b)T]+1}^{T} (\Delta_{t-[aT]}^{d} 1)(\Delta_{t-[bT]}^{d} 1)$, reflects the estimation effects of d_0 and μ_0. Lastly, notice that the subscript 2 in matrix $\tilde{\Lambda}_2$ above indicates that we test in the *second regime.*

Remark 5: When the LM test is implemented in the *first regime*, denoted as $\widetilde{LM}_{1,T}(\lambda)$, the derivations are similar, except that the sums in the score in the directions of θ and ν go from 1 to $[\lambda T]$ and the scaling factor $\tilde{\Lambda}_2$ is replaced by

$$\tilde{\Lambda}_1 = \Lambda_1(\lambda, \tilde{d}_{0T}) = \begin{pmatrix} 1 - \lambda & 0 \\ 0 & \frac{1 + L_T\left(\tilde{d}_{0T}; \lambda, \lambda\right) - 2L_T\left(\tilde{d}_{0T}; 0, \lambda\right)}{L_T\left(\tilde{d}_{0T}; \lambda, \lambda\right) - L_T^2\left(\tilde{d}_{0T}; 0, \lambda\right)} \end{pmatrix}.$$

Finally, the unknown break fraction is determined by

$$\tilde{\lambda}_T = \arg \max_{\lambda \in \Lambda} \widetilde{LM}_{2,T}(\lambda), \tag{9}$$

where $\Lambda = [\epsilon, 1 - \epsilon], \ 0 < \epsilon < 1/2.$

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 $q > \max\{4, \frac{2}{1-2d_0}\}$.

Assumption 2. $d_0 \in Int\left(D\right),\ D = \left[\underline{d}, \bar{d}\right]$, $-0.5 < \underline{d} < \bar{d} < 0.5$, and $\mu_0 \in Int\left(M\right),\ \lambda_0 \in \Lambda$.

Assumption 2a. $d_0, d_1 \in Int\left(D_1\right),\ D_1 = [\underline{d}, 0]$, $-0.5 < \underline{d} < 0$, and $\mu_0, \mu_1 \in Int\left(M\right),\ \lambda_0 \in \Lambda$.

Assumption 2b. $d_0, d_1 \in Int\left(D_2\right),\ D_2 = \left[0, \bar{d}\right],\ 0 < \bar{d} < 0.5$ and $\mu_0, \mu_1 \in Int\left(M\right),\ \lambda_0 \in \Lambda$.

REMARK 6: The restriction that d belongs to either the D_1 or D_2 sets in Assumptions 2a and 2b ensures that the maximum distance (in absolute value) between d_0 and d_1 is less than 0.5 under fixed alternatives, allowing to establish a proper probability limit of the estimates of d in this case. Moreover, assuming that d_0 and d_1 share the same sign (negative in D_1 and positive in D_2) avoids having to characterize a variety of memory transitions that may affect the rate of divergence of the test statistics, leading to radically different persistence properties in the two regimes.[10]

To assess the asymptotic null distribution and local power of the LM test, we analyze its properties under the following local break alternatives,

$$H_{1T}(\lambda_0) : \theta_0 = \delta/T^{1/2}, \nu_0 = \eta/T^{1/2-d_0}, \tag{10}$$

for some $\lambda_0 \in \Lambda$, where the convergence rate in the direction of the level differs from the standard $T^{1/2}$ rate, and the null is recovered by setting $\delta = \eta = 0$ while leaving λ_0 unspecified.

We next derive the asymptotic distribution of the $\widetilde{LM}_{2,T}$ test in (4) in the case of an unknown break fraction λ, which therefore needs to be estimated jointly with d and μ. The limiting distribution is a function of both standard Brownian Motion (BM) and a variant of fractional BM (fBM). Define,

$$\mathcal{A}_d^0\left(\lambda, a\right) = \frac{1}{\lambda\left(1-\lambda\right)}\left(B\left(\lambda\right) - \lambda B\left(1\right) + a\left[\lambda\left(1-\lambda_0\right) - \left(\lambda-\lambda_0\right)_+\right]\right)^2$$

and

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

where $L\left(d; a, b\right) = \left(1-2d\right)\int_{\max(a,b)}^1 \left(s-a\right)^{-d}\left(s-b\right)^{-d} ds$ (so that $L\left(d; a, a\right) \equiv \left(1-a\right)^{1-2d}$ and $L\left(0; a, b\right) = 1 - \max\left(a, b\right)$) and where $\tilde{\mathcal{W}}_{d_0}\left(a, b\right) = \left(1-2d_0\right)^{1/2}\int_a^b \left(s-a\right)^{-d_0} dB'\left(s\right)$ is a variant of fBM, $B\left(s\right)$ and $B'\left(s\right)$ are two independent BM. Notice that $\tilde{\mathcal{W}}_{d_0}\left(a, b\right)$ differs from the standard

[10] However, in practice, the restriction that the break in d is smaller in absolute value that 0.5 does not seem to be too relevant, as illustrated in our Monte Carlo study where it is shown that power increases when breaks in d exceed 0.5.

fBM, $\mathcal{W}_{d_0}(\lambda) = (1-2d_0)^{1/2} \int_0^\lambda (\lambda - s)^{-d_0} dB'(s)$, by its particular covariance structure, given by $Cov\left(\tilde{\mathcal{W}}_{d_0}(s,1), \tilde{\mathcal{W}}_{d_0}(t,1)\right) = L(d_0; s, t)$.

Then, with $\omega^2 = \frac{\pi^2}{6}$, the following result holds.

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

$$\sup_{\lambda\in\Lambda} \widetilde{LM}_{2,T}(\lambda) \xrightarrow{d} \sup_{\lambda\in\Lambda}\left\{\mathcal{A}_d^0(\lambda, \delta\omega) + \mathcal{A}_\mu^0\left(d_0, \lambda, \frac{\eta/\sigma_0}{\sqrt{1-2d_0}\Gamma(1-d_0)}\right)\right\}.$$

Remark 7. The distribution of the sup-$\widetilde{LM}_{2,T}(\lambda)$ test under the null is then given by

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

which only depends on d_0, but not on λ_0 due to the lack of identification of the break fraction under the null of no breaks in any of the two parameters. Critical values of such limiting distribution will be provided below in Section 7, for a grid of values of d_0 and ϵ.

Remark 8. Under local alternatives, the two components A_d^0 and A_μ^0 in the asymptotic distribution of the sup- $\widetilde{LM}_{2,T}(\lambda)$ test capture the contributions of the local shifts of the memory parameter and of the level, respectively. It is noteworthy that, while $\mathcal{A}_d^0(\lambda, \delta\omega)$ is symmetric around the break fraction $\lambda_0 = 0.5$, $\mathcal{A}_\mu^0\left(d_0, \lambda, \eta/(\sigma_0\sqrt{1-2d_0}\Gamma(1-d_0))\right)$ turns out to be positively (negatively) skewed if $d_0 > 0$ ($d_0 < 0$). Hence, if there is only a break in d, 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 and μ, and if $d_0 > 0$ ($d_0 < 0$), then local power would be highest for some $\lambda_0 < 0.5$ ($\lambda_0 > 0.5$).[11]

Theorem 1 also nests the two special cases where one tests exclusively for a break in d (so that A_μ^0 drops) or only in μ (so that A_d^0 drops) under H_0, reflecting that these two tests are asymptotically independent. Notice that if it is assumed that only one of the parameters breaks, a testing procedure which does not allow for a break in the other parameter could lead to better power properties in finite samples. In contrast, estimation of the model under this null could lead to misleading conclusions if the other parameter is the one that actually changes while the tested one is constant (see Section 3.2 for an alternative method which is robust to this problem).

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

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

$$\widetilde{LM}_{2,T}(\lambda_0) \xrightarrow{d} \chi_2^2(c),$$

[11] In Section 3.3, we will show that allowing for an autoregressive lag polynomial in the model reduces the local drift of the memory break component, but it does not affect the component corresponding to testing for a break in μ.

with non-centrality parameter

$$c = \delta^2 \lambda_0 (1 - \lambda_0) \omega + \frac{\eta^2}{\sigma_0^2} \frac{L(d_0; \lambda_0, \lambda_0) - L^2(d_0; 0, \lambda_0)}{(1 - 2d_0) \Gamma^2 (1 - d_0)} \equiv c_d + c_\mu.$$

As expected, when λ_0 is known, the asymptotic distribution becomes a $\chi_2^2(c)$ with a non-centrality parameter c which depends on the two drifts under local alternatives, namely c_d and c_μ. Moreover, as in the case of unknown λ_0, Corollary 1 nests the cases of testing for a break in only one of the parameters: (i) if we test for a break only in d, the limiting distribution becomes $\chi_1^2(c_d)$, where c_μ drops even if $\eta \neq 0$, and (ii) if we test for a break only in μ, the limiting distribution becomes $\chi_1^2(c_\mu)$, where c_d drops even when $\delta \neq 0$.

Remark 9. Martins and Rodriguez (2014) and Hassler and Meller (2014) have recently proposed similar LM test statistics to ours 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.[12] 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 and 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, despite acknowledging that the asymptotic distribution would be altered due to the lower rate of convergence of this estimator, they do not derive it.

3.2 Consistency of the LM test and break source identification

In this section we prove the consistency of the LM tests for either breaks in only one of the two parameters or breaks in both. In particular, as regards the $\widetilde{LM}_{2,T}$ tests for the null $H_0 : \theta_0 = \nu_0 = 0$ (i.e., no breaks in d or μ), we consider the following three alternative fixed hypotheses,

$$H_1^d(\lambda_0) : \theta \neq 0 \text{ and } \nu = 0,$$
$$H_1^\mu(\lambda_0) : \theta = 0 \text{ and } \nu \neq 0,$$
$$H_1^{d,\mu}(\lambda_0) : \theta \neq 0 \text{ and } \nu \neq 0,$$

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

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

three aforementioned fixed alternatives. To achieve break-source identification, it is convenient to derive individual LM tests under the following two simple null hypotheses,

$$H_0^d(\lambda_0) : \theta = 0,$$
$$H_0^\mu(\lambda_0) : \nu = 0,$$

which 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 using the joint $\widetilde{LM}_{2,T}$ tests and, in case of rejection, to identify the specific source of the break. In effect, following a rejection of the null at the first stage, then a test of the null $H_0^d(\lambda_0) : \theta = 0$ (resp. $H_0^\mu(\lambda_0) : \nu = 0$) could be applied at a second stage to confirm if d (resp. μ) is actually breaking, irrespectively of whether the other parameter shifts or not. As will be discussed below (see Remark 10), testing for $H_0^d(\lambda_0)$ and $H_0^\mu(\lambda_0)$ by means of the individual versions of the $\widetilde{LM_{2,T}}$ tests, with the non-tested parameter always taken to be stable under the null, could lead to incorrect identification of the source of the break. Thus, to robustify these 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 $H_0^d(\lambda_0) : \theta = 0$, 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 (7), the following $H_0^d(\tilde{\lambda}_T)$-restricted estimates should be considered

$$(\bar{d}_{0T}, \bar{\mu}_{0T}, \bar{\nu}_{0T}) = \arg \min_{d \in D, \mu, \mu+\nu \in M} \sum_{t=1}^{T} \left(\Delta_t^d \left(y_t - \mu - \nu R_t^{(2)} \left(\tilde{\lambda}_T \right) \right) \right)^2, \tag{11}$$

where different levels are allowed in each of the two regimes, and λ_0 is replaced by the estimate of the break date $\tilde{\lambda}_T$ obtained from the first step. Below is shown that this procedure yields a consistent estimator of λ_0 under any of the three alternatives. Then, the robust individual version of the LM test for $H_0^d(\lambda_0)$, denoted as $\overline{LM}_{2,T}^d(\tilde{\lambda}_T)$, is given by

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

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 -restricted residuals $\bar{\varepsilon}_t = \varepsilon_t\left(\bar{\psi}_T\right) = \Delta_t^{\bar{d}_{0T}} \left(y_t - \bar{\mu}_{0T} - \bar{\nu}_{0T} R_t^{(2)} \left(\tilde{\lambda}_T \right) \right)$. In a similar way, we can define $\overline{LM}_{2,T}^\mu(\tilde{\lambda}_T)$ to test H_0^μ based on the corresponding $H_0^\mu\left(\tilde{\lambda}_T\right)$-restricted estimation, where this time $\bar{\psi}_T = (\bar{\theta}_{0T}, 0, \bar{d}_{0T}, \bar{\mu}_{0T}, \bar{\sigma}_T^2)'$.

Under the null, both $\widetilde{LM}$ and $\overline{LM}$ tests have the same asymptotic distribution whereas, under the alternatives, the following result holds.

Proposition 1 *Under Assumptions 1 and 2a or 2b, then:*

(a) The estimator $\tilde{\lambda}_T$ defined in (9) estimates the break fraction λ_0 super-consistently at rate T under $H_1^d(\lambda_0)$, $H_1^{\mu}(\lambda_0)$ and $H_1^{d,\mu}(\lambda_0)$.

(b) The LM test statistics for a break in both parameters, $\widetilde{LM}_{2,T}^{d,\mu}(\lambda_0)$ and $\sup_\lambda \widetilde{LM}_{2,T}^{d,\mu}(\lambda)$, diverge: (i) at rate T under either $H_1^{d,\mu}(\lambda_0)$ or $H_1^d(\lambda_0)$, and (ii) at rate T^{1-2d_0} (resp. T) under Assumption 2b (resp. Assumption 2a) and $H_1^{\mu}(\lambda_0)$. The same results hold for $\widetilde{LM}_{1,T}^{d,\mu}(\lambda_0)$ and $\sup_\lambda \widetilde{LM}_{1,T}^{d,\mu}(\lambda)$.

(c) The robust $\overline{LM}$ test statistic for a break only in the memory, $\overline{LM}_{2,T}^{d}(\tilde{\lambda}_T)$, 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_1^2 distribution under $H_1^{\mu}(\lambda_0)$, namely, when only μ breaks. The same results hold for $\overline{LM}_{1,T}^{d}(\tilde{\lambda}_T)$.

(d) The robust $\overline{LM}$ test statistic for a break in the level, $\overline{LM}_{2,T}^{\mu}(\tilde{\lambda}_T)$, diverges: (i) at rate T^{1-2d_0} (resp. T) under Assumption 2b (resp. Assumption 2a) and $H_1^{\mu}(\lambda_0)$; (ii) at rate T^{1-2d_1} (resp. T) under Assumption 2b (resp. Assumption 2a) and $H_1^{d,\mu}(\lambda_0)$. By contrast, it converges to a χ_1^2 distribution under $H_1^d(\lambda_0)$, namely, when only d breaks. The same results hold for $\overline{LM}_{1,T}^{d}(\tilde{\lambda}_T)$, except that the divergence rate in (ii) is T^{1-2d_0} under Assumption 2b and $H_1^{d,\mu}(\lambda_0)$.

Proposition 1 illustrates why our identification strategy of the sources of potential breaks works well. On the one hand, parts (a) and (b) of Proposition 1 imply that using the joint $\widetilde{LM}$ tests allows one to detect consistently any breakpoint (at a known or unknown date). On the other hand, upon rejection of parameter constancy, the robust individual tests $\overline{LM}_{2,T}^{d}(\tilde{\lambda}_T)$ and $\overline{LM}_{2,T}^{\mu}(\tilde{\lambda}_T)$ 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 the rates of divergence of the test statistics $\widetilde{LM}_{2,T}^{d,\mu}$ and $\overline{LM}_{2,T}^{\mu}(\tilde{\lambda}_T)$ 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. Furthermore, as expected, they decrease with the level of persistence.

Remark 10. As will be explained in more detail in Section 6, it is also worth noticing that the drift terms under $H_1^{d,\mu}(\lambda_0)$ and positive d differ between $\widetilde{LM}_{1,T}^{d}(\tilde{\lambda}_T)$ and $\widetilde{LM}_{2,T}^{d}(\tilde{\lambda}_T)$, being higher (resp. lower) when the test is implemented in the *first* (resp. *second*) regime and $d_0 < d_1$ (resp. $d_0 > d_1$), that is, when d shifts upwards (resp. downwards). The insight for this result comes from one of the steps in the proof of Proposition 1 (see online Appendix) where, under a break in d, it is shown that the probability limit of the restricted estimate of $\widetilde{d}_{0T}$ obtained from (7) is given by

$$d_A = \arg\min_{d \in D_i} \left\{ \lambda_0 \frac{\Gamma(1-2(d_0-d))}{\Gamma^2(d-d_0+1)} + (1-\lambda_0) \frac{\Gamma(1-2(d_1-d))}{\Gamma^2(d-d_1+1)} \right\}, \tag{12}$$

where D_i is either D_1 or D_2, and d_A represents a weighted average of memory in both regimes. From (12), it can be checked that, when d increases (resp. decreases) from the first to the second regime, d_A becomes closer to d_1 (resp. d_0). Thus, since the $\widetilde{LM}_{2,T}$ test for the null of d stable exploits non-zero correlations between $\log \Delta_t \widetilde{\varepsilon}_t$ and $\widetilde{\varepsilon}_t$ in the *second* regime, the smaller the distance between d_1 and d_A implies that the power of the $\widetilde{LM}_{2,T}$ test will be lower than the power of the $\widetilde{LM}_{1,T}$ test. Conversely, when d decreases, the converse result holds since, in this case, d_A becomes closer to d_0.

REMARK 11. In contrast with the results in parts (c) and (d) of Proposition 1, it is important to remark that the use of individual $\widetilde{LM}$ tests for breaks in a single parameter may lead to spurious rejections when the non-tested parameter happens to be the only one breaking. For instance, as in Rachinger (2016), it can be shown that, under $H_1^{\mu}(\lambda_0)$ (when only μ breaks), a non-robust individual $\widetilde{LM}$ test for a break in d diverges.[13] Since a similar individual $\widetilde{LM}$ test for a break in μ under $H_1^{\mu}(\lambda_0)$ will asymptotically reject the null, the conclusion drawn from the second stage of the sequential procedure (based on joint and individual $\widetilde{LM}$ tests) would be that both parameters break, when in fact only μ does so.

3.3 Allowing for stable short-run dynamics in LM tests

We are now in disposition to discuss how the previous results change when $\alpha(L) \neq 1$ but the autoregressive parameters are assumed to be stable. The restricted estimates of the parameters used to compute $\tilde{\varepsilon}_t$ and $\tilde{\sigma}_T^2$ now result from the minimization of the following conditional sum of squares estimator (CSS),

$$\left(\tilde{d}_{0T}, \tilde{\mu}_{0T}, \tilde{\boldsymbol{\alpha}}_{0T}\right) = \arg \min_{d \in D, \mu \in M, \alpha \in A} \sum_{t=1}^{T} [\alpha(L) \Delta_t^d (y_t - \mu)]^2, \tag{13}$$

where $\boldsymbol{\alpha}_0 \in Int(A)$ and the compact set A excludes roots of $\alpha(L)$ on or inside the unit circle. Let $\omega_\alpha^2 = \frac{\pi^2}{6} - \kappa' \Phi^{-1} \kappa$, such that $\kappa = (\kappa_1, \ldots, \kappa_p)'$ 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 $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$ under Gaussianity.

Corollary 2 *The conclusions of Theorem 1 and Corollary 1 hold for the general model with autoregressive component replacing ω^2 by ω_α^2.*

Since $\omega^2 > \omega_\alpha^2$, the test now has lower local power due to the estimation effect of the autoregressive parameters. However, in the same way as the test for a break in μ was not affected by the estimation of d_0, this test with $\alpha_0(L) \neq 1$ is neither affected by the presence of short- memory

[13] It does so at a fairly slow rate $T^{1-4d_0} \log^2 T$ when $0 < d_0 \leq 0.25$ and at the faster rate T when $d_0 \leq 0$.

parameters.

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 $d \neq d_0$.[14] Moreover, they argue that, while remaining asymptotically equivalent under local alternatives, the LMW-type EFDF test has considerably higher power than the LM test under fixed alternatives. Thus, relying upon this approach, a similar test for joint breaks in d and μ is proposed here, focusing on the null hypothesis (H0) in model (2).

For simplicity, we start by considering the case where the parameters in the first regime, d_0 and μ_0 are assumed to be known, so that the data in the *second regime* satisfies for $t = [T\lambda_0]+1, \ldots, T$,

$$\begin{aligned} \Delta_t^{d_0}(y_t - \mu_0) &= (\Delta_t^{d_0} - \Delta_t^{d_1})(y_t - \mu_0) + \Delta_t^{d_1} R_t^{(2)}(\lambda_0)(\mu_1 - \mu_0) + \varepsilon_t \\ &= (1 - \Delta_t^{\theta})\Delta_t^{d_0}(y_t - \mu_0) + \nu \Delta_t^{d_1} R_t^{(2)}(\lambda_0) + \varepsilon_t, \end{aligned}$$

where recall $d_1 = d_0 + \theta$ and $\mu_1 = \mu_0 + \nu$ 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 $\theta = \nu = 0$ can be constructed by means of a joint test of

$$H_0 : \vartheta_1 = \vartheta_2 = 0$$

in the following regression model,[15]

$$\Delta_t^{d_0}(y_t - \mu_0) = \vartheta_1 \left[\frac{1 - \Delta_t^{\theta R_t^{(2)}(\lambda)}}{\theta} \right] \Delta_t^{d_0}(y_t - \mu_0) + \vartheta_2 \Delta_t^{d_1} R_t^{(2)}(\lambda) + \varepsilon_t, \tag{14}$$

for $t = 1, \ldots, T,$ and each λ and θ. Defining $\Theta = (\vartheta_1, \vartheta_2)'$, $Y_t^0 = \Delta_t^{d_0}(y_t - \mu_0)$ and

$$X_t(\lambda) = X_t(\lambda, \theta, d, \mu) = \left(\left[\frac{1 - \Delta_t^{\theta R_t^{(2)}(\lambda)}}{\theta} \right] \Delta_t^{d}(y_t - \mu), \quad \Delta_t^{d+\theta} R_t^{(2)}(\lambda) \right)',$$

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

[15] 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.

the regression (14) can be rewritten in a more compact way as

$$Y_t^0 = \Theta' X_t^0(\lambda) + \varepsilon_t, \tag{15}$$

with $X_t^0(\lambda) = X_t(\lambda, \theta, d_0, \mu_0)$.

Under the more realistic assumption of unknown d_0 and μ_0, running regression (15) requires the estimation of those two parameters, on top of θ. For d_0 and μ_0, one can use the restricted estimates $\tilde{d}_{0T}$ and $\tilde{\mu}_{0T}$ obtained under the null using the whole sample. As for θ, one can set $\hat{\theta}_T(\lambda) = \hat{d}_{1T}(\lambda) - \tilde{d}_{0T}$, where $\hat{d}_{1T}(\lambda)$ is the CSS estimate obtained from the second subsample, defined by a given λ. The main justification for the estimation of (d_0, μ_0) based on the minimization of (7) with observations for the whole sample, rather than just for the first regime, is that it facilitates comparison of the LM and LMW-type tests, since both use the same parameter estimates under the null.[16]

This leads to the following feasible regression model

$$\tilde{Y}_t = \Theta' \tilde{X}_t(\lambda) + e_t \tag{16}$$

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

Testing for breaks in both parameters corresponds to the joint null hypothesis of $\vartheta_1 = \vartheta_2 = 0$ in (16), while testing for a break only in d (resp. only in μ) corresponds to the null hypothesis of $\vartheta_1 = 0$ (resp. $\vartheta_2 = 0$).[17] Then, the LMW-type test statistic (implemented in the *second regime*) from regression (16) for the joint hypothesis $H_0 : \Theta = 0$ is defined as

$$\widetilde{LMW}_{2,T}(\lambda) = \tilde{\Theta}_T(\lambda)' \tilde{V}_T^{-1}(\lambda) \tilde{\Theta}_T(\lambda), \tag{17}$$

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

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

where

$$\hat{\sigma}_T^2(\lambda) = \frac{1}{T} \sum_{t=1}^{[T\lambda]} \left(\Delta_t^{\hat{d}_0}(y_t - \hat{\mu}_0)\right)^2 + \frac{1}{T} \sum_{t=[T\lambda]+1}^{T} \left(\Delta_{t-[T\lambda]}^{\hat{d}_1}(y_t - \hat{\mu}_1)\right)^2 \tag{18}$$

with $\left(\hat{d}_0, \hat{\mu}_0\right)$ and $\left(\hat{d}_1, \hat{\mu}_1\right)$ being the CSS estimators for the first and second regime respectively, and $\tilde{\Lambda}_2^{1/2}$ defined in (8). In the construction of the LMW-type test, we use the variance estimate under the alternative to improve its power properties.

[16] 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.

[17] As before, note that if it assumed that only one of the two 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 (14) when testing for a break in the memory, that is $H_0 : \vartheta_1 = 0$).

From the discussion in Wooldridge (2002) and LV (2007), it follows that, when (d_0, μ_0) are taken as known, the estimation of θ by $\hat{\theta}_T(\lambda)$ does not affect the null asymptotic distribution of the Wald-type test derived from (15). However, this is no longer true when (d_0, μ_0) need to be estimated since these parameters affect the left-hand-side variable in regression (16). As with the LM test, notice that estimation of the parameters (d_0, μ_0) increases the variance of the LMW-type test statistic. This is reflected in the need to pre- and post-multiply by $\tilde{\Lambda}_2^{1/2}$ in the definition of $\tilde{V}_T(\lambda)$ compared to the usual LS expression. In the case when λ is unknown, we set

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

and our test statistic becomes $\widetilde{LMW}_{2,T}(\tilde{\lambda}_T) = \sup_{\lambda\in\Lambda} \widetilde{LMW}_{2,T}(\lambda)$.

Remark 12. For the more general model with $\alpha(L) \neq 1$, in line with LV (2007) we propose a two-step procedure where the data is filtered from the short-run dynamics prior to running the testing regression. We start again with the DGP under the alternative, given by $\alpha_0(L)\Delta_t^{d_0}(y_t - \mu_0) = \varepsilon_t$, $t = 1, \ldots, [T\lambda_0]$ and

$$\alpha_0(L)\Delta_t^{d_0}(y_t - \mu_0) = (1 - \Delta_t^{\theta})\alpha_0(L)\Delta_t^{d_0}(y_t - \mu_0) + \nu\alpha_0(L)\Delta_t^{d_1}R_t^{(2)}(\lambda) + \varepsilon_t$$

for $t = [T\lambda_0] + 1, \ldots, T$, which leads to the following regression model estimated by OLS

$$\tilde{\alpha}_{0T}(L)\Delta_t^{\tilde{d}_{0T}}(y_t - \tilde{\mu}_{0T}) = \vartheta_1\left[\frac{1 - \Delta_t^{\hat{\theta}(\lambda)R_t^{(2)}(\lambda)}}{\hat{\theta}(\lambda)}\right]\tilde{\alpha}_{0T}(L)\Delta_t^{\tilde{d}_{0T}}(y_t - \tilde{\mu}_{0T}) + \vartheta_2\tilde{\alpha}_{0T}(L)\Delta_t^{\tilde{d}_{0T}+\hat{\theta}_T(\lambda)}R_t^{(2)}(\lambda) + \tilde{\varepsilon}_t,$$

where, as before, the parameter estimates under the null, $\left(\tilde{d}_{0T}, \tilde{\mu}_{0T}, \tilde{\boldsymbol{\alpha}}_{0T}\right)$, are obtained over both regimes by (13). As a result, testing the joint null of a break in d and μ corresponds to $H_0 : \vartheta_1 = \vartheta_2 = 0$.

Remark 13. 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. For example, in the simpler case where $\alpha_0(L) = 1$, the previous regression model becomes

$$\begin{aligned}\Delta_t^{\tilde{d}_{0T}}(y_t - \tilde{\mu}_{0T}) &= \vartheta_1\left[\frac{1 - \Delta_t^{\hat{\theta}_T R_t^{(1)}(\lambda)}}{\hat{\theta}_T}\right]\left[\Delta_t^{\tilde{d}_{0T}}(y_t - \tilde{\mu}_{0T})\right] \\ &\quad + \vartheta_2\left[\Delta_t^{\hat{d}_{0T}}R_t^{(1)}(\lambda) + \left(\Delta_t^{\tilde{d}_{0T}} - \Delta_{t-[\lambda T]}^{\tilde{d}_{0T}}\right)\left(1 - R_t^{(1)}(\lambda)\right)\right] + \tilde{\varepsilon}_t,\end{aligned}$$

We next show that, using estimates $(\tilde{d}_{0T}, \tilde{\mu}_{0T})$ in place of (d_0, μ_0), 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 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 $\widetilde{LMW}_{1,T}(\lambda_0)$.

(c) Corollary 2 applies when autoregressive components are fitted.

4.1 Consistency of the LMW-type test and break source identification

We next discuss the consistency of the LMW-type test for breaks in d and/or μ under fixed alternatives and, in case of rejection, of the corresponding individual tests for break source identification.

To derive individual LMW-type tests, we can proceed as with the $\overline{LM}$ tests derived in Section 3.2. For example, when testing $H_0^d(\lambda_0)$, we use $H_0^d(\tilde{\lambda}_T)$-restricted estimates $(\bar{d}_{0T}, \bar{\mu}_{0T}, \bar{\nu}_{0T})$ as in (11) to define $\bar{Y}_t = \Delta_t^{\bar{d}_{0T}}(y_t - \bar{\mu}_{0T})$ and $\bar{X}_t = X_t(\tilde{\lambda}_T, \bar{\theta}_T, \bar{d}_{0T}, \bar{\mu}_{0T})$, where $\bar{\theta}_T = \bar{\theta}_T(\tilde{\lambda}_T) = \hat{d}_{1T}(\tilde{\lambda}_T) - \bar{d}_{0T}$ compares memory estimates for the two regimes allowing also for different levels. As a result, when regressing $\bar{Y}_t$ on $\bar{X}_t$, we will only test for the significance of the first element of Θ by means of the Wald statistic $\overline{LMW}_{2,T}^d\left(\tilde{\lambda}_T\right) = \tilde{\vartheta}_{1T}^2\left(\tilde{\lambda}_T\right)/\tilde{V}_{1,1,T}\left(\tilde{\lambda}_T\right)$ which is asymptotically distributed as a χ_1^2 under H_0^d when $\tilde{\lambda}_T \to_p \lambda_0$.

Under $H_0^\mu(\lambda_0)$, an equivalent LMW-type test can be derived using $H_0^\mu\left(\tilde{\lambda}_T\right)$-restricted estimates to define $\bar{Y}_t$ and $\bar{X}_t$ and $\overline{LMW}_{2,T}^\mu\left(\tilde{\lambda}_T\right) = \tilde{\vartheta}_{2T}^2\left(\tilde{\lambda}_T\right)/\tilde{V}_{2,2,T}\left(\tilde{\lambda}_T\right)$. If the tests were to be implemented in the first regime, then similar definitions would apply to $\overline{LMW}_{1,T}^d$ and $\overline{LMW}_{1,T}^\mu$.

Then, the following result holds.

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

(b) The LMW-type test for a break in the memory, $\overline{LMW}_{2,T}^d\left(\tilde{\lambda}_T\right)$, behaves as the $\overline{LM}_{2,T}^d\left(\tilde{\lambda}_T\right)$ test in Proposition 1 c), and idem for $\overline{LMW}_{1,T}^d(\lambda_0)$.

(c) The LMW-type test for a break in the level, $\overline{LMW}_{2,T}^\mu\left(\tilde{\lambda}_T\right)$, behaves as the $\overline{LM}_{2,T}^\mu\left(\tilde{\lambda}_T\right)$ test in Proposition 1 d), and idem for $\overline{LMW}_{1,T}^\mu(\lambda_0)$.

Notice that the remarks below Proposition 1 apply here as well. The main difference between the LM test and the LMW-type test is that while the former uses the filter $-\log\Delta$, the latter uses

$(1-\Delta_t^{\theta})/\theta$, which can be very different when θ does not converge to zero (under a fixed alternative). This leads to different drift terms although it does not affect the rates of divergence of the tests. See Section 6 for more details.

5 Models with changing autoregressive parameters

In this section, we consider the richer class of models where $\alpha(L) \neq 1$ and changes in this autoregressive structure of the DGP (as well as in d and μ) are subject to testing. In particular, we consider the following DGP

$$\begin{aligned} \alpha_0(L)\,\Delta_t^{d_0}(y_t-\mu_0) &= \varepsilon_t,\ \ t=1,\ldots,[\lambda_0 T] \\ (\alpha_0(L)+\beta(L))\,\Delta_t^{d_0+\theta}\left(y_t-\mu_0-\nu R_t^{(2)}(\lambda_0)\right) &= \varepsilon_t,\ \ t=[\lambda_0 T]+1,\ldots,T, \end{aligned} \tag{19}$$

where $\beta(L)$ is a lag polynomial of order p, with $\beta(0)=0$, and $\alpha_1(L)=\alpha_0(L)+\beta(L)$ is a stable autoregressive polynomial with coefficients also changing at $[\lambda_0 T]$.[18] Defining $\boldsymbol{\beta}=\left(\beta_1,\ldots,\beta_p\right)'$, this leads to the following likelihood,

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

with $\psi^*=\left(\theta,\nu,\boldsymbol{\beta}',d,\mu,\boldsymbol{\alpha}',\sigma^2\right)'$ and

$$\varepsilon_t\left(\psi^*,\lambda\right)=\left(\alpha(L)+\beta(L)\,R_t^{(2)}(\lambda)\right)\Delta_t^{d+\theta R_t^{(2)}(\lambda)}\left(y_t-\mu-\nu R_t^{(2)}(\lambda)\right).$$

(A) LM Test: To define the corresponding LM test statistic $\widetilde{LM}_{2,T}(\lambda)$ for $H_0:\left(\theta,\boldsymbol{\beta}',\nu\right)'=0$, the derivative of $\mathcal{L}_T$ in the direction of $\boldsymbol{\beta}$, evaluated at the restricted estimates $\tilde{\psi}_T^*=(0,0,0,\tilde{d}_{0T},\tilde{\mu}_{0T},\tilde{\boldsymbol{\alpha}}_{0T}',\tilde{\sigma}_T^2)'$ and $\tilde{\varepsilon}_t=\varepsilon_t(\tilde{\psi}_T^*)=\tilde{\alpha}_{0T}(L)\Delta_t^{\tilde{d}_{0T}}\left(y_t-\tilde{\mu}_{0T}\right)$, is given by

$$\left.\frac{\partial\mathcal{L}_T(\psi,\lambda)}{\partial\boldsymbol{\beta}}\right|_{\psi^*=\tilde{\psi}_T^*}=\frac{1}{\tilde{\sigma}_T^2}\sum_{t=[\lambda T]+1}^{T}\begin{pmatrix}\tilde{\alpha}_T^{-1}(L)\,\tilde{\varepsilon}_{t-1}\\ \cdots\\ \tilde{\alpha}_T^{-1}(L)\,\tilde{\varepsilon}_{t-p}\end{pmatrix}\tilde{\varepsilon}_t.$$

(B) LMW-type Test: Next, we can now modify the LMW-type test regression in order to allow for breaks in the short-run dynamics. In particular, the DGP (19) for $t=[\lambda_0 T]+1,\ldots,T$ now becomes,

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

[18] Although, for simplicity, we have assumed the same lag -order for $\alpha(L)$ and $\beta(L)$, this model also would allow the order of the polynomial $\beta(L)$, p_1, to be different from p.

from which we can run the following regression,

$$\tilde{\alpha}_{0T}(L)\,\Delta_t^{\tilde{d}_{0T}}(y_t-\tilde{\mu}_{0T}) = \vartheta_1\tilde{\alpha}_{0T}(L)\left[\frac{1-\Delta_t^{\hat{\theta}_T R_t^{(2)}(\lambda)}}{\hat{\theta}_T}\right]\Delta_t^{\tilde{d}_{0T}}(y_t-\tilde{\mu}_{0T})$$
$$+\vartheta_2\tilde{\alpha}_{0T}(L)\,\Delta_t^{\tilde{d}_{0T}+\hat{\theta}_T R_t^{(2)}(\lambda)}R_t^{(2)}(\lambda)$$
$$+\sum_{j=1}^{p}\vartheta_{2+j}R_t^{(2)}(\lambda)\,\Delta_{t-j}^{\tilde{d}_{0T}+\hat{\theta}_T R_t^{(2)}(\lambda)}\left(y_{t-j}-\tilde{\mu}_{0T}-\hat{\nu}_T R_t^{(2)}(\lambda)\right)+\tilde{\varepsilon}_t,$$

where $\hat{\theta}_T$ is as before and $\hat{\nu}_T$ is obtained as the difference between $\widetilde{\mu}_{0T}$ and $\hat{\mu}_{1T}(\lambda)$. Then, we can define the LMW-type test statistic $\widetilde{LMW}_T(\lambda)$ as in (17) but replacing λ by λI_{p+1} in the upper left element in $\tilde{\Lambda}_2$, for the null

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

which considers a joint break in the $2+p$ parameters, while it is also possible to develop tests on subsets of the parameters to account for specific breaks.

The following proposition discusses the asymptotic behaviour of the LM and LMW-type tests for joint breaks in memory, level and short run dynamics under the local alternative

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

where $\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_p)'$. Let

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

with B_{p+1} being a $(p+1)$-dimensional standarised Brownian motion, while κ and Φ are defined in Section 3.3,

$$\mathcal{A}_{d,p}^0(\lambda,\delta,\boldsymbol{\gamma})=\frac{1}{\lambda(1-\lambda)}\varpi_p(\lambda,\delta,\boldsymbol{\gamma})'\,\Xi^{-1}\varpi_p(\lambda,\delta,\boldsymbol{\gamma})\ \text{ and }\ \omega_p^2(\delta,\boldsymbol{\gamma})=\begin{pmatrix}\delta\\ \boldsymbol{\gamma}\end{pmatrix}'\Xi\begin{pmatrix}\delta\\ \boldsymbol{\gamma}\end{pmatrix}.$$

Theorem 3 *(a) With an unknown break fraction λ_0, under Assumptions 1 and 2 for the DGP (19), under $H_{1,T}^{d,\alpha,\mu}(\lambda_0)$, then $\sup_\lambda \widetilde{LM}_{2,T}(\lambda)$ and $\sup_\lambda \widetilde{LMW}_{2,T}(\lambda)$ converge to*

$$\sup_{\lambda\in\Lambda}\left\{\mathcal{A}_{d,p}^0(\lambda,\delta,\boldsymbol{\gamma})+\mathcal{A}_\mu^0\left(d_0,\lambda,\frac{\eta/\sigma_0}{\sqrt{1-2d_0}\Gamma(1-d_0)}\right)\right\}.$$

(b) With a known break fraction λ_0, under Assumptions 1 and 2a or 2b for the DGP (19), and under $H_{1,T}^{d,\alpha,\mu}(\lambda_0)$, $\widetilde{LM}_{2,T}(\lambda_0)$ and $\widetilde{LMW}_{2,T}(\lambda_0)$ converge to a $\chi^2_{2+p}(c)$ distribution, where

$$c=\omega_p^2(\delta,\boldsymbol{\gamma})\,\lambda_0(1-\lambda_0)+\frac{\eta^2}{\sigma_0^2}\frac{L(d_0;\lambda_0,\lambda_0)-L^2(d_0;0,\lambda_0)}{(1-2d_0)\Gamma^2(1-d_0)}.$$

This theorem shows that the contributions of the tests for breaks in short-run dynamics parameters and level parameter are independent in the joint tests. However, interactions among the former parameters arise.

Notice that a similar analysis to the one developed in Section 3.2 to identify the source of break would be possible in the presence of short-run parameters. In effect, different robustness strategies could be developed generalizing H_1^d and H_1^μ by either setting $\gamma = 0$ and introducing $H_1^\alpha(\lambda_0) : \theta = \nu = 0$ and $\gamma \neq 0$, or by focusing instead on distinguishing H_1^μ from $H_1^{d,\alpha}(\lambda_0) : \nu = 0$, $(\theta, \gamma') \neq 0$. We omit details for the sake of space.

6 Asymptotic power comparisons

Our next step is to analyze the asymptotic behaviour of the LM and LMW-type tests under $H_1^d(\lambda_0)$, $H_1^\mu(\lambda_0)$ and $H_1^{d,\mu}(\lambda_0)$. In addition, given the discussion in Remark 11 about potential differences in power depending on whether tests are implemented in the *first* or *second* regime, we also distinguish between these two cases. For illustrative purposes, DGP (2)-(H1) with $\alpha_0(L) = 1$ and unknown break date is the setup considered here.

However, to do so, a preliminary observation is to notice that the relative power performance of the two tests may depend on the probability limits of the H_0-restricted estimates $\tilde{d}_{0T}$ and $\tilde{\mu}_{0T}$ under the previous set of fixed alternatives. On the one hand, recall that the probability limit of $\tilde{d}_{0T}$ when d breaks is given by d_A in (11). On the other hand, the corresponding probability limit of $\tilde{\mu}_{0T}$ when μ breaks is given by a weighted average of μ_0 and μ_1 whose value depends on which parameter changes and whether d is positive or negative,[19]

$$\mu_A = \mu_0 + L(d^*; 0, \lambda_0)(\mu_1 - \mu_0). \tag{20}$$

The value of d^* is selected as follows: (i) $d^* = d_0$ under Assumption 2b and H_1^μ, (ii) $d^* = d_A$ under Assumption 2b and $H_1^{d,\mu}$, and (iii) $d^* = 0$ under Assumption 2a and either $H_1^{d,\mu}(\lambda_0)$ or $H_1^\mu(\lambda_0)$.

6.1 Analytical derivation of drift terms

Given d_A and μ_A, we next derive the probability limits (denoted in short as *drift terms* in the sequel) of the rescaled LM and LMW-type test statistics in their Sup-type formats under the three different alternatives.

[19] As with d_A, the derivation of μ_A can be found in the proof of Proposition 1 (see online Appendix).

Breaks in memory: First, under $H_1^d(\lambda_0)$, let us first define the following terms,

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

$$C_{LMW}\left(d_i,d_A\right) = \frac{\Gamma\left(1+2\left(d_A-d_i\right)\right)}{\Gamma^2\left(1+\left(d_A-d_i\right)\right)}-1,$$

and

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

Then, the limiting behaviour of the two test statistics under $H_1^d(\lambda_0)$ can be derived as follows.

Proposition 3 *Under $H_1^d(\lambda_0)$,*

(a) the LM test (4) satisfies

$$\underset{T\to\infty}{p\lim}\frac{\sup_\lambda \widetilde{LM}_{1,T}\left(\lambda\right)}{T} = \frac{\lambda_0}{1-\lambda_0}C_{LM}\left(d_0,d_A\right)$$

$$\underset{T\to\infty}{p\lim}\frac{\sup_\lambda \widetilde{LM}_{2,T}\left(\lambda\right)}{T} = \frac{1-\lambda_0}{\lambda_0}C_{LM}\left(d_1,d_A\right);$$

(b) the LMW-type test (17) satisfies

$$\underset{T\to\infty}{p\lim}\frac{\sup_\lambda \widetilde{LMW}_{1,T}\left(\lambda\right)}{T} = \frac{\lambda_0}{1-\lambda_0}C_{LMW}\left(d_0,d_A\right)$$

$$\underset{T\to\infty}{p\lim}\frac{\sup_\lambda \widetilde{LMW}_{2,T}\left(\lambda\right)}{T} = \frac{1-\lambda_0}{\lambda_0}C_{LMW}\left(d_1,d_A\right).$$

From the definition of d_A in (12), it is easy to check that the smaller (larger) is λ_0, the closer d_A is to d_1 (d_0), and the smaller are $C_{LM}\left(d_1,d_A\right)$ and $C_{LMW}\left(d_1,d_A\right)$ (resp. $C_{LM}\left(d_0,d_A\right)$ and $C_{LMW}\left(d_0,d_A\right)$). As illustrated in Section 6.2 below, the drift terms satisfy $C_{LMW}\left(d_i,d_A\right) > C_{LM}\left(d_i,d_A\right)$. Therefore, under $H_1^d(\lambda_0)$, the LMW-type tests tend to dominate the LM tests in terms of asymptotic power.

Breaks in level: Secondly, under $H_1^\mu(\lambda_0)$, let us define the following term,

$$C^\mu = \left(\frac{\mu_1-\mu_0}{\sigma_0}\right)^2\frac{1}{\left(1-2d_0\right)\Gamma^2\left(1-d_0\right)}.$$

The following Proposition provides the drift terms under $H_1^\mu(\lambda_0)$. Two relevant cases are distinguished, dictated by the sign of d_0. First, when d_0 is positive (Assumption 2b) the following result holds, irrespective of the regime where the tests are implemented.

Proposition 4 *Under $H_1^{\mu}(\lambda_0)$ and Assumption 2b, the sup-type tests satisfy*

$$p\lim_{T\to\infty}\sup_{\lambda}\frac{\widetilde{LM}_{i,T}(\lambda)}{T^{1-2d_0}} = p\lim_{T\to\infty}\sup_{\lambda}\frac{\widetilde{LMW}_{i,T}(\lambda)}{T^{1-2d_0}} = \left(L(d_0;\lambda_0,\lambda_0) - L^2(d_0;0,\lambda_0)\right)C^{\mu},\ \ i=1,2.$$

Unlike the result under $H_1^d(\lambda_0)$, in this case the two tests exhibit the same asymptotic power for $d_0, d_1 > 0$ (Assumption 2b), the reason being that the asymptotic variance is equally efficiently estimated under the null and under this type of alternative.

However, when $d_0 < 0$ (Assumption 2a), the estimator $\widetilde{d}_{0T}$ is inconsistent and converges to 0 under $H_1^{\mu}(\lambda_0)$. Hence, as in Proposition 2, the rate of divergence is T rather than T^{1-2d_0} or T^{1-2d_1}. As a result, the drift terms of the two tests are no longer identical, and the limit of the estimator of the variance in the LMW-type-test, $\bar{\sigma}^2_{LMW} = \sigma_0^2$, is smaller than the corresponding limit of the LM test, $\bar{\sigma}^2_{\mu,LM}$, given by,

$$\bar{\sigma}^2_{\mu,LM} = \sigma_0^2\frac{\Gamma(1-2d_0)}{\Gamma^2(1-d_0)} + \lambda_0(1-\lambda_0)\left(\frac{\mu_1-\mu_0}{\sigma_0}\right)^2.$$

Hence, the LMW-type test has better asymptotic power than the LM test when $d < 0$ and the alternative involves a break in μ. The following proposition summarizes the discussion above.

Proposition 5 *Under $H_1^{\mu}(\lambda_0)$ and Assumption 2a,*

(a) the LM test (4) satisfies

$$\begin{aligned} p\lim_{T\to\infty}\sup_{\lambda}\frac{\widetilde{LM}_{1,T}}{T} &= \frac{\lambda_0}{1-\lambda_0}C_{LM}(d_0,0) + \left(\frac{\mu_1-\mu_0}{\bar{\sigma}_{\mu,LM}}\right)^2\lambda_0(1-\lambda_0), \\ p\lim_{T\to\infty}\sup_{\lambda}\frac{\widetilde{LM}_{2,T}}{T} &= \frac{1-\lambda_0}{\lambda_0}C_{LM}(d_0,0) + \left(\frac{\mu_1-\mu_0}{\bar{\sigma}_{\mu,LM}}\right)^2\lambda_0(1-\lambda_0); \end{aligned}$$

(b) the LMW-type test (17) satisfies

$$\begin{aligned} p\lim_{T\to\infty}\sup_{\lambda}\frac{\widetilde{LMW}_{1,T}}{T} &= \frac{\lambda_0}{1-\lambda_0}C_{LMW}(d_0,0) + \left(\frac{\mu_1-\mu_0}{\sigma_0}\right)^2\lambda_0(1-\lambda_0), \\ p\lim_{T\to\infty}\sup_{\lambda}\frac{\widetilde{LMW}_{2,T}}{T} &= \frac{1-\lambda_0}{\lambda_0}C_{LMW}(d_0,0) + \left(\frac{\mu_1-\mu_0}{\sigma_0}\right)^2\lambda_0(1-\lambda_0). \end{aligned}$$

Breaks in memory and level: Finally, under $H_1^{d,\mu}(\lambda_0)$, the break in d dominates for Assumption 2b, and $T^{-1}\sup\widetilde{LM}_{i,T}$ and $T^{-1}\sup\widetilde{LMW}_{i,T}$ $(i=1,2)$ behave as in Proposition 3. Alternatively, under Assumption 2a, the variance in the LM test becomes

$$\bar{\sigma}^2_{(d,\mu),LM} = \sigma_0^2\left(\lambda_0\frac{\Gamma(1-2d_0)}{\Gamma^2(1-d_0)} + (1-\lambda_0)\frac{\Gamma(1-2d_1)}{\Gamma^2(1-d_1)}\right) + \lambda_0(1-\lambda_0)\left(\frac{\mu_1-\mu_0}{\sigma_0}\right)^2,$$

and the test statistics diverge at rate T. The drift term is the sum of the respective terms in Proposition 3 with $d_A = 0$ and the respective second terms in Proposition 5 with $\bar{\sigma}_{\mu,LM}$ replaced by $\bar{\sigma}_{(d,\mu),LM}$.

6.2 Graphical representation of drift terms

To illustrate some of the results on asymptotic power gathered in Propositions 3 to 5, we next proceed to plot a few graphs of the drift terms in the two tests. We start by considering $H_1^{\mu}(\lambda_0)$ as the alternative hypothesis. For illustrative purposes we only display results for LM and LMW-type tests implemented in the second regime.

Since the drift terms depends on the memory (d_0), the break fraction (λ_0) and the size of the break in the level $(\frac{\mu_1-\mu_0}{\sigma_0})$, we proceed by varying one parameter at a time, while keeping fixed the remaining two parameters. Thus, while the varying parameter is represented in the horizontal axis, the next three graphs share the same vertical axis. This axis displays the drift multiplied by the factor T^{1-2d_0} which allows us to compare drifts for different memory parameters.

Figure 1.a displays the break magnitudes $(\mu_1-\mu_0)/\sigma_0$ in the horizontal axis, for $T=200$, $\lambda_0 = 0.5$, $d_0 \in \{-0.25, 0, 0.25\}$, $\mu_0 = 0$ and $\sigma_0 = 1$. In view of Proposition 5, the drift is increasing and symmetric in the break magnitude, reaching a minimum at $\mu_1 - \mu_0 = 0$. Furthermore, whereas the drifts of the LM and LMW-type tests in the second regime (solid and dashed lines) coincide under Assumption 2b $(0 < d_0 < 0.5)$, that is not the case under Assumption 2a $(-0.5 < d_0 < 0)$, where the drift of the LMW-type (crosses) test is uniformly larger than the drift of the LM test (squares). For $d_0 < 0$, notice that there is a discontinuity at 0 arising from the fact that, in this case, $\tilde{d}_{0T}$ converges to 0, rather than to d_0, thus providing additional power. Under the null of $\mu_0 = \mu_1$, $\tilde{d}_{0T}$ converges to d_0 and thus this additional drift vanishes.

Figure 1.b in turn displays the break fractions λ_0 in the horizontal axis for $T = 200$, $d_0 \in \{-0.25, 0, 0.25\}$ and $(\mu_1 - \mu_0)/\sigma_0 = 1$. Notice that the drift of the test in the second regime is no longer symmetric around $\lambda_0 = 0.5$ since, as argued earlier, the function $\left(L\left(d_0; \lambda, \lambda\right) - L^2\left(d_0; 0, \lambda\right)\right)$ does not satisfy this property. Again, for $-0.5 < d_0 < 0$, the drift of the LM test is smaller than the drift of the LMW-type test. In this case, the drift decreases in the break fraction as a result of the factor $\frac{1-\lambda_0}{\lambda_0}$ and the limit of the memory estimate $d_A = 0$. Further, given that the estimated variance in the LM test exceeds the true one, depending on $(\frac{\mu_1-\mu_0}{\sigma_0})$, the drift of the LM test is not necessarily decreasing in d_0. The drift of the tests in the first regime, are the same as in Figure 1.b under Assumption 2b and increasing in λ_0 under Assumption 2a.

Finally, Figure 1.c displays different values of d in the horizontal axis for $T = 200$, $\lambda_0 = 0.5$ and $(\mu_1 - \mu_0)/\sigma_0 = 1$. In this case the drift is decreasing in the memory. Interestingly, as before, there is a discontinuity (at $d_0 = 0$) arising from the inefficient estimation of the variance of the LM test for $d_0 < 0$.

[Figures 1.a, 1.b and 1.c about here]

We next analyze the drift terms when $H_1^{d}(\lambda_0)$ becomes the alternative. Figure 2.a and 2.b displays the drift term of the test for a break in d under $H_1^d(\lambda_0)$ when testing in the second and first regime, respectively. In both cases, the memory changes from $d_0 = 0$ to $d_1 \in (-0.5, 0.5)$ at $\lambda_0 = 0.5$. The estimator (12) is closer to d_0 (d_1) if $d_1 < d_0$ ($> d_0$) leading to higher (lower) power when the memory decreases (increases) in the second regime.

[Figures 2a and 2.b about here]

Overall, the evidence above indicates that, under fixed alternatives, the LMW-type test potentially exhibits better power properties than the LM tests when the alternative involves a break in d either exclusively or jointly with a break in μ, and that power may depend on whether d_0 goes up or down.

6.3 Symmetric tests

To mitigate the above-mentioned dependence of power on the direction in which the long-memory parameter d breaks, one possible suggestion is to take a simple average of the tests implemented in the first and second regime. In what follows, these tests will be denoted as $\widetilde{LM}_{1+2,T}$ and $\widetilde{LMW}_{1+2,T}$, respectively.[20] 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

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

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

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 local alternative, the asymptotic properties of the symmetric tests mimic those of the corresponding individual tests derived in Theorems 1 and 2, so that they can be implemented using the same critical values. The reason is that, since both

[20] Although for brevity we focus on the $\widetilde{LM}$ and $\widetilde{LMW}$-type versions of the tests, notice that similar definitions can be applied to their robust versions, namely, the $\overline{LM}$ and $\overline{LMW}$-type tests, whose finite sample properties will be discussed in Section 7 below.

individual 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 distribution when it is known). Thus, their average converges to the same individual limit under the null, while the drift equals the average drift of the individual tests under the alternative.

7 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.[21] First, consider the case of a known break fraction of $\lambda_0 = 0.5$. The significance level is 0.05 and the sample sizes are $T = 200, 500$ and $1,000$ when considering size, and $T = 200$ as regards power. We assume an error variance $\sigma_0^2 = 1$ 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 1 (panels a and b) displays the size of the symmetric LM and LMW-type tests for breaks in both d and μ, respectively. The main finding is that the $\widetilde{LM}$ test (though slightly undersized), and the $\widetilde{LMW}$-type tests (slightly oversized) control size fairly well and approach 5% as the sample size grows. Table 1 (panels c and d) displays the power results of the two symmetric tests for a break in d and/or μ at $\lambda_0 = 0.5$. Figures in bold characters correspond to size. Our simulation results confirm that there are some power gains from using the LMW-type tests in finite samples. As can be inspected, power is increasing in the magnitude of the shifts in d and μ, respectively.[22] For example, looking at the middle block of panel d, for $\mu_0 = 0$, a shift in d from 0 to 0.2 increases the power of the LMW-type test in the first regime by 23.8 pp. ($= 32.0 - 8.2$) whereas, for $d_0 = 0$, a shift in μ from 0 to 0.25 raises power by 28.5 pp. ($= 36.7 - 8.2$). The corresponding rise in power when d shifts to 0.4 (for $\mu_0 = 0$) and when μ shifts to 0.5 (for $d_0 = 0$) are 79.3 pp. and 79.0 pp., respectively. Finally, as expected, the power arising from breaks in μ is lower the higher d. For

[21] Results on simulated size and power of the individual LM and LMW-type tests implemented in the first and second regimes, 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 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.

[22] As mentioned earlier, notice that the dip in power in some of the entries is due to those cases where the alternative coincides with the null. For example, the cells for $d_1 = -0.2$ and $\mu_1 = 0$ in the first block of panels d and e, correspond to the DGP under the null, so that the rejection rates equal the sizes of the respective tests.

instance, using the latter shifts in μ, when $d_0 = 0.2$ rather than $d_0 = 0$, only increases the power of the LMW-type test by 8.2 pp. ($= 15.5 - 7.3$) and 33.4 pp. respectively.

[Table 1 about here]

Table 2 reports simulation results regarding size and power of symmetric $\overline{LM}$ and $\overline{LMW}$-type tests for the case of a break in μ which is robust to breaks in d (panels a and b), as well as for the case of a break in d which is robust to breaks in μ (panels c and d). As before, the break fraction $\lambda_0 = 0.5$ is assumed to be known. The significance level is again 0.05 and the sample size is $T = 200$. 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 $\mu_1 \in \{0, 0.25.0.5, 1\}$. The number of simulations is $10,000$. Several results stand out. First, in panels a and b, the $\overline{LMW}$-type test has both higher size and higher power than the $\overline{LM}$ test. Further, for all values of d_0, the tests control the size well, even though they are often slightly undersized (bold numbers); thus, both tests are robust to breaks in the non-tested parameter, d. Power increases with the break magnitude in μ (moving in the rows to the right). As expected, power decreases with d_0 (moving in the blocks from left to right) and d_1 (moving down the rows). When comparing these findings with the corresponding ones in Table 1, both size and power in the robust tests are slightly smaller than the corresponding ones in the joint tests. Second, in panels c and d, both size and power of the $\overline{LM}$ and $\overline{LMW}$-type test are very similar. For all values of d_0, the tests control the size well, even though they are slightly oversized (bold numbers); hence they are robust to breaks in the non-tested parameter, μ. Power increases with break magnitude in μ (moving up and down the rows), while it remains relatively constant over different level break magnitudes (comparing the columns). When comparing these findings with the corresponding ones in Table 1, again both size and power in the robust tests are comparable to the corresponding ones in the joint tests.

[Table 2 about here]

Next, we report simulation results regarding size and power of the symmetric $\widetilde{LM}$ and $\widetilde{LMW}$-type test 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. As discussed in Remark 7, the critical values of both tests for several values of d_0, which are shown in Table 3, have been generated from Theorem 1 using 2,000 grid points for the break fraction, and 20,000 simulations. To find the critical values for an unknown d_0, we interpolate between these values and replace d_0 by $\tilde{d}_{0T}$ as in (7) (see Giraitis et al., 2006, for a similar solution).

[Table 3 about here]

Table 4 shows that both tests have satisfactory size properties. In terms of power, again the LMW-type test often performs better, especially for larger values of d_0 and d_1.

[Table 4 about here]

Finally, we conduct a small Monte Carlo for the size and power properties of the symmetric $\widetilde{LM}$ and $\widetilde{LMW}$-type test with known λ_0 and with short run dynamics which are potentially changing. Note that when incorporating short run dynamics, the $\widetilde{LMW}$-type test with values of d in the interval $(-0.5, 0.5)$ both under the null and under alternative, leads to uncontrolled size for some parameter combinations. Therefore, we restrict the estimates $\hat{d}_{1T}$ and $\hat{d}_{2T}$ to lie in the interval $[\tilde{d}_{0T} - 0.25, \tilde{d}_{0T} + 0.25]$. Naturally, by doing so, we lose power in some circumstances. 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 5 (panel a) shows that for the LM test the size is relatively well controlled. Table 6 (panel a) illustrates that the $\widetilde{LMW}$-type test is still oversized for $T = 200$ but, as the sample size gets larger, it gets closer to the nominal size. Again panels b in both tables illustrate that there can be some gains in power from using the $\widetilde{LMW}$-type test, especially for positive autoregressive components.

[Table 5 and 6 about here]

8 Discussion

The starting point of this paper is to stress that the joint modeling of breaks in the memory 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 twofold. On the one hand, we extend the well-known LM test for breaks only in the memory parameter to also account for breaks in the level, as well as propose a novel regression-based LMW-type test for $FI(d)$ processes that also accounts for these shifts. Under the alternative, a sequential procedure combining tests for joint and individual (robust) breaks can consistently identify the source(s) of shifts. 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. Furthermore, both tests are easily amendable to also account for breaks in short-run dynamics. Our analytical results and Monte-Carlo simulations show in particular that LMW-type tests for joint breaks can yield power gains relative to LM tests in some instances.

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 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.

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

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

with

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

where $R_t^{(i+1)} (\boldsymbol{\lambda}) = R_t^{(i+1)} (\lambda_i, \lambda_{i+1}) = 1 ([\lambda_i T] < t \leq [\lambda_{i+1} T]), \lambda_0 = 0, \lambda_m = 1, \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{m-1})'$. 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

$$\mathcal{L}_T \left(\theta_1, \theta_2, d_0, \nu_1, \nu_2, \mu_0, \sigma^2, \lambda_1, \lambda_2\right) = -\frac{T}{2} \log \left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \varepsilon_t (\psi, \boldsymbol{\lambda})^2,$$

with

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

where $R_t^{(2)} (\boldsymbol{\lambda}) = 1 ([\lambda_1 T] < t \leq [\lambda_2 T])$, $R_t^{(3)} (\boldsymbol{\lambda}) = 1 (t > [\lambda_2 T])$ and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$. The LM test is constructed as in (4) 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,

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

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 $H_0' : \vartheta_1 = \vartheta_3 = 0$ ($\vartheta_2 = \vartheta_4 = 0$) corresponds to testing for two breaks only in d (μ).

References

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

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

[3] Beran, J. and Terrin, N., 1999 "Correction note to "Testing for a change of the long-memory parameter" (Beran and Terrin, 1996)". *Biometrika*, Vol. 86, No. 1, p. 233.

[4] Breitung, J. and Hassler, U., 2002. "Inference on the cointegration rank in fractionally integrated processes," *Journal of Econometrics*, Elsevier, vol. 110(2), pages 167-185, October.

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

[6] 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.

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

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

[9] 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.

[10] 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.

[11] 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.

[12] 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

[13] 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.

[14] Gil-Alana, L. A., 2008. Fractional integration and structural breaks at unknown periods of time. *Journal of Time Series Analysis* 29, 163-185.

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

[16] 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.

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

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

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

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

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

[22] 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.

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

[24] 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.

[25] 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.

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

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

[28] 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.

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

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

[31] 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.

[32] Rachinger, H., 2016. "Multiple breaks in long memory time series". University of Vienna, mimeo. https://sites.google.com/site/heikorachinger/Rachinger2016.pdf

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

[34] Robinson, P.M., 1994. ”Efficient tests of nonstationary hypotheses”. *Journal of the American Statistical Association*, 89, 1420-1437

[35] 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.

[36] 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.

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

[38] Tanaka, K., 1999. "The nonstationary fractional unit root". *Econometric Theory* 15, 549-582.

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

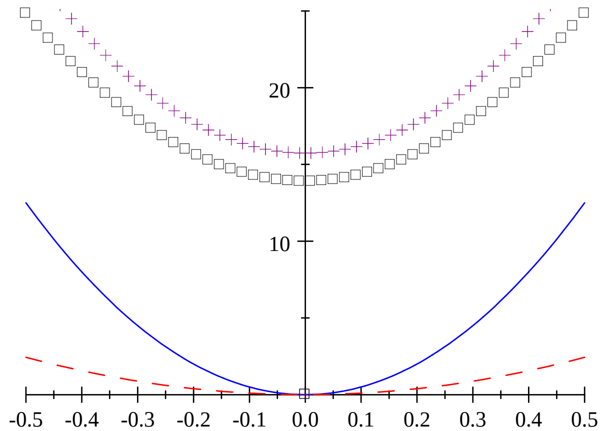
[40] Zaffaroni P, 2004, Contemporaneous aggregation of linear dynamic models in large economies, *Journal of Econometrics*, 120, 75-102

9 FIGURES AND TABLES

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

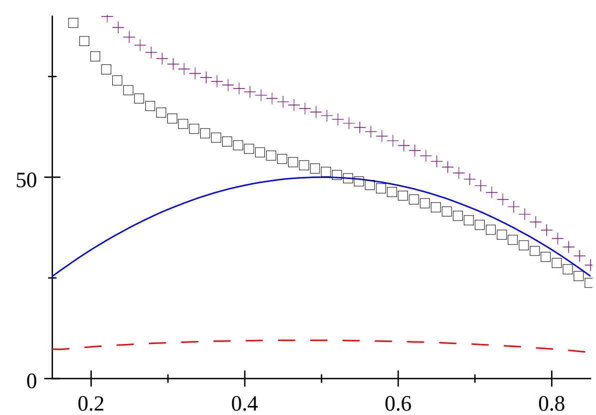
a) Drift of the tests as a function of the break magnitude $\frac{\mu_1-\mu_0}{\sigma_0}$ (λ_0= 0.5).

d_0= −0.25 ($\widetilde{LM}_2$: squares; $\widetilde{LMW}_2$-type: crosses), 0 (solid), 0.25 (dashed).



b) Drift of the tests in the second regime as a function of the break fraction ($T = 200$, $\left|\frac{\mu_1-\mu_0}{\sigma_0}\right| = 1$).

d_0= −0.25 ($\widetilde{LM}_2$ on second regime: squares; $\widetilde{LMW}_2$-type: crosses), 0 (solid), 0.25 (dashed)



c) Drift of the tests as a function of the memory parameter (λ_0= 0.5, $\left|\frac{\mu_1-\mu_0}{\sigma_0}\right| = 1$ and $T = 200$).

$\widetilde{LM}_2$ test (solid and dashed); $\widetilde{LMW}_2$-type test (solid and squares).

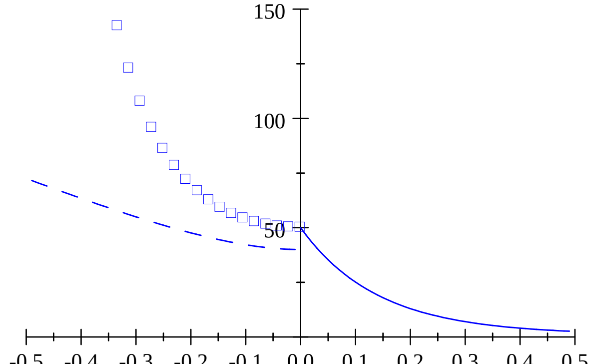
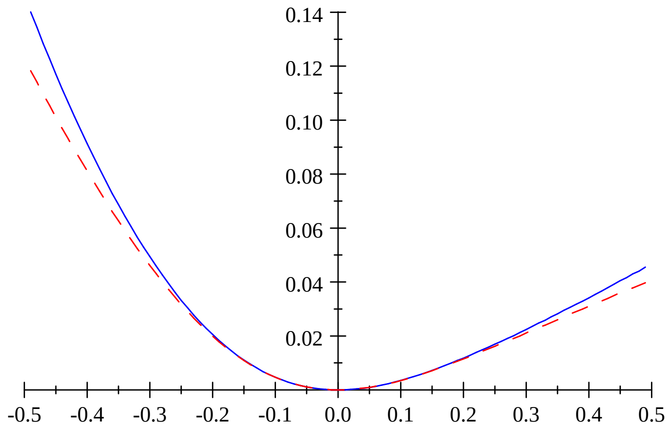


Figure 2: **Drift of the tests for a break in the memory**

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) ($\lambda_0 = 0.5$, $d_0 = 0$).

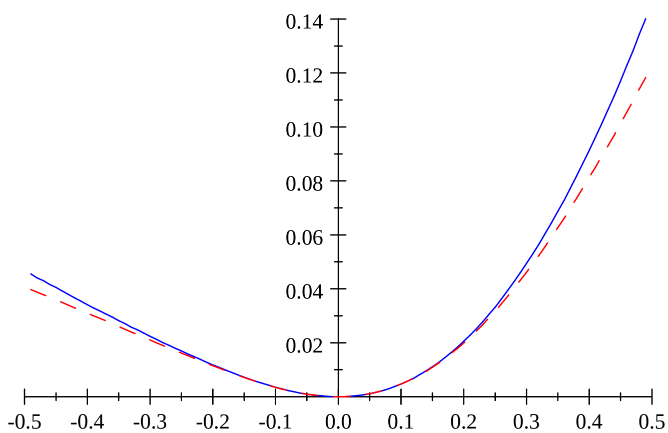


Table 1: **Simulated size and power of symmetric LM and LMW-type tests for a joint break in memory and level.**

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	4.0	4.0	4.0	4.8	6.3
500	5.1	4.8	4.7	4.2	4.5	4.3	4.2	5.1	5.8
1000	4.9	5.1	4.9	4.6	4.6	4.5	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	6.8	7.5	7.7	8.2	7.8	7.3	7.6	8.4
500	6.2	6.1	6.3	5.9	6.5	6.1	6.2	6.6	6.8
1000	5.7	5.9	5.9	5.6	5.9	5.7	6.3	5.7	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	**4.0**	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	**4.0**	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	**7.5**	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	**7.2**	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	**7.3**	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 2: **Simulated size and power of symmetric robust LM and LMW-type tests for breaks in memory and level.**

a) Symmetric $\overline{LM}$ test for a break in level robust to a break in memory

d_0	-0.2				0				0.2			
$d_1 \setminus \mu_1$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
-0.4	**4.8**	92.9	100	100	**3.3**	58.4	95.5	99.9	**1.9**	32.5	61.5	86.6
-0.2	**3.2**	71.9	97.4	99.9	**4.3**	33.3	81.1	99.3	**4.3**	13.9	39.0	85.7
0	**2.4**	33.5	82.1	98.1	**2.5**	20.8	66.0	95.2	**2.6**	8.6	26.3	72.6
0.2	**3.1**	10.0	32.1	80.1	**2.3**	10.0	32.1	80.1	**2.5**	7.2	24.6	69.5
0.4	**4.6**	6.7	13.5	43.1	**4.4**	6.5	13.5	43.3	**3.8**	6.1	13.4	42.7

b) Symmetric $\overline{LMW}$-type test for a break in level robust to a break in memory

d_0	-0.2				0				0.2			
$d_1 \setminus \mu_1$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
-0.4	**8.2**	97.5	100	100	**6.3**	72.8	98.9	100	**2.2**	39.7	76.7	94.5
-0.2	**5.0**	82.9	99.7	100	**7.6**	46.4	91.8	99.9	**8.6**	24.0	55.3	95.5
0	**3.1**	36.8	85.3	99.0	**3.2**	25.4	74.1	98.8	**4.1**	11.9	35.1	84.0
0.2	**3.2**	10.3	33.0	82.7	**2.7**	10.5	33.4	83.3	**2.7**	7.7	26.3	75.5
0.4	**4.8**	7.1	14.1	44.4	**4.5**	6.8	14.0	44.8	**4.0**	6.5	13.8	44.2

c) Symmetric $\overline{LM}$ test for break in memory robust to a break in level

d_0	-0.2				0				0.2			
$d_1 \setminus \mu_1$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
-0.4	30.2	29.9	30.9	30.6	79.4	79.4	78.4	78.8	93.0	92.7	93.1	92.9
-0.2	**6.0**	**6.0**	**6.3**	**6.0**	35.4	35.2	34.8	35.2	85.3	85.3	85.4	86.3
0	37.0	36.8	37.4	37.2	**6.6**	**6.3**	**6.1**	**6.1**	36.6	36.7	37.5	37.4
0.2	87.3	87.6	88.0	88.1	36.9	36.3	36.4	37.6	**6.3**	**6.1**	**6.2**	**6.3**
0.4	99.4	99.5	99.5	99.5	87.9	87.9	87.5	88.1	37.4	36.6	37.1	37.0

d) Symmetric $\overline{LMW}$-type test for break in memory robust to a break in level

d_0	-0.2				0				0.2			
$d_1 \setminus \mu_1$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
-0.4	30.0	29.8	30.6	30.5	79.5	79.6	78.7	78.8	93.4	93.1	93.3	93.2
-0.2	**6.0**	**6.1**	**6.4**	**5.8**	35.4	35.2	34.8	35.2	85.6	85.6	85.6	86.6
0	37.3	37.4	37.9	37.9	**6.6**	**6.3**	**6.2**	**6.3**	36.8	36.8	37.8	37.4
0.2	87.7	87.9	88.3	88.3	37.4	36.7	36.7	37.9	**6.4**	**6.1**	**6.4**	**6.5**
0.4	99.4	99.6	99.5	99.6	88.2	88.3	87.9	88.6	37.8	36.8	37.5	37.3

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

Table 3: **Critical Values of LM-tests for unknown λ for breaks in μ and d.**

$\epsilon \ \backslash \ d_0$	-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	11.0	11.0	11.4	11.7	12.2	13.1	14.5	16.1	17.5
0.2	10.5	10.6	10.6	10.7	11.0	11.2	11.9	12.9	14.2	15.6	17.2
0.25	9.6	10.0	10.1	10.2	10.5	10.5	11.2	11.9	13.1	14.4	15.6

Table 4: **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

$T \ \backslash \ d_0$	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4
200	9.0	7.1	6.5	6.4	4.8	3.9	3.4	3.0	7.0
500	6.2	5.6	5.6	5.1	4.9	4.7	4.4	4.8	5.2

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	8.3	7.4	7.1	7.3	6.9	7.0	6.7	5.8	8.7
500	6.5	6.0	6.5	5.9	6.3	6.1	6.9	6.5	6.1

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			
$\mu_1 \ \backslash \ d_1$	-0.4	-0.2	0	0.2	-0.4	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4
0	37.2	**6.5**	14.7	58.4	85.6	33.1	**4.8**	13.9	64.8	78.4	24.6	**3.4**	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			
$\mu_1 \ \backslash \ d_1$	-0.4	-0.2	0	0.2	-0.4	-0.2	0	0.2	0.4	-0.2	0	0.2	0.4
0	35.5	**7.1**	15.7	58.9	86.0	37.9	**6.9**	17.4	68.8	83.2	32.5	**6.7**	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 5: **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	8.8	4.2	7.3	7.9	3.7	7.1	7.5	3.9
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	$\alpha_1 \setminus d_1$	-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	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	**8.8**	37.6	89.4	85.7	31.5	**4.2**	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	**7.9**	32.1	86.5	79.0	29.9	**3.7**	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	**3.9**	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 6: **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	8.6	8.0	6.8	8.8	6.0	7.3	9.2	7.8
500	6.2	6.0	7.8	6.1	7.3	7.2	6.4	8.0	7.4
1000	5.9	5.8	8.0	5.6	6.3	8.3	6.0	6.8	7.9

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	**8.6**	24.2	78.9	86.9	33.7	**8.0**	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	**8.8**	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	**7.3**	29.2	81.3	85.1	29.9	**9.2**	30.8	83.1	83.7	29.9	**7.8**	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.