Multiple Breaks in Long Memory Time Series[1]

Job Market Paper

Heiko Rachinger[2]

Department of Economics
Universidad Carlos III de Madrid

November 15, 2011

Abstract. We analyze least squares (LS) estimation of breaks in long memory time series. We show that the estimator of the break fraction is consistent and converges at rate T when there is a break in the mean, in the memory or in both parameters. Further, we analyze tests for the number of breaks. When testing for breaks in the memory, the asymptotic results correspond to the standard ones in the literature. When testing for breaks in the mean and when testing for breaks in both parameters, the results differ in terms of the asymptotic distribution of the test statistic. In this case, the LS-procedure loses some of its nice properties, such as asymptotic pivotality. In a simulation exercise, we find that the tests based on asymptotic critical values are oversized in finite samples. Therefore, we suggest using the bootstrap, for which we derive validity and consistency, and we confirm its better size properties. Finally, we use the method to test for breaks in the U.S. inflation rate.

JEL Classification: C13, C22
Keywords: Structural Breaks, Fractional Integration, Least Squares Estimation, Testing, Bootstrap

1. INTRODUCTION

Macroeconomic and financial time series are in general persistent and display long memory characteristics such as hyperbolically decaying autocorrelation functions. There has been a long discussion whether these time series can be described as fractionally integrated models or whether their long memory is spurious due to breaks in their mean (Granger and Hyung, 2004). Recently, Perron and Qu (2010) discuss that many time series are more likely generated by stationary processes with a break

[1] I thank Carlos Velasco for his valuable advice. I further thank Vanessa Berenguer-Rico, Miguel Delgado, Juan José Dolado, Jesús Gonzalo, Uwe Hassler, Javier Hidalgo, Peter M. Robinson, Abderrahim Taamouti and participants of seminars at London School of Economics and Universidad Carlos III, ERCIM 2010 and ESEM 2011 for helpful comments and the London School of Economics for the hospitality during my stay in Spring 2010. I acknowledge financial support from the Spanish Ministerio de Educación y Ciencia, Ref. no. SEJ2007-62908/ECON.

[2] Department of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe, Spain. E-mail: hraching@eco.uc3m.es

in their mean rather than by long memory models. However, processes with breaks in the long memory parameter can also generate those series (McCloskey, 2010).

The aim of this paper is to provide a method to detect the presence of breaks in memory and in mean and to distinguish between them. We propose a unified approach for modeling breaks in the mean and the memory. In particular, we extend Bai and Perron (1998) methodology to the long memory context and analyze least squares estimation of breaks in long memory time series. In their short memory framework, they discuss a linear model with multiple breaks. They derive consistency and $T-$rate convergence of the break fraction estimate and the asymptotic distribution of the parameter estimates in the regimes. Finally, they provide a series of tests for the existence and number of breaks. Boldea and Hall (2010) extend Bai and Perron's (1998) analysis into a nonlinear setting. They show that the results of Bai and Perron (1998) do not change, even though the proofs become more involved. By considering nonlinear models, they encompass several ergodic models but not long memory time series models.

Hsu and Kuan (1998) and Lavielle and Moulines (2000) analyze the LS procedure for a process with a break in the mean and a stationary long memory error term, yet without breaks in the memory. Since they do not integrate explicitly the memory parameter in their analysis, they find different asymptotics. Further, Gil-Alana (2008) analyzes a similar methodology as ours. Nevertheless, he works with a data generating process that is not a typical long memory process. He also does not derive rigorously the asymptotic distributions of the estimates and statistics. He conjectures that the asymptotic properties resemble the ones found in Bai and Perron (1998). However, we show that the critical values employed in Gil-Alana (2008) are not the correct ones for testing for breaks in the mean. Besides, Gil-Alana (2008) is not specific about the impact coming from the estimation of the memory parameter d. Taking the latter into account, the problem becomes a nonlinear one and we have to consider specific arguments to derive the asymptotic properties.

In this paper, we derive consistency and T-rate convergence of the break fraction estimator and the asymptotic distribution of the parameter estimates when there are breaks in the memory and/or the mean. We assess the power of break tests by considering local breaks in the memory and in the mean. The asymptotic distribution of these tests differ from the ones of Bai and Perron (1998) and the procedure loses some of its nice properties, such as asymptotic pivotality. We discuss tests for determining which parameter is the changing one. Since the tests based on asymptotic critical values suffer from some size distortions in finite samples, we suggest using the bootstrap for which we derive validity and consistency.

Another strand of literature focuses on testing for the presence and the number of breaks in the memory parameter in time series with long memory. Beran and Terrin (1996, 1999) use parametric Whittle estimators to test for a break in the memory. Hassler and Meller (2009) introduce an augmented Lagrange Multiplier test to test semiparametrically for breaks in the memory, allowing for breaks in the mean. Hassler and Scheithauer (2011) show that tests for the null hypothesis of I(0) series against alternatives of a change from I(0) to I(1), discussed by Kim, Belaire-Franch and Amador (2002) and Busetti and Taylor (2004), are also consistent for a change from I(0) to I(d), for $d > 0$. Sibbertsen and Kruse (2009) derive a CUSUM of squares-based test. Martins and Rodrigues (2010) use recursive forward and backward estimation of a LM test. McCloskey (2010) uses a modified ratio of weighted partial sums to test semiparametrically for breaks in the memory.

In Section 2, we discuss the model and the least squares estimation of an unstable process. In Section 3, we derive the asymptotic behavior of the estimators in the presence of breaks. In Section 4, we analyze tests for the number of breaks and examine the behavior of these tests in finite samples. In Section 5, we propose a sequential testing strategy to determine which parameter is changing. In Section 6, we analyze the bootstrap. In Section 7, we apply the methodology to the U.S. inflation series and test for breaks in memory and mean in this series. Finally in Section 8, we conclude. Some Lemmata and additional Propositions which are needed for the analysis are provided in Appendix A. The proofs are collected in Appendix B.

2. PRELIMINARIES

We consider the following model with m breaks in $(T_1^0, T_2^0, ..., T_m^0)$ ($m+1$ regimes),

$$y_t = \mu_j^0 + \Delta_t^{-d_j^0} u_t, \ t = T_{j-1}^0 + 1, ..., T_j^0, \ j = 1, ..., m+1. \tag{1}$$

The coefficients of interest $\theta_j^0 = (\mu_j^0, d_j^0)$ lie in some set $\Theta_j = M_j \times D_j$. The process consists of an intercept and a Type II fractionally integrated disturbance,

$$\Delta_t^{-d_j^0} u_t = \sum_{k=0}^{t-1} \pi_k(-d_j^0) u_{t-k}, \tag{2}$$

where Δ_t^{-d} denotes the truncated fractional differencing filter with memory d and where

$$\pi_k(-d) = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}, k = 0, ..., t-1,$$

denote the sequence of coefficients of the expansion of Δ_t^{-d}. In this and in the next section, we assume that the number of breaks, m, is known but the actual break points, $\left(T_1^0, T_2^0, ..., T_m^0\right)$, are unknown. The latter will be estimated together with the parameter vector $(\theta_j^0)_{j=1}^{m+1}$. We consider equally the cases of a pure structural change model, in which both coefficients change, and a partial structural change model, in which some coefficient does not change.

For obtaining the conditional sum of squares (CSS) estimator in a stable context, it suffices to apply the filter Δ_t^d to the process since for $d = d^0$, the resulting residuals are u_t. Nevertheless, for a unstable process, it is not correct to apply the filter $\Delta_t^{d_j^0}$ to the process (1), as it is done in Gil-Alana (2008), because $\Delta_t^{d_j^0} y_t$ is a weighted sum of $I(d_1)$ to $I(d_j)$ terms rather than u_t. In order to avoid this problem, Dolado *et al.* (2009) define the process implicitly as

$$\Delta_t^{d_j^0}\left(y_t - \mu_j^0\right) = u_t,\ t = T_{j-1}^0 + 1, ..., T_j^0. \tag{3}$$

In this case it suffices to apply the fractional differencing filter $\Delta_t^{d_j^0}$ to obtain $I(0)$ residuals and the whole analysis simplifies considerably. However, the process defined in (3) is not strictly a $I(d_j^0)$ process in $t > T_1^0$. Therefore, we rather apply a filter to (1) that restricts the filtered data to lie in the interval of the corresponding regime. First, we define a break fraction λ_i and the true break fraction λ_i^0 as T_i/T and T_i^0/T respectively. In particular, we set the residuals

$$\hat{u}_t\left(\lambda_{j-1}, \theta_j\right) = \Delta_{t-[\lambda_{j-1}T]}^{d_j}\left(y_t - \mu_j\right),\ t = T_{j-1} + 1, ..., T_j. \tag{4}$$

Since the fractional differencing filter for regime j is restricted to the observations of this regime, this filter avoids the aforementioned mixing of observations from different regimes. The resulting residuals in (4) are close to $I(0)$, if break fraction and coefficients are estimated close to the true ones. However, apart from terms coming from the distance between estimate and true break fraction and coefficients, there are also some additional terms coming from the fact that the applied fractional filter is too short. These terms are similar in nature to the terms that show up when applying a truncated Type II fractional filter to a untruncated Type I process. The technical difficulties arise from showing that all these terms are asymptotically negligible.

In particular, assume the process has m breaks at $\left(T_1^0, ..., T_m^0\right)$, where the true number of breaks m is known. We estimate the break fractions $\lambda_j = T_j/T$ together with the coefficients in the regimes by conditional sum of squares (CSS) estimation. Let

$$S_T\left(\boldsymbol{\lambda}, \boldsymbol{\theta}\right) = \sum_{i=1}^{m+1} S_{i,T}\left(\lambda_{i-1}, \lambda_i, \theta_i\right) = \sum_{i=1}^{m+1} \sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_i T]} \hat{u}_t\left(\lambda_{i-1}, \theta_i\right)^2, \tag{5}$$

where $\hat{u}_t$ is defined in (4). For simplicity, we illustrate the procedure for $m = 1$, the general case follows equally. For a given break fraction λ_1 with $T_1 = [\lambda_1 T]$ and (d_1, d_2),

$$\left\{\hat{\mu}_i\left(d_i, \lambda_1\right)\right\}_{i=1,2} = \underset{\mu_1, \mu_2 \in M_1 \times M_2}{\operatorname{argmin}} \left\{S_{1,T}\left(0, \lambda_1, \mu_1, d_1\right) + S_{2,T}\left(\lambda_1, 1, \mu_2, d_2\right)\right\}.$$

Substituting the estimator $\{\hat{\mu}_i(d_i, \lambda_1)\}_{i=1,2}$ into the objective function, we obtain the conditional memory estimator

$$\{\hat{d}_i(\lambda_1)\}_{i=1,2} = \underset{d_1,d_2 \in D_1 \times D_2}{\operatorname{argmin}} \{S_{1,T}(0, \lambda_1, \hat{\mu}_1(d_1), d_1) + S_{2,T}(\lambda_1, 1, \hat{\mu}_2(d_2), d_2)\}.$$

Finally, we minimize the objective function with respect to λ_1 and obtain an estimator for the break fraction as

$$\hat{\lambda}_1 = \arg\min_{\lambda_1} S_{1,T}\left(0, \lambda_1, \hat{\mu}_1\left(\hat{d}_1(\lambda_1), \lambda_1\right), \hat{d}_1(\lambda_1)\right) + S_{2,T}\left(\lambda_1, 1, \hat{\mu}_2\left(\hat{d}_2(\lambda_1), \lambda_1\right), \hat{d}_2(\lambda_1)\right)$$

The estimator for the parameters d_i and μ_i $(i = 1, 2)$ are

$$\hat{d}_i(\hat{\lambda}_1) \text{ and } \hat{\mu}_i(\hat{d}_i(\hat{\lambda}_1), \hat{\lambda}_1).$$

The truncated filter (4) is attractive because it estimates the parameters in the different regimes separately. Therefore, considering m breaks is conceptionally not more involved than considering *one* break. Besides, it extends easily to a Type I process DGP, $\Delta_{\infty}^{-d_i^0} u_t$. The only difference is that for a Type I process, the truncated part is $\sum_{j=0}^{t-1} \pi_j(d) \Delta_{\infty}^{-d_i^0} u_{t-j}$ rather than $\sum_{j=1}^{t} \pi_j(d) \Delta_{t-j}^{-d_i^0} u_{t-j}$.

For the subsequent analysis we need the following assumptions:

Assumption 1.

(i) The error term u_t is *iid* $\left(0, \sigma^2\right)$.

(ii) $E|u_t|^s < \infty$, $s > \frac{3}{2(1-2\max(d_i^0))}$.

Assumption 2. The common parameter space $\Theta = M \times D = ([\underline{\mu}, \bar{\mu}], [0, 1/2 - \varepsilon])$, $0 < \varepsilon < 1/2$, is compact and $\theta^0 \in \Theta$.

Assumption 3. $T_i^0 = \left[T\lambda_i^0\right]$, $i = 1, ..., m$, where $0 < \lambda_1^0 < ... < \lambda_m^0 < 1$.

Assumption 1 implies that the errors are independent from the regression function $f_t(\theta) = (\Delta_{t-T_{i-1}}^{d_i} - 1)(y_t - \mu_i)$, $E[u_t f_t(\theta)] = 0$ for all θ and t. In contrast to Boldea and Hall (2009), our regressor is not strictly stationary $\alpha - mixing$ but fractionally integrated. For further generalizations of the error term, we could assume a different variance in the different regimes or a short memory error process, $u_t = w_\theta(L)\varepsilon_t$. In the former, for $m = 1$, let $u_t^{(1)}$ and $u_t^{(2)}$ denote the errors of the two regimes. The variance of the mean estimator of the second regime depends then on both error variances. For the latter, the analysis is complicated by the correlation between the estimators of $w_\theta(L)$ and d. Hualde and Robinson (2010) analyze the case of this estimator in a stable context with short term component but without mean. In the following sections, we also consider the case of a stable autoregressive structure. Further, we discuss shortly the case of testing for a changing short term component and conjecture that the asymptotic distributions follow from combining Boldea and Hall's (2010) approach with ours. Assumption 1 Part (ii) is needed for weak convergence of partial sums of products of the regressor and the error term. Assumption 3 is a standard assumption in the break literature.

For the following analysis of the estimators in the presence of structural breaks, we need to analyze the behavior of the CSS estimator of one parameter if the other one is not consistently estimated. For simplicity, we consider the stable case. First, the CSS estimator of the memory works well when there is no deterministic component or when it is known or consistently estimated at rate $T^{1/2-d^0}$. On the other hand, if the mean is not consistently estimated, the memory estimator can have a huge bias in finite samples (Chung and Baillie, 1993). But there are no asymptotic results for this case to my best knowledge. Proposition 1a) delivers these results. Equally, we analyze the properties of the mean estimation when the memory is inconsistently estimated. Proposition 1b) shows that consistency and rate of convergence of the mean estimation are asymptotically not affected by the memory estimation.

PROPOSITION 1. *(Behavior of the CSS estimator)*
a) For the memory estimator given μ, for $d^0 \in Int(D)$,

$$\hat{d}\,(\mu) - d^0 = O_p(T^{-1/2}) \text{ uniformly in } \mu.$$

b) For the mean estimator given d,

$$\hat{\mu}\,(d) - \mu^0 = O_p(T^{d^0-1/2}) \text{ uniformly in } d \in D.$$

It turns out that the estimation is inconsistent for $d^0 = 0$ but still consistent for $d^0 \in Int(D)$. The finite sample effects depend on $d^0, (\mu^0 - \mu)$ and T. Especially, for d^0 close to 0, the estimate can be highly upward biased in finite samples. The same argument applies if we do not estimate μ, just set $\hat{\mu} = 0$.

In the following sections, we analyze long memory time series with a break only in the mean μ, only in the memory d or in both parameters.

3. ASYMPTOTIC BEHAVIOR OF ESTIMATES IN THE PRESENCE OF BREAKS

Given the nonlinear nature of our problem, our approach is closer to Boldea and Hall (2010) rather than to Bai and Perron (1998). However, our process is fractionally integrated and does not meet their conditions. In the following, we have to derive most of the results newly.

The break fraction estimate is consistent for breaks in the memory, in the mean and in both parameters.

THEOREM 1. *(Consistency of the break fraction estimator)*
Let $\hat{\lambda}_i$ be such that $\hat{T}_i = [T\hat{\lambda}_i]$. Then under Assumptions 1-3,

$$\hat{\lambda}_i \xrightarrow{p} \lambda_i^0.$$

Using consistency of the break fraction estimates, we establish their rate of convergence.

THEOREM 2. *(Rate of convergence of the break fraction estimator)*
For every $\eta > 0$, there exists a finite $C > 0$ such that for all large T,

$$P\left(T|\hat{\lambda}_i - \lambda_i^0| > C\right) < \eta.$$

We find $T-$rate convergence for the break fraction estimator when there are breaks in the memory, in the mean or in both parameters. This $T-$rate corresponds to the one found in Lavielle and Moulines (2000) for a break in the mean in a process with Type I long memory error but is faster than the one found in Hsu and Kuan (1998). Given the T rate convergence of the break fraction estimates, Theorem 3 provides consistency, the rate of convergence and the limiting distribution of the parameter estimates. The estimators d_i and d_j are independent and the estimators μ_i and μ_j are dependent.

THEOREM 3. *(Asymptotic distribution of the CSS estimators)*
Under Assumptions 1-3, with $\theta^0 \in Int(\Theta)$,

$$diag\left(T^{1/2}, T^{1/2-d_i^0}\right)\left(\hat{\theta}_i - \theta_i^0\right) \xrightarrow{d} N\left(\underline{0}, D_i\left(d_i^0, \lambda_i^0, \lambda_{i-1}^0\right)\right)$$

where

$$D_i\left(\lambda_{i-1}^0, \lambda_i^0, d_i^0\right) = \begin{pmatrix} \frac{6}{\pi^2}\left(\lambda_i^0 - \lambda_{i-1}^0\right)^{-1} & 0 \\ 0 & \sigma^2\left(\frac{\Gamma^2(1-d_i^0)(1-2d_i^0)}{\left(\lambda_i^0-\lambda_{i-1}^0\right)^{1-2d_i^0}} + D_{ii}^{\mu}\left(\lambda_{i-1}^0, \lambda_i^0, d_i^0\right)\right) \end{pmatrix}$$

where $\hat{d}_i$ and $\hat{\mu}_j$ are uncorrelated for $i, j = 1, 2$, and $\hat{d}_i$ and $\hat{d}_j$ are uncorrelated and $\hat{\mu}_i$ and $\hat{\mu}_j$ are correlated for $i \neq j$.

$D_{ii}^{\mu}\left(\lambda_{i-1}^0, \lambda_i^0, d_i^0\right)$ is the variance component arising from applying the too short differencing filter on the fractionally integrated error series

$$D_{ii}^{\mu}\left(\lambda_{i-1}^0, \lambda_i^0, d_i^0\right) = \frac{\Gamma^4\left(1-d_i^0\right)\left(1-2d_i^0\right)^2}{\left(\lambda_i^0 - \lambda_{i-1}^0\right)^{2-4d_i^0}} A_i^{\mu}\left(\lambda_{i-1}^0, \lambda_i^0, d_i^0\right), \tag{6}$$

where

$$A_i^{\mu}\left(\lambda_{i-1}^0, \lambda_i^0, d_i^0\right) = \tag{7}$$
$$\lim_{T\to\infty} T^{-1} \sum_{k=1}^{[\lambda_{i-1}^0 T]} \left(T^{d_i^0} \sum_{t=1}^{[(\lambda_i^0-\lambda_{i-1}^0)T]} \pi_{t-1}\left(d_i^0 - 1\right) \sum_{l=0}^{t} \pi_l\left(d_i^0\right) \pi_{[\lambda_{i-1}^0 T]+t-l-k}\left(-d_i^0\right)\right)^2.$$

The covariance $\sigma^2 D_{ij}^{\mu}(\{\lambda_{k-1}^0, \lambda_k^0, d_k^0\}_{k=i,j})$ defined as (32) in the Appendix as well as the variance component $D_{ii}^{\mu}\left(\lambda_{i-1}^0, \lambda_i^0, d_i^0\right)$ of the mean estimators μ_i and μ_j are functions of $\{\lambda_{k-1}^0, \lambda_k^0, d_k^0\}_{k=i,j}$ and have to be numerically approximated. We estimate the covariance matrix of the estimator by replacing $\{\lambda_{i-1}^0, \lambda_i^0, d_i^0\}$, D_{ii}^{μ} and D_{ij}^{μ} by their estimates and $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\hat{u}_t^2$.

Finally, if there is some short run dynamics in the form of a stable and known causal AR(p) structure,

$$\alpha(L)\left(y_t - \mu_i^0\right) = \Delta_t^{-d_i^0}\varepsilon_t,\ T_{i-1}^0 < t \leq T_i^0, \tag{8}$$

the mean estimation behaves as in Theorem 3. The memory estimator is correlated with the estimator of the AR component. In particular,

$$Var\left(T^{1/2}(\hat{d}_i - d_i^0)\right) = \omega^{-2}\left(\lambda_i^0 - \lambda_{i-1}^0\right)^{-1},$$

where $\omega^2 = \frac{\pi^2}{6} - \kappa'\Phi^{-1}\kappa$ is defined as in Lobato and Velasco (2007). $\kappa = (\kappa_1, ..., \kappa_p)'$ and $\kappa_k = \sum_{j=k}^{\infty} j^{-1}c_{j-k}, k = 1, ..., p$ where c_j are the coefficients of L^j in the expansion of $1/\alpha(L)$. $\Phi = [\Phi_{k,j}], \Phi_{k,j} = \sum_{j=0}^{\infty} c_t c_{t+|k-j|}, k, j = 1, ..., p$ denotes the Fisher information matrix for α under Gaussianity. The proof follows from combining Hualde and Robinson (2010) and our Theorem 3.

4. TESTS

Up to now, we have assumed that the number of breaks is known. In the following, we analyze some tests for determining the number of breaks if this number is unknown.

4.1. F-test of 0 versus k breaks

First, we consider the hypothesis of no breaks and the alternative of k breaks, where in practice k is a small number:

$$H_0 : m = 0 \text{ vs. } H_1 : m = k.$$

Let $\boldsymbol{\lambda}$ denote a break fraction partition satisfying the standard assumption of asymptotic distinctiveness and distance to the end-points. In particular, $\boldsymbol{\lambda}$ belongs to the subset

$$\Lambda_\epsilon = \{\boldsymbol{\lambda} \equiv (\lambda_1, ..., \lambda_k) : |\lambda_{i+1} - \lambda_i| \geq \epsilon, \lambda_i \geq \epsilon, \lambda_i \leq 1 - \epsilon\}$$

with $\epsilon > 0$. Given a break partition $\boldsymbol{\lambda}$**,** let

$$SSR_k(\boldsymbol{\lambda}) = \min_{\theta_1,...,\theta_{k+1}} \sum_{i=1}^{k+1} \sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_i T]} \left(\Delta_t^{d_i}(y_t - \mu_i)\right)^2 \tag{9}$$

denote the minimized sum of squared residuals under the alternative hypothesis of k breaks. Note that this filter differs from the previous filter (4) in being truncated at 1 rather than at $[\lambda_{i-1}T]$. This filter is the appropriate one under H_0. In consequence, also the test statistic will be constructed under the assumption that H_0 is true. From (9), we obtain the unconstrained estimators in the $k+1$ regimes, $(\hat{\theta}_1, ..., \hat{\theta}_{k+1})$, given the break partition $\boldsymbol{\lambda}$. Equally, SSR_0 denotes the minimized sum of squares under the hypothesis of no breaks. As in Bai and Perron (1998) and Boldea and Hall (2010), we use a sup F-type test

$$\sup_{\boldsymbol{\lambda}\in\Lambda_\in} F_T^{\vartheta}(\boldsymbol{\lambda}, k; p) = \sup_{\boldsymbol{\lambda}\in\Lambda_\in} \frac{(SSR_0 - SSR_k(\boldsymbol{\lambda}))/kp}{SSR_k(\boldsymbol{\lambda})/[T-(k+1)p]}. \tag{10}$$

The number of changing parameters p is *one* or *two*. The superscript $\vartheta \in \{d, \mu, (d, \mu)\}$ denotes the parameter in which we are testing for breaks. ϵ is a fixed small number. The larger ϵ is, the larger is the power, but the test might become inconsistent, if Λ_ϵ does not contain the true break fraction under the alternative. For the break only in the memory (mean), $SSR_k(\boldsymbol{\lambda})$ constraints the mean (memory) to be constant over the regimes.

Since from (9), the same μ_i is subtracted from observations with true mean μ_j^0 of all regimes $j \le i$, the mean μ_i, $i > 1$, is inconsistently estimated under the alternative hypothesis. This does not happen for the memory estimator d_i since the terms arising from applying the wrong filter are negligible. Alternatively, the filter (4) from Sections 2 and 3 would solve this problem of inconsistent estimation under the alternative. However, for determining the asymptotic distribution of $\sup_{\boldsymbol{\lambda}} F_T$ under H_0, the filter in expression (9) is more appropriate. The asymptotic distribution resembles the one of Bai and Perron (1998) and the size properties are better. Despite the estimators are inconsistent, this test has power, as we show in Theorem 5.

We consider the following local alternative for assessing the power of the tests for processes close to H_0,

$$H_{1,T} : d_t^0 = d_1^0 + T^{-1/2} h_d\left(\frac{t}{T}\right) \ \ and \ \ \mu_t^0 = \mu_1^0 + T^{d_1^0 - 1/2} h_\mu\left(\frac{t}{T}\right).$$

As in Lazarová (2005), $h_j(\frac{t}{T}), j = d, \mu$, is a bounded variation function on [0,1]. This local alternative comprises many types of structural change models. A function $h(\tau) = \sum_{j=1}^{i} \delta_j I\left(\lambda_j^0 \le \tau\right)$ describes abrupt breaks of size δ_i at time $[\lambda_i^0 T]$. A function h consisting of constant segments connected by smooth curves describes a smooth transition between the different levels of the parameter. Finally, a general smooth function of h describes continual change of the parameters.

Let

$$\tilde{W}_{1/2-d_1^0}(\lambda) = \int_0^{\lambda} s^{-d_1^0} dB(s) \tag{11}$$

be a variant of a fractional Brownian Motion with a particular covariance structure,

$$Cov\left(\tilde{W}_{1/2-d_1^0}(\lambda_i), \tilde{W}_{1/2-d_1^0}(\lambda_{i-1})\right) = \frac{\lambda_{i-1}^{1-2d_1^0}}{\Gamma\left(1-d_1^0\right)\left(1-2d_1^0\right)}. \tag{12}$$

Further, let

$$B^h(\lambda_i) = B(\lambda_i) - \frac{\pi}{\sqrt{6}} \int_0^{\lambda_i} h_d(u)\, du \tag{13}$$

and

$$\tilde{W}^h(\lambda_i) = \tilde{W}_{1/2-d_1^0}(\lambda_i) - \frac{\int_0^{\lambda_i} u^{-2d_1^0} h_\mu(u)\, du}{\Gamma\left(1-d_1^0\right)\sqrt{1-d_1^0}}, \tag{14}$$

where the second terms reflect the local drift for the break in memory and in mean respectively. Finally, let

$$F_i^d(\boldsymbol{\lambda},k,1) = \frac{\left(\lambda_i B^h(\lambda_{i+1}) - \lambda_{i+1} B^h(\lambda_i)\right)^2}{\lambda_i \lambda_{i+1}(\lambda_{i+1}-\lambda_i)}, \tag{15}$$

$$F_i^\mu(\boldsymbol{\lambda},k,1) = \frac{\left(\lambda_i^{1-2d_1^0}\tilde{W}^h(\lambda_{i+1}) - \lambda_{i+1}^{1-2d_1^0}\tilde{W}^h(\lambda_i)\right)^2}{\lambda_i^{1-2d_1^0}\lambda_{i+1}^{1-2d_1^0}\left(\lambda_{i+1}^{1-2d_1^0} - \lambda_i^{1-2d_1^0}\right)} \quad \text{and} \tag{16}$$

$$F_i^{(d,\mu)}(\boldsymbol{\lambda},k,2) = F_i^d(\boldsymbol{\lambda},k,1) + F_i^\mu(\boldsymbol{\lambda},k,1). \tag{17}$$

Theorem 4 provides the asymptotic distribution of the test statistic for breaks in both parameters under the local alternative.

THEOREM 4. *(Asymptotic distribution of the test)*
Under Assumptions 1-2 and under $H_{1,T}$,

$$\sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon} F_T^\vartheta(\boldsymbol{\lambda},k;p) \xrightarrow{d} \sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon} \frac{1}{pk}\sum_{i=1}^k F_i^\vartheta(\boldsymbol{\lambda},k,p),$$

where the superscript $\vartheta \in \{d, \mu, (d,\mu)\}$ denotes the parameters in which we are testing for breaks.

For the local alternative $H_{1,T}$, the distribution of the test statistic depends on the shape of the $h-$functions and depends therefore on the true break fractions if the h-functions depend, e.g. for h being a *stepfunction* in the break fractions λ_i^0.

The asymptotic distribution of the test differs from the one in Bai and Perron (1998) and depends on both standard and fractional Brownian Motion. The terms corresponding to the estimation of memory and mean are additive because of their independent estimation. If we test for breaks only in the memory, $F_i^d(\boldsymbol{\lambda},k,1)$ corresponds to the one of Bai and Perron (1998) and if we test for breaks only in the mean, the limit distribution $F_i^\mu(\boldsymbol{\lambda},k,1)$ depends on the nuisance parameter d_1^0. $F_i^\mu(\boldsymbol{\lambda},k,1)$ resembles the one for a break in the memory with fractional rather than standard Brownian Motions. In practice, we estimate the memory and compare the test statistic to critical values obtained from simulating the test statistic for a grid of different values of d and fitting a polynomial in d. The validity of this approach follows from Giraitis *et al.* (2003).

Corollary 1 provides the distribution of the test statistic for one break in both parameters under the specific local break hypothesis

$$H'_{1,T} : h_\vartheta(\tau) = \delta_\vartheta I\left(\lambda_j^0 \leq \tau\right), \vartheta = \{d, \mu, (d, \mu)\}.$$

COROLLARY 1. *Under Assumptions 1-2 and under $H'_{1,T}$,*

$$\sup_{\boldsymbol{\lambda}\in\Lambda_\in} F_T^{d,\mu}(\boldsymbol{\lambda},1;2) \xrightarrow{d} \sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon} \frac{\left[\lambda B(1) - B(\lambda) - \delta_d \frac{\pi}{\sqrt{6}}\left(\min\{\lambda, \lambda_1^0\}(1-\max\{\lambda,\lambda_1^0\}\right)\right]^2}{\lambda(1-\lambda)}$$
$$+\frac{\left[\lambda \tilde{W}_{1/2-d_1^0}(1) - \tilde{W}_{1/2-d_1^0}(\lambda) - \delta_\mu \frac{\left((\min\{\lambda,\lambda_1^0\})^{1-2d_1^0}\left(1-(\max\{\lambda,\lambda_1^0\})^{1-2d_1^0}\right)\right)}{\Gamma(1-d_1^0)\sqrt{1-d_1^0}}\right]^2}{\lambda^{1-2d_1^0}\left(1-\lambda^{1-2d_1^0}\right)}.$$

The proof follows from substituting $h_\vartheta(\tau) = \delta_\vartheta I\left(\lambda_j^0 \leq \tau\right), \vartheta = \{d, \mu, (d, \mu)\}$, in Theorem 4. From Corollary 1, because of symmetry, the local power is highest for $\lambda_1^0 = 1/2$.

We focus on tests for *one* break and we simulate the critical values for a grid of d^0 for $\alpha = 0.05$ and $\epsilon = 0.15$. For a break in both parameters they are shown in the first line of Table 1 and for a break only in the mean, they are shown in the second line. For a break only in the memory, the critical value corresponds to the one in Bai and Perron (1998), $CV_d = 8.57$.

For establishing the consistency of the test, we have to analyze the estimator using the filter in expression (9) under H_1. Similar to Theorems 1 and 2, the break fractions are also consistently estimated at rate T. Thus, we can treat them as if they were known. Next, while the memory estimators $\hat{d}_1, ..., \hat{d}_{k+1}$ are still consistent, the mean estimators $\hat{\mu}_2, ..., \hat{\mu}_{k+1}$ are inconsistent because the applied filters mix observations of the different regimes and converge to weighted averages of the true means of the corresponding and the preceding regimes. Using these results, Theorem 5 provides the consistency of the test.

THEOREM 5. *(Consistency of the test)*

Under Assumptions 1-3 for $k > 0$ breaks,

a) The test for breaks in both parameters diverges at rate T under $H_1^{d,\mu}$ and under H_1^d, and diverges at rate T^{1-2d^0} under H_1^μ.

b) The test for breaks in the memory diverges at rate T under $H_1^{d,\mu}$ and H_1^d.

c) The test for breaks in the mean diverges at rate T^{1-2d^0} under H_1^μ and at rate $T^{1-2\min\{d_1^0,d_2^0\}}$ under $H_1^{d,\mu}$.

Thus, the tests are consistent with a rate of divergence that depends on which parameters are changing. In consequence, for a d^0 close to 1/2, the test for a break only in the mean has low power under the alternative.

TABLE 1
Critical Values of F-test for breaks in μ and d and only in μ.

d^0	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.49
CV	11.6	11.6	11.5	11.5	11.4	11.4	11.4	11.3	11.2	11.2	11.1
CV_μ	8.6	8.5	8.5	8.4	8.4	8.2	8.2	8.1	8.0	7.9	7.9

Finally, if the error term has the stable and known short run dynamics structure ARFI(p,d) in (8), expression (13) in Theorem 4 becomes

$$B^h(\lambda_i) = B(\lambda_i) - \omega \int_0^{\lambda_i} h_d(u)\, du,$$

where $\omega^2 = \frac{\pi^2}{6} - \kappa'\Phi^{-1}\kappa$ is defined in the end of Section 3. A solution to an unknown stable structure is discussed in the empirical application in Section 6.

4.2. F-test of ℓ versus $\ell + 1$ breaks

We consider the following hypothesis

$$H_0 : m = \ell \text{ vs. } H_A : m = \ell + 1.$$

Technically, we impose ℓ breaks and test each segment for an additional break. The test statistic corresponds to the one in Bai and Perron (1998),

$$F_T(\ell + 1|\ell) = \max_{1 \le i \le \ell} \frac{1}{\hat{\sigma}_i^2} \left\{ S_T(\hat{T}_{i-1}, \hat{T}_i) - \inf_{\tau \in \Delta_{i,l}} S_T(\hat{T}_{i-1}, \tau, \hat{T}_i) \right\}$$

where

$$\Delta_{i,l} = \left[\tau : \hat{T}_{i-1} + (\hat{T}_i - \hat{T}_{i-1})\eta \le \tau \le \hat{T}_i - (\hat{T}_i - \hat{T}_{i-1})\eta \right]$$

and

$$\hat{\sigma}_i^2 \xrightarrow{p} \sigma_i^2 = \sigma^2.$$

Following the same logic as in the test of *zero* against k breaks, we choose the filter truncated at $\hat{T}_{i-1}$ which is appropriate under H_0. The underlying constrained estimator (assuming one regime for the interval $[\hat{T}_{i-1} + 1, \hat{T}_i]$) is the one discussed in Theorem 3. For estimating the regime $[\tau, \hat{T}_i]$, the filter is still truncated at $\hat{T}_{i-1}$ rather than at τ and thus differs from the one used in Sections 2-3. Therefore, similar to Section 4.1, the mean estimate is not consistent under the alternative. Yet, the test is still consistent.

We consider again a local break in all regimes: *For* $i = 1, ..., \ell$ and $t = T_{i-1}^0 + 1, ..., T_i^0$,

$$\begin{aligned} H_{1T}^{\ell} \quad : \quad & d_{i,t}^0 = d_i^0 + T^{-1/2} h_d\left(\frac{t - T_{i-1}^0}{T_i^0 - T_{i-1}^0}\right) \textit{ and} \\ & \mu_{i,t}^0 = \mu_i^0 + T^{d_i^0 - 1/2} h_\mu\left(\frac{t - T_{i-1}^0}{T_i^0 - T_{i-1}^0}\right). \end{aligned}$$

There is a local break in all regimes with $h_d(\tau)$ and $h_\mu(\tau)$ as defined in $H_{1,T}$.

First,

$$T^{-1/2}\sum_{k=1}^{T_{i-1}^0}\left(T^{d_i^0}\sum_{t=1}^{\left[\gamma\left(T_i^0-T_{i-1}^0\right)\right]}\pi_{t-1}\left(d_i^0-1\right)\sum_{l=0}^{t}\pi_l\left(d_i^0\right)\pi_{T_{i-1}^0+t-l-k}\left(-d_i^0\right)\right)u_k \quad (18)$$

converges in distribution to $C\left(\lambda_{i-1}^0,\gamma,d_i^0\right)$, a Gaussian process with mean zero and variance (7) with $\lambda_i^0=\gamma T_i^0/T+(1-\gamma)T_{i-1}^0/T$. Tightness in γ follows from arguments similar to the ones in Lemma 1. Next let $G_{2,\eta}^{(d,\mu);(i)}(x)$ be the distribution function of

$$\sup_{\eta\leq\gamma\leq1-\eta}\left\{\frac{\left(B^h(\gamma)-\gamma B^h(1)\right)^2}{\gamma(1-\gamma)}+\frac{\hat{W}_i^h(\gamma)-\gamma^{1-2d_i^0}\hat{W}_i^h(1)}{\gamma^{1-2d_i^0}\left(1-\gamma^{1-2d_i^0}\right)}\right\}, \quad (19)$$

where $B^h(\gamma)$ is defined in (13) and where

$$\hat{W}_i^h(\gamma)=\tilde{W}^h(\gamma)+\frac{\left(1-2d_i^0\right)\Gamma^2\left(1-d_i^0\right)C\left(\lambda_{i-1}^0,\gamma,d_i^0\right)}{\left(\lambda_i^0-\lambda_{i-1}^0\right)^{1-2d_i^0}}. \quad (20)$$

The first term of (20) corresponds to (14) with one local break. For $G_{2,\eta}^{d;(i)}(x)$, the second term in (19) drops and for $G_{2,\eta}^{\mu;(i)}(x)$ the first term in (19) drops. Theorem 6 provides the asymptotic distribution for testing for a $(\ell+1)$'s break in both parameters.

Theorem 6. *(Asymptotic distribution of the test for ℓ vs. $\ell+1$ breaks) Under Assumptions 1,2 and under H_{1T}^{ℓ},*

$$\lim_{T\to\infty}P\left(F_T(\ell+1|\ell)\leq x\right)=\Pi_{i=1}^{\ell+1}G_{p,\eta}^{\vartheta,(i)}(x),\ \vartheta\in d,\mu,(d,\mu).$$

For the test for a break only in the memory, the test statistic behaves as the one in Bai and Perron (1998). The critical value x_α is the value x for which $G_{p,\eta}^{d}(x)=\alpha^{1/(l+1)}$ and the critical values are the ones tabulated in Bai and Perron (1998). For the test for a break only in the mean, the distribution depends on the variant of fractional Brownian motion (11) plus the additional term coming from applying the too short filter. For this test and for the test for a break in both parameters, $G_{p,\eta}^{\vartheta,(i)}(x),\vartheta=\{\mu,(d,\mu)\}$, differs between the regimes and, consequently, the critical value x_α is the value x for which $\Pi_{i=1}^{\ell+1}G_{p,\eta}^{\vartheta,(i)}(x)=\alpha$. The asymptotic distribution depends on $(d_1^0,...,d_{\ell+1}^0)$ and $(\lambda_1^0,...,\lambda_\ell^0)$. As a consequence, the critical values are obtained on a case-by-case basis given the estimated break partition and memory parameters. Further, the additional term in (20) introduces some dependence between the distribution function in the different regimes that has to be taken into account when simulating the critical values. Therefore, it is clear that using

TABLE 2

Test for a joint break in memory and mean.

a) Size. Rejection probabilities when there is no break.

T\d^0	0.05	0.15	0.25	0.35	0.45
200	2.2	6.7	10.0	11.3	13.0
500	3.4	7.5	9.6	9.7	8.8
1000	3.5	6.5	7.0	8.0	7.3

b) Power. Rejection probabilities when there is a break at the half of the sample.

		T=200				T=500			
	$d_2^0 \backslash\backslash \mu_2^0$	0.5	**1**	1.5	2	0.5	**1**	1.5	2
d_1^0=0.05	**0.05**	48.2	2.7	50.9	98.1	91.3	3.8	90.9	100.0
	0.10	44.0	4.3	41.7	95.5	83.7	5.4	82.8	100.0
	0.25	45.8	21.8	45.0	84.1	78.7	56.9	81.0	98.7
	0.45	78.6	75.7	80.8	86.5	99.8	99.4	99.2	99.6
d_1^0=0.25	0.05	49.6	21.4	46.8	90.8	85.5	56.5	86.1	99.7
	0.25	22.9	10.7	23.1	54.8	28.8	9.6	27.8	75.3
	0.30	23.1	13.8	23.0	48.9	27.3	13.2	28.0	64.5
	0.45	35.6	31.0	35.7	50.5	65.7	61.5	63.7	73.6
d_1^0=0.45	0.05	83.6	77.6	83.7	91.3	99.8	99.6	99.6	100.0
	0.25	38.9	31.6	38.3	56.8	71.9	61.1	68.4	81.2
	0.40	18.8	16.5	17.9	28.4	17.1	15.1	17.7	27.6
	0.45	16.5	14.8	16.8	26.5	12.3	9.8	12.2	22.4

this test is unfeasible. To overcome this problem in practice, we suggest using the bootstrap, which we discuss in the next section.

The consistency and rates of divergence of $F_T\left(l+1|l\right)$ follow from using a similar argument as the one for the consistency of the $\sup_\lambda F\left(\lambda,1,p\right)$ test for the segment that contains the additional break in Theorem 5.

5. MONTE CARLO ANALYSIS USING ASYMPTOTIC CRITICAL VALUES

In this section, we analyze size and power of the three tests discussed in Section 4.1, $\sup_{\boldsymbol{\lambda}} F_T^d$, $\sup_{\boldsymbol{\lambda}} F_T^\mu$, $\sup_{\boldsymbol{\lambda}} F_T^{d,\mu}$. For simplicity we analyze the case of one break, using the critical values provided in Table 1. In all following simulations the number of simulations is $1,000$, the distance to the endpoints of the sample $\varepsilon = 0.15$, the significance level $\alpha = 0.05$ and the sample sizes are $T = 200, 500$ and $1,000$ for the size and 200 and 500 for the power. We assume an error variance $\sigma^2 = 1$. Since asymptotic results are invariant to the level of the mean, we take $\mu^0 = 1$ if the mean is constant and $\mu_1^0 = 1$ for the mean in the first regime if it is changing. For the size, we analyze $d^0 = 0.05, 0.15, 0.25, 0.35$ and 0.45. For the power, we consider breaks in the mean from $d_1^0 = 0.05$ to $d_2^0 = 0.1$, 0.25 and 0.45, from $d_1^0 = 0.25$ to $d_2^0 = 0.05, 0.3$ and 0.45 and from $d_1^0 = 0.45$ to $d_2^0 = 0.05, 0.25$ and 0.4. Further, we consider breaks in the mean from $\mu_1^0 = 1$ to $\mu_2^0 = 0.5, 1.5$ and 2. The break fraction is always at the

TABLE 3

Test for a break in the memory.

a) Size. Rejection probabilities when there is no break.

T\d^0	0.05	0.15	0.25	0.35	0.45
200	1.5	4.2	7.3	9.2	7.4
500	2.3	6.3	8.3	8.4	5.5
1000	2.7	6.5	6.8	7.2	4.8

b) Power. Rejection probabilities when there is a break at the half of the sample.

	d_1^0 =0.05				d_1^0 =0.25				d_1^0 =0.45			
$T\backslash\backslash d_2^0$	**0.05**	0.1	0.25	0.45	0.05	**0.25**	0.3	0.45	0.05	0.25	0.4	**0.45**
200	1.2	1.8	20.8	84.3	25.6	7.7	8.9	29.3	84.5	36.6	10.6	8.7
500	1.2	4.5	64.1	99.6	66.6	8.9	12.3	67.9	99.8	67.9	11.2	7.1

half of the sample ($\lambda_1^0 = 0.5$).

First, Table 2a) shows the size of a test for a break in both parameters. The estimator of the memory is constrained to lie in the interval $[0, 1/2)$ which naturally has a negative effect on the size in finite samples. This negative effect is largest for $d = 0.05$ and decreases as the sample size increases. For larger memory parameter, the test is oversized in finite samples. This happens because even if the estimation of memory and mean is asymptotically uncorrelated, in finite samples it is still correlated. Table 2b) analyzes the power of this test. The power increases in the sample size. In general, a break in the memory is only detectable for larger break sizes. Further, the detectability of a break in the mean decreases considerably in d_0^2 since the higher the true memory in the two regimes, the less precisely the means are estimated. For a non-changing memory of 0.45, the break in the mean is not detected even for larger samples.

Next, we analyze the behavior of the test for a break only in the memory. Table 3a) shows the size properties of this test. For $d^0 = 0.05$, the size is too low because of the constrained estimation of the memory. This size distortion vanishes slowly. For larger memory parameters, the test is again slightly oversized. Next, Table 3b) shows that the test has power for detecting a break for not too small breaks in the memory. Since the size of a test for a break only in the memory is smaller than the one of a test for a break in both parameters, its power is also smaller.

Finally, we analyze size and power of a test for a break only in the mean. Table 4a) displays the size properties of such a test. This test is also slightly oversized. Finally, Table 4b) displays the power. Because of the imprecise estimation, a test of a break in the mean has low power when the true memory is close to 0.5. This confirms the lower rate of divergence in Theorem 5.

TABLE 4

Test for a break in the mean.

a) Size. Rejection probabilities when there is no break.

T\d^0	0.05	0.15	0.25	0.35	0.45
200	6.5	11.2	11.5	11.0	12.8
500	7.3	8.3	8.7	8.2	8.5
1000	7.1	7.1	6.4	7.3	6.0

b) Power. Rejection probabilities when there is a break at the half of the sample.

	T=200				T=500			
$d^0\backslash\backslash\mu_2^0$	0.5	**1**	1.5	2	0.5	**1**	1.5	2
0.05	65.1	7.9	70.8	99.6	95.8	8.5	96.8	100.0
0.25	29.7	12.5	28.2	65.3	36.2	10.0	33.7	82.3
0.45	20.2	14.6	14.9	22.6	15.5	10.0	14.5	22.1

TABLE 5

Robustness of tests for a break in one parameter.

a) Size of test for a break in the memory if there is a break in the mean.

	T=200				T=500			
$d^0\backslash\backslash\mu_2^0$	0.5	**1**	1.5	2	0.5	**1**	1.5	2
0.05	13.2	6.3	14.0	46.0	20.6	4.5	21.1	79.3
0.25	15.5	8.8	13.2	23.7	10.7	7.9	11.0	22.5
0.45	11.9	11.8	12.3	13.1	6.5	8.7	8.5	8.2

b) Size of test for a break in the mean if there is a break in the memory.

	d_1^0 =0.05				d_1^0 =0.25				d_1^0 =0.45			
T$\backslash\backslash d_2^0$	**0.05**	0.1	0.25	0.45	0.05	**0.25**	0.3	0.45	0.05	0.25	0.4	**0.45**
200	7.9	10.0	19.7	39.2	16.6	13.3	13.6	25.5	23.5	14.3	11.9	14.1
500	8.5	11.7	21.2	41.5	15.2	11.4	12.0	25.6	25.8	12.7	9.1	10.5

6. IDENTIFIABILITY OF CHANGING PARAMETERS

Up to now, we have analyzed the behavior of tests in situations for which they are designed. In this section, we analyze tests for breaks in one parameter for the case that the other parameter is changing. Table 5a) shows that the test for a break in the memory is highly oversized if the mean is changing. The reason is that as mentioned in the end of Section 2, a break in the mean affects the estimation of the memory in finite samples. Table 5b) shows that the same is true when testing for a break in the mean if the memory is changing. The intuition is that the mean is estimated at different rates of convergence under the alternative and therefore the difference between SSR_0 and SSR_1^μ becomes too large. Therefore, we cannot distinguish between breaks in the memory and breaks in the mean and it is not possible to identify the changing parameter.

First, we focus on testing for a break in the memory when the mean is changing.

To solve the mentioned problem we suggest a Chow type test. Let

$$SSR'_k(\boldsymbol{\lambda}) = \min_{\theta_1,\dots,\theta_{k+1}} \sum_{i=1}^{k+1} \sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_i T]} \left(\Delta_{t-[\lambda_{i-1}T]}^{d_i}(y_t-\mu_i)\right)^2$$

denote the minimized sum of squares under the alternative of a break in the memory and in the mean given a partition $\boldsymbol{\lambda}$. The estimate of the corresponding break fraction is

$$\hat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda}} SSR'_k(\boldsymbol{\lambda}).$$

As in Sections 2 and 3, the filter is truncated at $[\hat{\lambda}_{i-1}T]$ rather than at 1. Next, we use the estimated partition $\hat{\boldsymbol{\lambda}}$ to estimate under the null a constant memory and a changing mean with the corresponding minimized sum of squares

$$SSR_0^d\left(\hat{\boldsymbol{\lambda}}\right) = \min_{d,\mu_1,\dots,\mu_{k+1}} \sum_{i=1}^{k+1} \sum_{t=[\hat{\lambda}_{i-1}T]+1}^{[\hat{\lambda}_i T]} \left(\Delta_{t-[\hat{\lambda}_{i-1}T]}^{d}(y_t-\mu_i)\right)^2.$$

For testing for a break in the mean, we estimate under the null of a constant mean and a changing memory with the corresponding minimized sum of squares $SSR_0^{\mu}(\hat{\boldsymbol{\lambda}})$. For simplicity, we consider the case of one break. Let

$$F_T^{\vartheta}\left(1;1|\hat{\lambda}_1\right) = \frac{\left(SSR_0^{\vartheta}\left(\hat{\boldsymbol{\lambda}}\right) - SSR'_k\left(\hat{\boldsymbol{\lambda}}\right)\right)}{SSR'_k\left(\hat{\boldsymbol{\lambda}}\right)/(T-2)},\ \vartheta = d,\mu, \tag{21}$$

be the test statistic for testing for a break in the memory and the mean respectively.

For testing for a break in the memory under the maintained hypothesis of a break in the mean, we assume a local break in the memory and a break in the mean

$$H_{1,T}^{d,\mu_1^0\neq\mu_2^0} : d_t^0 = d_1^0 + T^{-1/2}h_d\left(\frac{t}{T}\right).$$

For testing for a break in the mean under the maintained hypothesis of a break in the memory, we assume a local break in the mean and a break in the memory

$$\begin{aligned} H_{1,T}^{\mu,d_1^0>d_2^0} &: \mu_t^0 = \mu_1^0 + T^{-1/2+d_1^0}h_\mu\left(\frac{t}{T}\right) \text{ or} \\ H_{1,T}^{\mu,d_1^0<d_2^0} &: \mu_t^0 = \mu_1^0 + T^{-1/2+d_2^0}h_\mu\left(\frac{t}{T}\right). \end{aligned}$$

Proposition 2a) (b)) discusses the asymptotic distribution of the test for a break in the memory (mean) when the mean (memory) is changing.

Proposition 2. *(Asymptotic distribution of the test for a break in one parameter under the maintained hypothesis of break in other parameter)* *a) Under Assumptions 1-2 and under* $H_{1,T}^{d,\mu_1^0 \neq \mu_2^0}$,

$$F_T^d\left(1;1|\hat{\lambda}_1\right) \xrightarrow{d} \chi_1^2\left(c_1\right),$$

where $c_1 = \frac{\pi^2}{6} \frac{\left(\lambda_1^0 \int_0^1 h_d(u)du - \int_0^{\lambda_1^0} h_d(u)du\right)^2}{\lambda_1^0(1-\lambda_1^0)}$.

b) Under Assumptions 1-2 and under $H_{1,T}^{\mu,d_1^0 > d_2^0}$,

$$F_T^\mu\left(1;1|\hat{\lambda}_1\right) \xrightarrow{d} \chi_1^2\left(c_2\right).$$

where $c_2 = \frac{\left(\int_0^{\lambda_1^0} u^{-2d_1^0} h_\mu(u)du\right)^2}{(\lambda_1^0)^{1-2d_1^0}\Gamma^2(1-d_1^0)(1-2d_1^0)}$.

c) Under Assumptions 1-2 and under $H_{1,T}^{\mu,d_1^0 < d_2^0}$,

$$F_T^\mu\left(1;1|\hat{\lambda}_1\right) \xrightarrow{d} \left(1 + D_{22}^\mu\left(\lambda_1^0,1,d_2^0\right)\right)\chi_1^2\left(c_3\right).$$

where $D_{22}^\mu\left(\lambda_1^0,1,d_2^0\right)$ *is defined in (6) and* $c_3 = \frac{1}{(1+D_{22}^\mu(\lambda_1^0,1,d_2^0))} \frac{\left(\int_0^1 u^{-2d_2^0} h_\mu(u)du\right)^2}{\left(1-(\lambda_1^0)^{1-2d_2^0}\right)\Gamma^2(1-d_2^0)(1-2d_2^0)}$.

First, when there is a break in the mean, the estimator of the break partition $\hat{\boldsymbol{\lambda}}$ converges at rate T to the true break fraction (from Theorem 2). This rate is superconsistent and we can treat the break fraction as known. Therefore, the asymptotic distribution in Parts a) and b) corresponds to the one of a Chow test and the critical values are taken from a χ_1^2. For Part c), because of the too short filter, the asymptotic distribution is not distribution free and we have to simulate the critical values. Finally, if we do not know the direction of the break in the memory, in order to control the size, we choose the critical values from case c) since they are the larger ones. Proposition 2 can be generalized to k breaks. Since the test is also consistent, this procedure makes it possible to distinguish between a break in the memory (mean) and a break in both parameters.

On the other hand, if there are no breaks, $\hat{\boldsymbol{\lambda}}$ converges to a spurious limit and the test statistic behaves asymptotically not as in Proposition 2 but similar to the one in Theorem 4 (the difference comes from the different filters). The critical values from Proposition 2 are not the right ones for this case and we overreject. However, this case only happens with probability α (probability of erroneously rejecting $H_0 : d_1 = d_2$ & $\mu_1 = \mu_2$ in the first step). Thus, the size is controlled.

In practice, we can apply the following sequential testing strategy:

1) Test H_0 vs. $H_1 : d_1 \neq d_2$ and/or $\mu_1 \neq \mu_2$ (Corollary 1).
   (i) If do not reject $\rightarrow$ conclude there are no breaks. Stop.
   (ii) If reject $\rightarrow$ conclude there are breaks. $\rightarrow$ 2a) and 2b).

2a) Test $H_0^{d,\mu_1^0 \neq \mu_2^0}$ vs. $H_1 : d_1 \neq d_2$ & $\mu_1 \neq \mu_2$ (Prop. 2a))

(i) If do not reject $\rightarrow$ conclude the memory is not changing.

(ii) If reject $\rightarrow$ conclude the memory is changing.

2b) Test $H_0^{\mu,d_1^0 \neq d_2^0}$ vs. $H_1 : \mu_1 \neq \mu_2$ & $d_1 \neq d_2$ (Prop. 2b)/c))

(i) If do not reject $\rightarrow$ conclude the mean is not changing.

(ii) If reject $\rightarrow$ conclude the mean is changing.

All tests in this sequential procedure are consistent. The size is α for the tests in step 1 and in step 2a) and 2b) if the respective maintained hypothesis is true. If the mean (memory) is not changing in step 2a) (2b)), the size is $\beta_1 \cdot \alpha$ ($\beta_2 \cdot \alpha$) where β_1 (β_2) denotes the probability of rejecting in the step 2a) (2b)) after having rejected in step 1). This probability lies between α and 1 and depends on the relative strength of the signal in the first step. Therefore, the test of the null of $d_1 = d_2$ *versus* $d_1 \neq d_2$ has size $\leq \beta_1 \cdot \alpha \leq \alpha$ regardless of the memory and the test of the null of $\mu_1 = \mu_2$ *versus* $\mu_1 \neq \mu_2$ has size $\leq \beta_2 \cdot \alpha \leq \alpha$.

7. **BOOTSTRAP**

We propose bootstrap procedures for three different situations.

First, we use the bootstrap for the test of breaks in mean and/or memory as a solution to the encountered size distortions due to constrained estimation for d^0 close to 0 and for a higher memory in Tables 2, 3 and 4. For simplicity, we again consider the case of *one* break. We apply the following residual bootstrap for testing for breaks in memory and mean:

1. From the estimation under the null, obtain $\hat{d}, \hat{\mu}$ and $\hat{u}_t$.
2. Resample the residuals $\hat{u}_t$ to obtain u_t^*, and generate

$$y_t^* = \hat{\mu} + \Delta_t^{-\hat{d}} u_t^*.$$

3. From the estimation under the null and alternative for the new series y_t^*, obtain the test statistic

$$\sup_{\boldsymbol{\lambda} \in \Lambda_\epsilon} F_T^* (\boldsymbol{\lambda}, k; p) = \sup_{\boldsymbol{\lambda} \in \Lambda_\epsilon} \frac{(SSR_0^* - SSR_k^* (\boldsymbol{\lambda}))/kp}{SSR_k^* (\boldsymbol{\lambda}) / (T - (k+1)p)}, \tag{22}$$

with

$$SSR_k^* (\boldsymbol{\lambda}) = \sum_{i=1}^{k+1} \frac{1}{T} \sum_{t=T_{i-1}+1}^{T_i} \left(\Delta_t^{\hat{d}_i^*} (y_t - \hat{\mu}_i^*) \right)^2.$$

4. Repeat 2-3 B times and obtain from the empirical distribution the bootstrap critical values.

The obtained residuals are asymptotically *close to iid* under H_0. Since the memory is estimated, we integrate the residuals with $\hat{d}$ rather than with d. Therefore, even under H_0 we cannot use a simple resampling under *iid* but we use instead results of Kapetanios (2010), who analyzes the Sieve bootstrap in a similar context, and his remark about the applicability of the CSS estimator. In contrast to his DGP, ours is more restricted because we do not have a short memory component. Theorem 7 proves the validity of the bootstrap in our context where the difficulty arises from the fact that the memory is estimated.

THEOREM 7. *(Asymptotic behavior of the bootstrap test)*
Under Assumptions 1 and 2 and under H_0 or $H_{1,T}$, the bootstrap based test satisfies

$$P\left(\sup_{\boldsymbol{\lambda}} F_T^*\left(\boldsymbol{\lambda}, k, 2\right) \leq x | y_1, ... y_T\right) \xrightarrow{p} P\left(\sup_{\boldsymbol{\lambda}} F\left(\boldsymbol{\lambda}, k, 2\right) \leq x\right)$$

and the test is consistent.

In practice, we use the unconstrained estimator rather than the constrained one to obtain the residuals in the first step. By doing so, we expect better power properties. This is valid because of Proposition 3.

PROPOSITION 3. *Under H_0,*

$$1) \sup_{\lambda \in [\tau, 1-\tau]} T^{1/2}\left(d_i\left(\lambda\right) - d^0\right) = O_p\left(1\right), \ i = 1, 2.$$
$$2) \sup_{\lambda \in [\tau, 1-\tau]} T^{1/2-d^0}\left(\mu_i\left(\lambda\right) - \mu^0\right) = O_p\left(1\right), \ i = 1, 2.$$

Table 6a) displays Monte Carlo simulations of the size properties of the bootstrap critical values for testing for a break in both parameters. We apply the Warp bootstrap (Giacomini *et al.*, 2007) for all simulations. Not surprisingly, the size properties of the test for breaks in both parameters with bootstrap critical values is closer to the nominal level. Table 6b) provides the power of this test. For testing for breaks only in the memory and only in the mean, we construct corresponding bootstrap procedures.

Finally, if there are short run dynamics of a stable and known ARFI(p, d) structure, the first two steps of the bootstrap change to

1. From the estimation under the null, obtain $\hat{d}, \hat{\mu}, \hat{\alpha}\left(L\right)$ and the residuals

$$\hat{v}_t = \Delta_t^{\hat{d}} \hat{\alpha}^{-1}\left(L\right)\left(y_t - \hat{\mu}\right).$$

2. Resample the residuals $\hat{v}_t$ to obtain v_t^* and generate

$$y_t^* = \hat{\mu} + \hat{\alpha}^{-1}\left(L\right)\Delta_t^{-\hat{d}} v_t^*.$$

TABLE 6

Bootstrap test for a break in memory and mean.

a) Size. Rejection probabilities when there is no break.

T\\d^0	0.05	0.15	0.25	0.35	0.45
200	5.0	6.1	4.2	3.8	4.9
500	5.6	6.2	5.9	5.7	4.4
1000	4.7	4.8	5.3	4.0	4.9

b) Power. Rejection probabilities when there is a break at the half of the sample.

		T=200				T=500			
	$d_2^0 \backslash \mu_2^0$	0.5	**1**	1.5	2	0.5	**1**	1.5	2
d_1^0=0.05	**0.05**	50.1	6.4	53.5	97.6	90.8	5.8	89.8	100.0
	0.10	42.2	5.7	44.6	91.8	79.6	5.5	82.7	99.9
	0.25	36.1	21.8	32.7	74.1	74.2	45.9	77.5	97.3
	0.45	63.7	62.5	71.0	77.8	98.9	99.1	98.6	99.1
d_1^0=0.25	0.05	41.6	20.6	40.7	86.2	78.6	48.7	84.7	98.8
	0.25	17.4	6.4	15.2	42.9	16.1	6.5	23.3	57.7
	0.30	14.9	9.5	17.1	41.5	15.0	7.7	19.4	47.6
	0.45	22.0	20.5	23.2	32.5	52.6	50.1	54.7	61.0
d_1^0=0.45	0.05	72.9	67.3	71.2	85.7	99.5	99.0	99.7	99.8
	0.25	26.6	18.3	25.1	41.0	56.2	49.8	60.9	62.5
	0.40	10.9	9.2	10.2	15.9	8.9	10.3	12.7	15.6
	0.45	7.7	6.6	8.6	17.1	6.4	5.6	10.7	11.7

Second, we analyze a bootstrap procedure for a test for a break in the memory (mean) that is robust to the presence of a break in the mean (memory). Such tests are necessary since the tests defined in Theorem 4 suffer from the size distortions shown in Table 5, and the tests in Proposition 2 require a break in the not tested parameter. For the test for a break in the memory that is robust to the presence of a break in the mean, we apply the following residual bootstrap:

1. From the estimation under the null, minimizing SSR_0^{μ}, obtain $\hat{\lambda}_1, \hat{d}, \hat{\mu}_1, \hat{\mu}_2$ and the residuals $\hat{u}_t$. In line with the procedure described in Sections 2-3, use the filter (4).

2. Resample the residuals $\hat{u}_t$ to obtain u_t^*, and generate

$$y_t^* = \begin{cases} \hat{\mu}_1 + \Delta_t^{-\hat{d}} u_t^*, \ t \leq \hat{\lambda}_1 T \\ \hat{\mu}_2 + \Delta_t^{-\hat{d}} u_t^*, \ t > \hat{\lambda}_1 T. \end{cases}$$

3. From the estimation under the null and the alternative for y_t^*, obtain a bootstrap version of the test statistic (21).

4. Repeat 2-3 B times and obtain from the empirical distribution the bootstrap critical values.

Proposition 4 discusses validity and consistency of the bootstrap procedures in both cases. If the not tested parameter is not changing, the behavior follows from combining Theorem 7 and Proposition 3. If the not tested parameter is changing, the behavior follows from similar arguments as the ones in Proposition 2.

PROPOSITION 4. *(Asymptotic behavior of the robust bootstrap test)*
a) Under Assumptions 1-2, for testing for a break in the memory, the bootstrap based test, corresponding to (22), satisfies under $H_{1,T}$*,*

$$P\left(\sup_{\boldsymbol{\lambda}} F_T^*\left(\boldsymbol{\lambda}, k, 1\right) \leq x | y_1, ..., y_T\right) \xrightarrow{p} P(\sup_{\boldsymbol{\lambda}} F^d\left(\boldsymbol{\lambda}, k, 1\right) \leq x),$$

under $H_{1,T}^{d,\mu_1^0 \neq \mu_2^0}$*,*

$$P\left(\sup_{\boldsymbol{\lambda}} F_T^*\left(\boldsymbol{\lambda}, k, 1\right) \leq x | y_1, ..., y_T\right) \xrightarrow{p} \chi_1^2.$$

Further, the test is consistent.

b) Under Assumptions 1-2, for testing for a break in the mean, the bootstrap based test satisfies under H_0 *and* $H_{1,T}$*,*

$$P\left(\sup_{\boldsymbol{\lambda}} F_T^*\left(\boldsymbol{\lambda}, k, 1\right) \leq x | y_1, ..., y_T\right) \xrightarrow{p} P\left(\sup_{\boldsymbol{\lambda}} F^{\mu}\left(\boldsymbol{\lambda}, k, 1\right) \leq x\right).$$

and under $H_{1,T}^{\mu,d_1^0 > d_2^0}$*,*

$$P\left(\sup_{\boldsymbol{\lambda}} F_T^*\left(\boldsymbol{\lambda}, k, 1\right) \leq x | y_1, ..., y_T\right) \xrightarrow{p} \chi_1^2$$

and under $H_{1,T}^{\mu,d_1^0 < d_2^0}$*,*

$$P\left(\sup_{\boldsymbol{\lambda}} F_T^*\left(\boldsymbol{\lambda}, k, 1\right) \leq x | y_1, ..., y_T\right) \xrightarrow{p} \left(1 + D_{22}^{\mu}\left(\lambda_1^0, 1, d_2^0\right)\right) \chi_1^2$$

Further, the test is consistent.

$F^d\left(\boldsymbol{\lambda}, k, 1\right)$ *and* $F^{\mu}\left(\boldsymbol{\lambda}, k, 1\right)$ *are both defined in Theorem 4.*

As discussed in Section 4.1, the asymptotic distribution of the test statistic for testing for a break in the memory (mean) differs between the case when the mean (memory) changes and the case when it does not change. The bootstrap based test has to take this into account and converges in probability to the corresponding asymptotic distributions. If there is a break in the mean, $\hat{\lambda}_1$ converges to the true break fraction and due to the superconsistency, the test behaves as a Chow test (Proposition 2). If there is no break in the mean, $\hat{\lambda}_1$ has a spurious limit and the asymptotic behavior corresponds to the first term of Corollary 1. Table 7a) displays the size of this alternative bootstrap procedure. It turns out that the test is still slightly oversized when the mean is not changing (2nd column). In this case, we estimate a changing mean with a spurious break point. Thus, the generated series has a changing mean at this spurious break point and we frequently estimate a break at this point. For larger sample sizes, the size gets closer to the nominal level. The power is clearly larger than the one for an alternative conservative strategy of using always critical values from Theorem 4. This robust bootstrap test also improves

TABLE 7
Size of robust bootstrap tests.

a) Size of a bootstrap test for a break in d that is robust to a break in μ.

$d^0 \backslash\backslash \mu_2^0$	T=200 0.5	**1**	1.5	2	T=500 0.5	**1**	1.5	2	T=1000 0.5	**1**	1.5	2
0.05	5.8	9.7	5.0	4.7	7.1	8.9	7.1	7.1	7.0	7.3	6.5	5.7
0.15	11.9	9.3	8.0	7.1	5.0	8.6	5.2	5.1	4.7	5.9	6.0	5.9
0.25	5.7	8.6	9.0	6.6	6.6	6.5	4.6	4.2	6.3	6.2	6.1	4.9
0.35	5.7	6.4	6.3	6.2	4.6	8.4	4.6	4.6	7.0	6.7	7.0	3.4
0.45	5.1	6.9	5.6	4.5	6.1	6.0	6.1	4.7	5.5	5.5	5.9	3.7

b) Size of a bootstrap test for a break in μ that is robust to a break in d.

$T \backslash\backslash d_2^0$	d_1^0 =0.05 **0.05**	0.1	0.25	0.45	d_1^0 =0.25 0.05	**0.25**	0.3	0.45	d_1^0 =0.45 0.05	0.25	0.4	**0.45**
200	4.1	9.4	4.2	9.7	5.8	6.5	4.0	10.8	6.0	5.0	4.4	7.0
500	6.8	6.4	6.0	10.1	6.9	7.1	5.0	8.2	5.7	4.5	4.8	8.7
1000	6.2	5.2	6.7	10.2	4.0	4.8	8.4	10.3	4.0	5.5	8.2	6.3

steps 2a) and 2b) in the sequential procedure in Section 6. Table 7b) provides the size of the test for a break in the mean that is robust to the break in the memory. The test is still oversized when the memory is close to 0.5 since in this case, the mean is imprecisely estimated.

Finally, we analyze a bootstrap procedure for testing ℓ versus $\ell + 1$ breaks to solve the problems described in the previous section. For simplicity, we consider the case of *one* vs. *two* breaks. We apply the following residual bootstrap:

1. From the estimation under the null, described in Sections 2-3, obtain $\hat{\lambda}_1, \hat{d}_1, \hat{d}_2$, $\hat{\mu}_1, \hat{\mu}_2$ and the residuals $\hat{u}_t$.
2. Resample the residuals $\hat{u}_t$ to obtain u_t^*, and generate

$$y_t^* = \hat{\mu}_i + \Delta_t^{-\hat{d}_i} u_t^*,$$

3. From the estimation under the null and under the alternative for the new series y_t^*, obtain the test statistic from Theorem 4.
4. Repeat 2-3 B times and obtain from the empirical distribution the bootstrap critical values.

This bootstrap test is valid for similar reasons as the ones in Theorem 7 and avoids the problem of obtaining the asymptotic critical values on a case by case basis.

8. EMPIRICAL APPLICATION

In the previous sections, we have assumed that the short run dynamics structure is known. For the empirical application this assumption has to be relaxed. Since the consistency of the parametric memory estimation depends on the knowledge of this autoregressive structure, we need a preliminary estimate of the memory. For a stable fractionally integrated progress, Hualde and Robinson (2011) suggest using the following approach: First, obtain a preliminary memory estimate from a semiparametric estimation (e.g. the local Whittle estimator (Robinson, 1995)) and use this estimate to filter the series to obtain (approximately) short memory. Next, choose the orders p, q of the short memory $ARMA\left(p,q\right)$ structure by minimizing an information criterion. Finally, the parameters of the $ARFIMA\left(p,d,q\right)$ are estimated parametrically.

In our case, we need to obtain the preliminary semiparametric estimate under the alternative rather than under H_0. Thus, as in Hsu (2005) and Hassler and Meller (2009), we use a modified version of the Exact Local Whittle estimator (Shimotsu and Phillips, 2005, Shimotsu, 2010) and we further modify it by allowing also for a break in the memory. In particular, we define the periodogram and the discrete Fourier transform of a time series x_t evaluated at the fundamental frequencies as

$$I_x\left(v_j\right) = \left|\zeta_x\left(v_j\right)\right|^2$$

and

$$\zeta_x\left(v_j\right) = \left(2\pi T\right)^{-1/2}\sum_{t=1}^{T} x_t e^{itv_j}, v_j = \frac{2\pi j}{T}.$$

Given a break fraction λ, the mean estimators are

$$\mu_1\left(\lambda\right) = \frac{1}{[\lambda T]}\sum_{t=1}^{[\lambda T]} y_t \quad \text{and} \quad \mu_2\left(\lambda\right) = \frac{1}{[(1-\lambda)T]}\sum_{t=[\lambda T]+1}^{T} y_t.$$

The memory estimator is

$$\hat{d}_i\left(\lambda\right) = \arg\min_{d_i} R\left(d_i, \lambda\right),$$

where for $n_T = T^{\alpha}$, $0 < \alpha < 1$,

$$R\left(d_i, \lambda\right) = \log \hat{G}\left(d_i, \lambda\right) - 2d_i \frac{1}{n_T}\sum_{j=1}^{n_T} \log v_j$$

and

$$\hat{G}\left(d_i, \lambda\right) = \frac{1}{n_T}\sum_{j=1}^{n_T} I_{u(\lambda)}\left(v_j\right),$$

where

$$u_t\left(\lambda\right) = \left\{ \begin{array}{l} \Delta_t^{d_1}\left(y_t - \mu_1\left(\lambda\right)\right), \ t \leq [\lambda T] \\ \Delta_{t-[\lambda T]}^{d_2}\left(y_t - \mu_2\left(\lambda\right)\right), \ t > [\lambda T] \end{array} \right. . \tag{23}$$

Finally, the break fraction is estimated as

$$\hat{\lambda} = \arg\min_{\lambda} \left\{ R\left(d_1\left(\lambda\right), \lambda\right) + R\left(d_2\left(\lambda\right), \lambda\right) \right\}.$$

From Lavielle and Ludeña (2000), such a break fraction estimator should estimate the break fraction at rate n_T. The subsequent estimators of the parameters in the two regimes behave as described in Shimotsu (2006). In the following, we choose $\alpha = 0.7$.

We filter the data using the semiparametric estimates $(\tilde{d}_1, \tilde{d}_2, \tilde{\mu}_1, \tilde{\mu}_2, \tilde{\lambda}_1)$ to obtain residuals that are close to $I\left(0\right)$. Then, we determine p in the $AR(p)$ structure using the Bayesian information criterion (BIC). Afterwards, we employ the parametric testing procedure described in Section 4 and 6. The extension to more breaks is straightforward.

If the short run dynamics is also changing, yet with a stable structure, we include $\alpha_1\left(L\right)$ and $\alpha_2\left(L\right)$ in the parametric estimation. This adds another dimension to the test, along the lines of Boldea and Hall (2010). The first component (13) consists now of a two dimensional Brownian Motion. Because the pre-estimation is semiparametric, we need to assume that $\alpha\left(L\right)$ is changing at the same point as the memory and/or the mean. In the following, we assume that $\alpha\left(L\right)$ and the memory are changing at the same time.

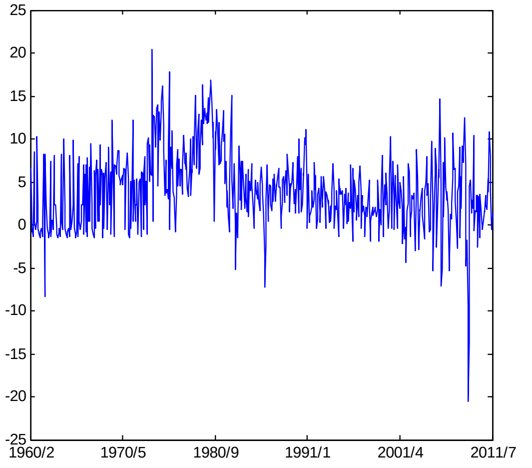
Next, we illustrate how the procedure works for a real data set. We consider the U.S. inflation time series which is already extensively analyzed in the literature. The literature is inconclusive about whether inflation is stationary, fractionally integrated or has a unit root and whether or not it has breaks in the deterministic part and/or the memory (See Martins and Rodrigues (2010) for a good summary of the results). Hsu (2005) finds two breaks in the mean in January 1973 and September 1981 when allowing for fractionally integrated errors. Hassler and Scheithauer (2011) and also Sibbertsen and Kruse (2009) find a break from a unit root to a memory smaller than 1 in the first quarter of 1982. Hassler and Meller (2009) conclude that there is one (or possibly two) break(s) in the memory. Mayoral (2011) concludes that the U.S. inflation is a fractionally integrated series with a memory around 0.6, though without testing for breaks in the memory parameter. Martins and Rodrigues (2010) find a break from a unit root to around 0.3 in July 1982, yet without taking into account potential breaks in the mean.

As in Hassler and Meller (2009), we analyze the monthly U.S. CPI data collected by the Organization for Economic Cooperation and Development (OECD). This series comprises 619 observations from January 1960 until July 2011. Inflation is computed as

$$\pi_t = 1200 \log\left(CPI_t / CPI_{t-1}\right).$$

Finally, we seasonally adjust the series by subtracting seasonal means and adding the overall mean. Figure 1 displays the seasonally adjusted inflation series.

FIG. 1 Seasonally Adjusted Monthly US Inflation



First, we apply the semiparametric procedure and find two breaks in November 1972 and in August 1981. Table 8a) displays memory and mean estimates in the regimes and the Bayesian information criterion (BIC) of AR(p) models for the filtered data in the regimes. Thus, we choose a AR(1) structure for the filtered data. Next, we apply the parametric testing procedure with an underlying ARFIMA(1,d,0) structure. In a first step, we determine sequentially the number of breaks in the memory parameter and/or the mean allowing for fractionally integrated errors under H_0 and H_1. In a second step, we identify whether the breaks are in the memory and/or the mean. Because of the size distortions mentioned in Section 5, we compare the test statistic to the bootstrap critical values. It turns out that for this data, the bootstrap critical values differ considerably from the asymptotic ones. We reject the hypothesis H_0 of no break at the 1% level. Thus, there is at least *one* break in October 1981. In the same way, we next test, whether there is an additional break in the periods before and after October 1981. Table 8b) displays the sequential tests for the number of breaks, the estimated break points, the test statistics and the bootstrap critical values. We conclude that there are two breaks, one in February 1973 and one in October 1981. The former, corresponds to the first oil crisis and the latter corresponds to the Volcker disinflation period, the end of the second oil crisis and the great moderation. The potential break in September 1990 is not found to be significant. Table 8c) summarizes the estimates of memory (with standard errors), mean and autoregressive parameter for the three regimes. At the first oil shock, the persistence increases and along with the Volcker disinflation and great moderation the persistence decreases considerably.

In the second step, we use the methodology in Proposition 2 to determine which

TABLE 8
Breaks in US Inflation Rate

a) Semiparametric pre-estimation: Memory, mean and BIC for order of AR(p)

Period	d	μ	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)
1960:02-1972:12	0.19	2.91	2.74	**2.69**	2.72	2.73	2.76	2.78
1973:01-1981:08	0.48	8.90	2.91	**2.76**	2.79	2.82	2.86	2.90
1981:09-2011:07	0.12	3.00	2.58	**2.53**	2.53	2.55	2.57	2.59

b) Sequential procedure: F-tests for breaks in both parameters.

Test	Break point	F	$CV^*_{0.95}$ ($CV^*_{0.99}$)
0 vs 1	1981:10	**55.50**	35.83 (41.60)
1 vs 2	1973:02	**25.09**	18.33 (22.84)
2 vs 3	1990:09	13.64	16.30

c) Parameter estimates in the regimes.

Period	d	μ	α
1960:02-1973:02	0.27 (0.09)	3.08	0.31
1973:03-1981:10	0.42 (0.11)	9.74	0.25
1981:11-2011:07	-0.07 (0.07)	2.98	-0.44

d) Sequential procedure: F-tests for identifying the changing parameter.

	Break in d		Break in μ	
Break point	F	$CV^*_{0.95}$	F	$CV^*_{0.95}$
1973:02	4.70	6.65	**9.82**	6.35
1981:10	**16.85**	6.58	**13.06**	5.89

parameter is the changing one for each break point. Table 8d) provides test statistics and bootstrap critical values for testing for a break in the memory (mean) under the maintained hypothesis of a break in the mean (memory). We conclude that both breaks are in the mean but only the one in October 1981 is also in the memory. Therefore, we reestimate a constant memory and the autoregressive parameter for the period 1960:01 to 1981:10 ($\hat{d} = 0.30$ (0.07) and $\hat{\alpha} = 0.29$).

Our memory estimates are considerably lower than the estimates in Martins and Rodrigues (2010), Hassler and Scheithauer (2011), Sibbertsen and Kruse (2009) and Mayoral (2011). However, these papers do not allow for breaks in the mean and, therefore, their memory estimates might be spuriously high. Hassler and Meller (2009) allow for breaks in the memory and obtain similar memory estimates as ours. However, they test for breaks in mean and memory sequentially rather than simultaneously. By testing for breaks in mean and memory simultaneously, we reduce spurious effects caused by the finite sample correlation between the respective estimates.

9. **FINAL REMARKS**

The analysis is extendable in several directions. First, we have analyzed breaks in (asymptotically) stationary time series with $0 \leq d_j^0 < 1/2$. The analysis also would hold for a memory in the interval $-1/2 < d_j^0 \leq 0$. In this case, the stronger signals come from the break in the mean rather than the break in the memory. Nevertheless, this is still too restrictive for many applications. For example, assume a series with a linear trend and with a nonstationary memory with $1/2 < d_j^0 \leq 1$ or $1 \leq d_j^0 < 3/2$,

$$y_t = \mu_j^0 + \beta_j^0 t + \Delta_t^{-d_j^0} u_t,\ t = T_{j-1}^0 + 1, ..., T_j^0.$$

In this case, we apply a first-differencing filter to the process to obtain

$$\Delta y_t = \beta_j^0 + \Delta_t^{1-d_j^0} u_t,\ t = T_{j-1}^0 + 1, ..., T_j^0.$$

The differenced process has a a changing mean and a new changing stationary memory parameter, $d_j^0 - 1 \in (-1/2, 0)$ for $1/2 < d_j^0 \leq 1$ and $d_j^0 - 1 \in (0, 1/2)$ for $1 \leq d_j^0 < 3/2$. For this interval for the memory, we have analyzed the methodology. Note that the original mean cannot be estimated and breaks in it are not identifiable and do not contribute to finding the break. Taylor *et al.* (2010) propose a test for a break in the mean that is robust for any d, including nonstationary ones. Next, if the process has a changing linear trend and a memory lying in Θ, the analysis increases by one further dimension. This analysis is beyond the scope of this paper.

In the previous analysis, we have assumed that the error follows (1). However, this so called Type II long memory process is not the only possibility of defining a long memory process. Alternatively, we could assume a Type I long memory error

$$\Delta_\infty^{-d_j^0} u_t = \sum_{j=0}^{\infty} \pi_j \left(d_j^0\right) u_{t-j}, 0 \leq d_j^0 < 1/2.$$

The estimation of the memory and of the short run dynamics is unaffected. The mean estimation, on the other hand, has an additional term that is similar to (6). In the tests, the variance is increased in a similar way as in Theorem 6. This increased variance would have to be taken into account. Further, since the mean is less precisely estimated, the resulting local power would be lower.

Finally, we have assumed one of two situations. Breaks are exclusively in one parameter or always simultaneously in both parameters. Nevertheless, the proposed procedure also works if the breaks are not simultaneous. Assume the true process has k_1 breaks in the memory and k_2 breaks in the mean at potentially different break points. Using the sequential testing in the lines of Bai and Perron (1998), we first detect $k = k_1 + k_2$ breaks. Next, using the sequential procedure in Section 6, we obtain for each of the k breaks, whether it is in the memory, in the mean or in both parameters.

REFERENCES

[1] Bai, J. & Perron, P., 1998. "Estimating and Testing Linear Models with Multiple Structural Changes," Econometrica, Econometric Society, vol. 66(1), pages 47-78, January.

[2] Boldea, O. & Hall, A. R., 2010. "Estimation and inference in unstable nonlinear least squares models," MPRA Paper 23150, University Library of Munich, Germany.

[3] Busetti, F. & Taylor, A. M. R., 2004. "Tests of stationarity against a change in persistence," Journal of Econometrics, Elsevier, vol. 123(1), pages 33-66, November.

[4] Chung, C.F. & Baillie, R. T., 1993. "Small Sample Bias in Conditional Sum-of-Squares Estimators of Fractionally Integrated ARMA Models," Empirical Economics, Springer, vol. 18(4), 791-806

[5] Dolado, J.J. & Gonzalo, J. & Mayoral, L. "Simple Wald Tests of the Fractional Integration Parameter: An Overview of New Results", The Methodology and Practice of Econometrics (A Festschrift in Honour of David Henry) (eds., J. Castle and N. Shepard) (OUP), 2009.

[6] Johansen, S & Nielsen, M.Ø., 2010. "Likelihood inference for a nonstationary fractional autoregressive model," Working Papers 1172, Queen's University, Department of Economics.

[7] Giacomini, R. & Politis, D. & White, H., 2007. "A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators," Working Paper

[8] Gil-Alana, L. A., 2008. "Fractional integration and structural breaks at unknown periods of time," Journal of Time Series Analysis, Blackwell Publishing, vol. 29(1), pages 163-185, 01.

[9] Gine, E. & Zinn, J., 1990., "Bootstrapping General Empirical Measures," Annals of Probability, Volume 18, Number 2, 851-869.

[10] Giraitis, L. & Kokoszka, P. & Leipus, R. & Teyssiere, G., 2003. "Rescaled variance and related tests for long memory in volatility and levels," Journal of Econometrics, Elsevier, vol. 112(2), pages 265-294, February.

[11] Hassler, U. & Meller, B., 2009. "Detecting a Change in Inflation Persistence in the Presence of Long Memory: Evidence from a New Approach". Available at SSRN: http://ssrn.com/abstract=1349129

[12] Hassler, U. & Scheithauer, J., 2011. "Detecting changes from short to long memory," Statistical Papers, Volume 52, Number 4, 847-870

[13] Hosoya, Y., 2005. "Fractional invariance principle," Journal of Time Series Analysis, Blackwell Publishing, vol.26, no.3, pp.463-486

[14] Hualde, J. & Robinson, P.M., 2010. "Gaussian Pseudo-Maximum Likelihood Estimation of Fractional Time Series Models," Working paper, Universidad de Navarra.

[15] Hsu, C.C., 2005. "Long memory or structural changes: An empirical examination on inflation rates, " Economics Letters, Volume 88, Issue 2, Pages 289-294

[16] Kapetanios, G., 2010. "A Generalization of a Sieve Bootstrap Invariance Principle to Long Memory Processes, " Quantitative and Qualitative Analysis in Social Sciences. Volume 4, Issue 1, 19-40

[17] Kim, J.-Y.& Belaire-Franch, J. & Amador, R.B., 2002. "Corrigendum to Detection of change in persistence of a linear time series", Journal of Econometrics 109, 389-392.

[18] Kuan, C.M. & Hsu, C.C., 1998. "Change-Point Estimation of Fractionally Integrated Processes," Journal of Time Series Analysis, Blackwell Publishing, Volume 19, Number 6, pp. 693-708(16)

[19] Lavielle, M. & Ludeña C., 2000. "The multiple change-points problem for the spectral distribution," Bernoulli. Volume 6, Number 5 (2000), 845-869.

[20] Lavielle, M. & Moulines E., 2000. "Least Squares estimation of an unknown number of shifts in a time series, " Journal of Time Series Analysis, Blackwell Publishing, vol. 21(1), pp. 33-59

[21] Lazarova, S., 2005. "Testing for structural change in regression with long memory processes," Journal of Econometrics, 129, issue 1-2, p. 329-372.

[22] Lobato, I. N. & Velasco, C., 2007. "Efficient Wald Tests for Fractional Unit Roots," Econometrica, Econometric Society, vol. 75(2), 575-589, 03.

[23] Marinucci, D. & Robinson, P.M., 1999. "Alternative forms of fractional Brownian motion," Journal of Statistical Planning and Inference, Elsevier, Vol. 80 (1), 111-122(12)

[24] Martins, L.F. & Rodrigues, P.M.M., 2010. "Testing for Persistence Change in Fractionally Integrated Models: An Application to World Inflation Rates," Working Papers w201030, Banco de Portugal, Economics and Research Department.

[25] Mayoral, L., 2011. "Testing for Fractional Integration Versus Short Memory with Structural Breaks," Oxford Bulletin of Economics and Statistics, Blackwell Publishing Ltd

[26] McCloskey, A., 2009. “Semiparametric Testing for Changes in Memory of Otherwise Stationary Time Series”, mimeo. March 2009.

[27] Perron, P. & Zhu, X., 2005. "Structural breaks with deterministic and stochastic trends," Journal of Econometrics, Elsevier, vol. 129(1-2), pages 65-119.

[28] Rachinger, H., 2011. "Supplemental Appendix to Multiple Breaks in Long Memory Time Series," Available at

https://sites.google.com/site/heikorachinger/Rachinger2011b.pdf

[29] Robinson, P. M., 1995. "Gaussian Semiparametric Estimation of Long Range Dependence," Annals of Statistics, 23 (5), 1630-1661

[30] Shimotsu, K., 2006. "Simple (but effective) tests of long memory versus structural breaks," Working Papers 1101, Queen’s University, Department of Economics.

[31] Shimotsu, K., 2010. "Exact Local Whittle Estimation Of Fractional Integration With Unknown Mean And Time Trend," Econometric Theory, Cambridge University Press, vol. 26(02), pages 501-540

[32] Katsumi S. & Phillips, P.C.B., 2005. "Exact local Whittle estimation of fractional integration," Annals of Statistics, Volume 33 (4), pages 1890-1933.

[33] Sibbertsen, P. & Kruse, R., 2009. "Testing for a break in persistence under long-range dependencies," Journal of Time Series Analysis, Wiley Blackwell, vol. 30(3), pages 263-285, 05.

[34] Taylor, R. & Harvey, D. & Leybourne, S., 2010. "Robust Methods for Detecting Multiple Level Breaks in Autocorrelated Time," Journal of Econometrics.157, pages 342-358.

[35] Wright, J.H., 1995. "Stochastic orders of magnitude associated with two-stage estimators of fractional arima systems." Journal of Time Series Analysis 16, 119–126.

[36] Wright, J. H., 1998. "Testing for a Structural Break at Unknown Date with Long-memory Disturbances," Journal of Time Series Analysis, 19, 369–376

APPENDIX A: LEMMATA AND PROPOSITIONS

A.1. Lemmata

LEMMA 1. *Under Assumptions 1-3, uniformly in $\theta \times r \in \Theta \times [0,1]$ and in s for $s - \lambda_{i-1}^0 = O\left(T^{-1}\right)$,*

$$a)\ T^{-\delta_i} \sum_{[sT]+1}^{[rT]} d_t^2\left(\theta\right) = O_p(1)$$

$$b)\ T^{-\delta_i} \sum_{[sT]+1}^{[rT]} u_t d_t\left(\theta\right) = o_p(1)$$

Proof. We have to show uniform convergence of $\sum_{t=[sT]+1}^{[rT]} d_t^2$ and $\sum_{t=[sT]+1}^{[rT]} d_t u_t$ for $\left(s - \lambda_{i-1}^0\right) = O\left(T^{-1}\right)$. The proofs of tightness use among other Lemma 15 and 16 of Johansen and Nielsen (2010). For Part a), we provide a sketch of the proof in Rachinger (2011). ∎

LEMMA 2. *If $\lambda_i^{(1)} < \lambda_i^0$, for some i then*

$$\begin{aligned}
(i) \sup{}_{\lambda_i^{(1)} < \lambda_i^0} T^{-\delta_i} \sum_{t=1}^{T} d_t u_t &= o_p\left(1\right) \\
(ii) \liminf P\left[T^{-\delta_i} \sum_{t=1}^{T} d_t^2 > C\right] &> \epsilon, \text{ for some } C > 0,\ \epsilon > 0.
\end{aligned}$$

For a break at T_i^0 in the memory and the mean or only in the memory : $\delta_i = 1$ *and for a break only in the mean*: $\delta_i = 1 - 2d_i^0$.

Proof. First, denote for $\lambda_{i-1}T < t \leq \lambda_i T$

$$d_t\left(\lambda_{i-1}, \theta_i\right) = \hat{u}_t\left(\lambda_{i-1}, \theta_i\right) - u_t. \tag{24}$$

We have to show that for any break fraction smaller than the true one, $\lambda_i^{(1)} < \lambda_i^0$, the term $T^{-\delta_i} \sum_{t=1}^{T} u_t d_t$ vanishes and $T^{-\delta_i} \sum_{t=1}^{T} d_t^2$ is of order $O_p^+\left(1\right)$.

ii) Assume m breaks and consider the break in $\lambda_i^0 T$ in (d, μ) or d. For $\lambda_i^{(1)} < \lambda_i^0$, we know from Lemma 1 that

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^{T} d_t^2 &\geq \frac{1}{T} \sum_{t=\lambda_{i-1}^{(1)}T+1}^{\lambda_i^0 T} d_t^2\left(\lambda_{i-1}, \theta_i\right) + \frac{1}{T} \sum_{t=\lambda_i^0 T+1}^{\lambda_i^{(1)}T+1} d_t^2(\lambda_{i-1}, \theta_i) \\
&\xrightarrow{p} \left(\lambda_i^0 - \lambda_{i-1}^0\right) \sigma^2 \sum_{j=1}^{\infty} \pi_j^2 (d_i - d_i^0) + \left(\lambda_i^{(1)} - \lambda_i^0\right) \sigma^2 \sum_{j=1}^{\infty} \pi_j^2 (d_i - d_{i+1}^0)
\end{aligned}$$

Similarly as in Boldea and Hall (2010), we can choose an η small enough so that the previous term bounds

$$
\begin{aligned}
&\eta\sigma^2 \inf_{d_i} \left[\sum_{j=1}^{\infty} \pi_j^2 (d_i - d_i^0) + \sum_{j=1}^{\infty} \pi_j^2 (d_i - d_{i+1}^0) \right] \\
\geq \quad &\eta\sigma^2 \left[(d_i - d_i^0)^2 \sum_{j=1}^{\infty} \pi_j^2(0) + (d_i - d_{i+1}^0)^2 \sum_{j=1}^{\infty} \pi_j^2(0) \right] \\
> \quad &\eta\sigma^2 \left(\frac{\pi^2}{6} - 1 \right) \left[(d_i - d_i^0)^2 + (d_i - d_{i+1}^0)^2 \right] > 0
\end{aligned}
$$

uniformly in d_i.

Next, we consider the consistency of the break fraction estimator, when there is only a break in the mean. For $d_i^0 > 0$, d_i and d_{i+1} converge at rate $T^{1/2}$ to d_i^0 and terms including $\left(d_j - d_i^0\right)$ vanish. From the proof of Lemma 1,

$$
\begin{aligned}
T^{2d_i - 1} \sum_{t=1}^{T} d_t^2 &\geq T^{2d_i - 1} \sum_{t=\lambda_{i-1}^{(1)} T + 1}^{\lambda_i^0 T} d_t^2 \left(\lambda_{i-1}^{(1)}, \theta_i \right) + T^{2d_i - 1} \sum_{t=\lambda_i^0 T + 1}^{\lambda_i^{(1)} T + 1} d_t^2 \left(\lambda_{i-1}^{(1)}, \theta_i \right) \\
&\geq T^{2d_i - 1} \sum_{t=\lambda_{i-1}^{(1)} T + 1}^{T_i^0} \left[\left(\mu_i^0 - \mu_i \right) \Delta_{t-\lambda_{i-1}^{(1)} T}^{d_i} 1 \right]^2 \\
&+ T^{2d_i - 1} \sum_{t=T_i^0 + 1}^{\lambda_i^{(1)} T} \left[\left(\mu_{i+1}^0 - \mu_i \right) \Delta_{t-T_i^0}^{d_i} 1 + \left(\mu_i^0 - \mu_i \right) \left(\Delta_{t-\lambda_i^{(1)} T}^{d_i} 1 - \Delta_{t-T_i^0}^{d_{i+1}} 1 \right) \right]^2 .
\end{aligned}
$$

First, both terms have a nonnegative limit. The first term's limit equals zero only if $\left(\mu_i^0 - \mu_i\right) = o_p(1)$. But in this case, the second term's limit is larger than zero. Therefore, uniformly in μ_i and d_i for $\left(d_i - d_i^0\right) = O_p(T^{-1/2})$, the term is positive. For the contradiction established for the break in T_i^0, the less favorable case is the one where all other breaks $j \neq i$ are consistently estimated at the rate established in Theorem 2. Therefore, it suffices to consider this case.

i) follows from Lemma 1. ∎

Lemma 3 states some properties for the regressor function and its derivative that are needed in the proofs. In Boldea and Hall (2010), they are assumed in their Assumptions 2-4. In our context, they are a consequence of Assumption 1 and 2.

LEMMA 3. *Define* $F_t(\theta) = \frac{\partial f_t(\theta)}{\partial \theta}$*, a px1 vector, a function of* θ_i *for* $t \epsilon \left[T_{i-1} + 1, T_i\right]$ *and* $F_{k,t}(\theta)$*,* $k = d, \mu$ *the derivative with respect to d and* μ *respectively. Further, define* $\bar{T}\left(d_i^0\right) = diag\left(T^{-1/2}, T^{d_i^0 - 1/2}\right)$

a) *Given the superconsistent rate of convergence of the break fractions,* $S_{i,T}\left(\lambda_{i-1}, \lambda_i, \theta_i\right)$ *defined in 5, appropriately standardized converges to a limit that is minimized in* $d_i = d_i^0$ *and* $\mu_i = \mu_i^0$*.*

b) *Evaluated at the true* θ_i^0 *and the true break fractions,*

$$\begin{aligned} D_{T,i}\left(\theta_i^0\right) &= \bar{T}\left(d_i^0\right)\sum_{t=T_{i-1}^0+1}^{T_i^0} F_t\left(\lambda_{i-1}^0,\theta_i^0\right)F_t\left(\lambda_{i-1}^0,\theta_i^0\right)'\bar{T}\left(d_i^0\right) \\ &\xrightarrow{p} \sigma^2 D_i^0\left(\lambda_{i-1}^0,\lambda_i^0,\theta_i^0\right) \end{aligned}$$

where

$$D_i^0\left(\lambda_{i-1}^0,\lambda_i^0,\theta_i^0\right)=\begin{pmatrix} \frac{\pi^2}{6}\left(\lambda_i^0-\lambda_{i-1}^0\right) & 0 \\ 0 & \frac{\left(\lambda_i^0-\lambda_{i-1}^0\right)^{1-2d_i^0}}{\left(1-2d_i^0\right)\Gamma^2\left(1-d_i^0\right)} \end{pmatrix}.$$

c) *Uniformly in* (s,r,θ) *for* $\left(s-\lambda_{i-1}^0\right)=O_p\left(T^{-1}\right)$ *and* $r>\lambda_i^0$*,*

$$D_{i,T}\left(\theta_i\right)=\bar{T}\left(d_i\right)\sum_{t=[sT]+1}^{[rT]}F_t\left(s,\theta\right)F_t\left(s,\theta\right)'\bar{T}\left(d_i\right)\xrightarrow{p}\sigma^2 D_i\left(s,r,\theta\right)$$

where

$$D_i\left(r,\theta\right)=\begin{pmatrix} \left(r-\lambda_i^0\right)\sigma^2\sum\limits_{j=0}^{\infty}\dot{\pi}_j^2(d-d_{i+1}^0)+\left(\lambda_i^0-\lambda_{i-1}^0\right)\sigma^2\sum\limits_{j=0}^{\infty}\dot{\pi}_j^2(d-d_i^0) & 0 \\ 0 & \frac{\left(r-\lambda_{i-1}^0\right)^{1-2d}}{(1-2d)\Gamma^2(1-d)} \end{pmatrix}.$$

d) *Evaluated at the true* d_i^0 *and the true break fractions*

$$A_i\left(\theta_i^0\right)=Var\left[diag\left(T^{-1/2},T^{d_i^0-1/2}\right)\sum\nolimits_{t\in I_i^0}u_t\left(\lambda_{i-1}^0,\theta_i^0\right)F_t\left(\lambda_{i-1}^0,\theta_i^0\right)\right]\xrightarrow{p}A\left(\lambda_{i-1}^0,\lambda_i^0,d_i^0\right)$$

where

$$A\left(\lambda_{i-1}^0,\lambda_i^0,d_i^0\right)=\begin{pmatrix} \sigma^4\frac{\pi^2}{6}\left(\lambda_i^0-\lambda_{i-1}^0\right) & 0 \\ 0 & \sigma^2\left(\frac{\left(\lambda_i^0-\lambda_{i-1}^0\right)^{1-2d_i^0}}{\Gamma^2(1-d_i^0)(1-2d_i^0)}+A_i^{\mu}\left(\lambda_{i-1}^0,\lambda_i^0,d_i^0\right)\right) \end{pmatrix}$$

with $A_i^{\mu}\left(\lambda_{i-1}^0,\lambda_i^0,d_i^0\right)$ *defined in (7). Because of the term* $A_i^{\mu}\left(\lambda_{i-1}^0,\lambda_i^0,d_i^0\right)$*,*

$$A\left(\lambda_i^0,\lambda_{i-1}^0,\theta_i^0\right)\neq D\left(\lambda_i^0,\lambda_{i-1}^0,\theta_i^0\right).$$

Proof. **Part a)** Write

$$\begin{aligned} S_{i,T}\left(\lambda_{i-1},\lambda_i,\theta_i\right) &= \sum_{t=T_{i-1}+1}^{T_i}\left(\Delta_{t-T_{i-1}}^{d_i}\Delta_t^{-d_i}u_t\right)^2+\sum_{t=T_{i-1}+1}^{T_i}\left(\left(\mu_i-\mu_i^0\right)\Delta_{t-T_{i-1}}^{d_i}1\right)^2 \\ &\quad -2\sum_{t=T_{i-1}+1}^{T_i}\Delta_{t-T_{i-1}}^{d_i}\Delta_t^{-d_i}u_t\left(\mu_i-\mu_i^0\right)\Delta_{t-T_{i-1}}^{d_i}1. \end{aligned}$$

For the first term uniformly in d_i and μ_i,

$$\frac{1}{T}\sum_{t=T_{i-1}+1}^{T_i}\left(\Delta_{t-T_{i-1}}^{d_i}\Delta_t^{-d_i}u_t\right)^2\xrightarrow{p}\left(\lambda_i^0-\lambda_{i-1}^0\right)\sum_{j=0}^{\infty}\pi_j^2(d_i-d_i^0),$$

a limit that has a unique minimum at d_i^0. The convergence follows from a law of large numbers and the last expression follows from (19) in Lobato and Velasco (2007). Uniformity, follows from a similar argument as the one in the proof of Lemma 1. For the second term uniformly in d_i and μ_i,

$$T^{2d_i-1}\sum_{t=T_{i-1}+1}^{T_i}\left(\left(\mu_i-\mu_i^0\right)\Delta_{t-T_{i-1}}^{d_i}1\right)^2 \to \left(\lambda_i^0-\lambda_{i-1}^0\right)^{1-2d_i^0}\frac{\left(\mu_i^0-\mu_i\right)^2}{\left(1-2d_i^0\right)\Gamma^2\left(1-d_i^0\right)},$$

a limit that has a unique minimum at $\mu_i=\mu_i^0$. Uniformity follows from the deterministic character. Finally, the third term multiplied by T^{d_i-1} is uniformly in d_i and μ_i of order $o_p(1)$.

Part b)

The derivative evaluated at true break points and true parameters, $F_t\left(\lambda_{i-1}^0,\theta_i^0\right)$, for $t=T_{i-1}^0+1,...,T_i^0$,

$$F_t\left(\lambda_{i-1}^0,\theta_i^0\right)=\begin{pmatrix} +\sum_{j=1}^{t-T_{i-1}^0-1} j^{-1}u_{t-j}+\dot{\Delta}_{t-T_{i-1}^0}^{d_i^0}\sum_{j=t-T_{i-1}^0}^{t-1}\pi_j\left(-d_i^0\right)u_{t-j} \\ -\Delta_{t-T_{i-1}^0}^{d_i}1 \end{pmatrix}. \tag{25}$$

First, the (1,1) element of $D_T\left(\theta_i^0\right)$ converges in mean square to $\left(\lambda_i^0-\lambda_{i-1}^0\right)\frac{\pi^2}{6}$ because the terms coming from the second term in $F_t\left(\lambda_{i-1}^0,\theta_i^0\right)$ are negligible. The (2,2) element of $D_T\left(\theta_i^0\right)$ converges to $\frac{\left(\lambda_i^0-\lambda_{i-1}^0\right)^{1-2d_i^0}}{\Gamma^2\left(1-d_i^0\right)\left(1-2d_i^0\right)}$. Finally, the (1,2) element is of smaller order.

Part c) Note that for a break fraction λ_{i-1}, the residuals for $t=T_{i-1}+1,...,T_i^0$ are

$$\hat{u}_t^{(i)}\left(\lambda_{i-1},\theta_i\right)=\Delta_{t-T_{i-1}}^{d_i}\left(\mu_i^0-\mu_i\right)+\Delta_{t-T_{i-1}}^{d_i-d_i^0}u_t+\Delta_{t-T_{i-1}}^{d_i}\sum_{j=t-T_{i-1}}^{t-1}\pi_j\left(-d_i^0\right)u_{t-j}.$$

The difficulty arises from showing that the last term of the first is asymptotically negligible. Similarly, the derivatives $F_t(\lambda,\theta)$ have a similar additional term. For the (2,2) element of $D_{i,T}\left(\theta,\theta_i^0\right)$, *fidi* convergence corresponds to the one in part b) since $\left(s-\lambda_{i-1}^0\right)=O_p\left(T^{-1}\right)$. Uniformity follows directly from the fact that the term is deterministic. For $D_{i,T}\left(\theta,\theta_i^0\right)_{(1,1)}$, we use that terms containing $\left(\mu_i^0-\mu_i\right)$ are of order $T^{1-2d_i^0}$. For uniformity, in (s,r,θ), the tightness of $D_T\left(\theta,\theta_i^0\right)$ can be proved using Johansen and Nielsen (2010).

Part d) The $(1x1)$ element of $A_i\left(\theta_i^0\right)$ is straightforward. For the $(2x2)$ element, we separate the second term into two uncorrelated terms

$$\Delta_{t-\lambda_{i-1}^0T}^{d_i}\Delta_t^{-d_i^0}u_t=\Delta_{t-\lambda_{i-1}^0T}^{d_i}\Delta_{t-\lambda_{i-1}^0T}^{-d_i^0}u_t+\Delta_{t-\lambda_{i-1}^0T}^{d_i}\sum_{k=t-T_{i-1}^0}^{t-1}\pi_k\left(-d_i^0\right)u_{t-k}.$$

The first term leads to a variance component of $\frac{\left(\lambda_i^0-\lambda_{i-1}^0\right)^{1-2d_i^0}}{\Gamma\left(1-d_i^0\right)\left(1-2d_i^0\right)}$. The one corresponding to the second term,

$$Var\left(T^{d_i-1/2}\sum_{t=T_{i-1}^0+1}^{T_i^0}\Delta_{t-T_{i-1}^0}^{d_i}1\Delta_{t-T_{i-1}^0}^{d_i}\sum_{k=t-T_{i-1}^0}^{t-1}\pi_k\left(-d_i^0\right)u_{t-k}\right)$$

$$= T^{2d_i-1}E\left[\sum_{k=1}^{T_{i-1}^0}\left(\sum_{t=1}^{T_i^0-T_{i-1}^0}\pi_{t-1}\left(d_i-1\right)\Delta_t^{d_i}\pi_{T_{i-1}^0+t-k}\left(-d_i^0\right)\right)u_k\right]^2,$$

converges to $\sigma^2 A_i^{\mu}\left(\lambda_{i-1}^0,\lambda_i^0,\theta_i^0\right)$. Combining the two terms leads to the result. ∎

Lemma 4 discusses the estimators for the partitions (T_1,T_2,T_3), $\left(T_1,T_2^0,T_3\right)$ and $\left(T_1,T_2,\,T_2^0,T_3\right)$.

Lemma 4. *(Behavior of estimators)*

a) For the estimator $\left(\theta_2^{***},\theta_2^{\delta},\theta_3^{***}\right)$ for $\left(T_1,T_2,T_2^0,T_3\right)$

$$\begin{aligned}\left(d_2^{***}-d_2^0,d_2^{\delta}-d_2^0,d_3^{***}-d_3^0\right) &= \left(O_p(T^{-1/2}),O_p(\Delta_2^{-1/2}),O_p(T^{-1/2})\right)\\ \left(\mu_2^{***}-\mu_2^0,\mu_2^{\delta}-\mu_2^0,\mu_3^{***}-\mu_3^0\right) &= \left(O_p(T^{-1/2+d_2^0}),O_p(\Delta_2^{-1/2+d_2^0}),O_p(T^{-1/2+d_3^0})\right)\end{aligned}$$

b) For the estimator $(\theta_2^{*},\theta_3^{**})$ for (T_1,T_2,T_3)

$$\begin{aligned}\left(d_2^{*}-d_2^0,d_3^{**}-d_3^0\right) &= \left(O_p(T^{-1/2}),O_p(T^{-1/2})\right)\\ \left(\mu_2^{*}-\mu_2^0,\mu_3^{**}-\mu_3^0\right) &= \left(O_p(T^{-1/2+d_2^0}),O_p(T^{-1/2+d_3^0})\right)\end{aligned}$$

c) For the estimator $(\theta_2^{**},\theta_3^{*})$ for $\left(T_1,T_2^0,T_3\right)$

$$\begin{aligned}\left(d_2^{**}-d_2^0,d_3^{*}-d_3^0\right) &= \left(O_p(T^{-1/2}),O_p(T^{-1/2})\right)\\ \left(\mu_2^{**}-\mu_2^0,\mu_3^{*}-\mu_3^0\right) &= \left(O_p(T^{-1/2+d_2^0}),O_p(T^{-1/2+d_3^0})\right)\end{aligned}$$

Lemmata 5 and 6 are needed for the proof of Theorem 2. We analyze the terms $\sum d_t^2,\sum d_t u_t$ multiplied by Δ_2^{-1} in the case of breaks in memory and mean or only in memory and by $\Delta_2^{-1+2d_2^{\delta}}$ in the case of a break only in the mean respectively. Both Lemmata use Lemma 4. The proofs of tightness are similar to the ones of Lemma 1 and use among others Lemma 15 and 16 of Johansen and Nielsen (2010). Further, we consider $T_2 < T_2^0$ and $\left(s-\lambda_1^0\right)=O_p\left(T^{-1}\right)$.

Lemma 5. *(Break in memory or in memory and mean.)*
a) *Behavior of* $\sum d_t^2$. *For* $r=\lambda_2<\lambda_2^0$

$$\Delta_2^{-1}\sum_{t=[sT]+1}^{[rT]}d_t^2=o_p\left(1\right),\Delta_2^{-1}\sum_{t=\lambda_2^0T+1}^{\lambda_3T}d_t^2=o_p\left(1\right).$$

$$\Delta_2^{-1}\sum_{t=[rT]+1}^{\lambda_2^0T}d_t^2\xrightarrow{p}\sum_{j=1}^{\infty}\pi_j^2\left(d_2^{\delta}-d_2^0\right)=O_p\left(1\right),$$

b) *Behavior of* $\sum d_t u_t$.

$$\Delta_2^{-1} \sum_{t=[sT]+1}^{[rT]} d_t u_t, \sum_{t=[rT]+1}^{\lambda_2^0 T} d_t u_t \text{ and } \Delta_2^{-1} \sum_{t=\lambda_2^0 T+1}^{\lambda_3 T} d_t u_t = o_p(1)$$

Proof. We use Cauchy Schwarz for the first and third in Part b). In particular

$$\left[\Delta_2^{-1} \sum_{t=[sT]+1}^{[rT]} d_t u_t\right]^2 \leq \Delta_2^{-1} \sum_{t=[sT]+1}^{[rT]} d_t^2 \Delta_2^{-1} \sum_{t=\lambda_1^0 T+1}^{[rT]} u_t^2$$

where the first term converges to zero from Part a). The proofs are similar to the one of Lemma 1 with the difference that the considered interval is constant rather than proportional to T. In particular, some tedious analysis shows that the terms converge uniformly. ∎

LEMMA 6. *For* $(d_2^\delta - d^0) = O_p(\Delta_2^{-1/2})$.
a) *Behavior of* $\sum d_t^2$

$$\Delta_2^{-1+2d_2^\delta} \sum_{t=[sT]+1}^{[rT]} d_t^2 = o_p(1), \Delta_2^{-1+2d_2^\delta} \sum_{t=\lambda_2^0 T+1}^{\lambda_3 T} d_t^2 = o_p(1)$$

$$\Delta_1^{-1+2d_2^\delta} \sum_{t=[rT]+1}^{\lambda_2^0 T} d_t^2 \xrightarrow{p} \frac{\left(\mu_2^0 - \mu_2^\delta\right)^2}{\Gamma^2(1-d^0)(1-2d^0)}$$

b) *Behavior of* $\sum d_t u_t$

$$\begin{aligned} \Delta_2^{-1+2d_2^\delta} \sum_{t=[sT]+1}^{[rT]} d_t u_t &= o_p(1), \Delta_2^{-1+2d_2^\delta} \sum_{t=[rT]+1}^{\lambda_2^0 T} d_t u_t = o_p(1) \text{ and} \\ \Delta_2^{-1+2d_2^\delta} \sum_{t=\lambda_2^0 T+1}^{\lambda_3 T} d_t u_t &= o_p(1) \end{aligned}$$

Proof. The terms including μ are deterministic, for the terms including d we can show that they converge uniformly at a faster rate and are, therefore, negligible at the present rate. Part b) follows from similar argument as the one in Part a). In addition we need also a uniform argument for the terms including μ. ∎

A.2. Propositions

Proposition 5 derives the asymptotic distribution of the estimators defined below (9) under the local alternative $H_{1,T}$.

PROPOSITION 5. *Under Assumptions 1-2, for $i = 1, ..., k+1$*

$$a) \qquad \bar{T}\left(d_1^0\right)\left(\hat{\theta}_{1,i} - \theta_1^0\right) \Longrightarrow \begin{pmatrix} \left(\lambda_i\right)^{-1}\left(\frac{\sqrt{6}}{\pi}B^h\left(\lambda_i\right)\right) \\ \frac{\sigma}{\lambda_i^{1-2d_1^0}}\left(\Gamma\left(1-d_1^0\right)\sqrt{1-2d_1^0}\tilde{W}^h\left(\lambda_i\right)\right) \end{pmatrix}.$$

$$b) \qquad \bar{T}\left(d_1^0\right)\left(\hat{\theta}_i - \theta_1^0\right) \Longrightarrow$$

$$\begin{pmatrix} \left(\lambda_i - \lambda_{i-1}\right)^{-1}\left(\frac{\sqrt{6}}{\pi}\left[B^h\left(\lambda_i\right) - B^h\left(\lambda_{i-1}\right)\right]\right) \\ \frac{1}{\lambda_i^{1-2d_1^0} - \lambda_{i-1}^{1-2d_1^0}}\left(\sigma\Gamma\left(1-d_1^0\right)\sqrt{1-2d_1^0}\left[\tilde{W}^h\left(\lambda_i\right) - \tilde{W}^h\left(\lambda_{i-1}\right)\right]\right) \end{pmatrix},$$

where $B^h(\lambda_i)$ and $\tilde{W}^h(\lambda)$ are defined in (13) and (14) respectively. θ_i and θ_j are asymptotically uncorrelated.

Proof. **Part a)** The consistency follows from combining Lemma 3a) and Robinson and Hualde (2010). For the asymptotic distribution, we analyze its denominator and numerator. For the denominator, we obtain uniformly,

$$\bar{T}\left(d_1^0\right)\sum_{1,i} F_t\left(0, \theta_{1,i}\right) F_t'\left(0, \theta_{1,i}\right)\bar{T}\left(d_1^0\right) \xrightarrow{p} \begin{pmatrix} \lambda_i\frac{\pi^2}{6} & 0 \\ 0 & \frac{\lambda_i^{1-2d_1^0}}{\Gamma^2\left(1-d_1^0\right)\left(1-2d_1^0\right)} \end{pmatrix}$$

and for the numerator, we obtain

$$\bar{T}\left(d_1^0\right)\sum_{1,i} u_t F_t\left(0, \theta_{1,i}\right) \Longrightarrow \begin{pmatrix} \frac{\pi}{\sqrt{6}}B^h\left(\lambda_i\right) \\ \frac{1}{\Gamma\left(1-d_1^0\right)\sqrt{1-2d_1^0}}\tilde{W}^h\left(\lambda_i\right) \end{pmatrix}$$

where the weak convergence to Brownian and fractional Brownian Motion follows from a FCLT and Marinucci and Robinson (1999) respectively. The fractional Brownian Motion $\tilde{W}_{1/2-d_1^0}(\lambda_i)$ has the same marginal distribution as the standard one $W_{1/2-d_1^0}(\lambda_i) = \int_0^{\lambda_i}(\lambda_i - r)\, dB(r)$. Because of the opposite order of summing the error terms, its covariance is (12) rather than the usual one,

$$\begin{aligned} Cov\left(W_{1/2-d_1^0}\left(\lambda_i\right), W_{1/2-d_1^0}\left(\lambda_{i-1}\right)\right) &= \frac{\lambda_i^{1-2d_1^0} + \lambda_{i-1}^{1-2d_1^0}}{\Gamma\left(1-d_1^0\right)\left(1-2d_1^0\right)} \\ &\quad - E\left[W_{1/2-d_1^0}\left(\lambda_i\right) - W_{1/2-d_1^0}\left(\lambda_{i-1}\right)\right]^2. \end{aligned}$$

In consequence, $\tilde{W}_{1/2-d_1^0}(.)$ has independent increments. The local drift of the memory estimator

$$\frac{\frac{1}{T}\sum_{j=1}^{i+1}\sum_{t=T_{j-1}+1}^{T_j} h_d\left(\frac{t}{T}\right)\left(\sum_{j=1}^{t-1}\dot{\pi}_j(0)u_{t-j}\right)^2}{T^{-1}\sum_{j=1}^{i+1}\sum_{t=T_{j-1}+1}^{T_j}\left(\sum_{j=1}^{t-1}\dot{\pi}_j(0)u_{t-j}\right)^2} \xrightarrow{p} \frac{\sigma^2}{\lambda_i}\int_0^{\lambda_i} h_d(u)\, du.$$

and the one of the mean

$$\frac{T^{2d_1^0-1}\sum_{j=1}^{i+1}\sum_{t=T_{j-1}+1}^{T_j}\left(\Delta_t^{d_{1,i}}\right)^2 h_\mu\left(\frac{t}{T}\right)}{T^{2d-1}\sum_{j=1}^{i+1}\sum_{t=T_{j-1}+1}^{T_j}\left(\Delta_t^{d_{1,i}}\right)^2} \rightarrow \frac{1}{\lambda_i^{1-2d_1^0}}\int_0^{\lambda_i} u^{-2d_1^0} h_\mu(u)\, du,$$

where we use that $(\Delta_t^{d_i} 1) \simeq (t-1)^{-d_i}$ and that $h_\mu(.)$ is a bounded variation function.

Part b) The proofs follow similar lines as the one of Part a). The variance of the estimator μ_i is $\frac{\lambda_i^{1-2d_1^0} - \lambda_{i-1}^{1-2d_1^0}}{\Gamma^2(1-d_1^0)(1-2d_1^0)}$. Further, the covariance of the two estimators μ_i and μ_j for $i<j$ is

$$Cov\left(T^{d_1^0-1/2}\left(\mu_i-\mu_i^0\right), T^{d_1^0-1/2}\left(\mu_j-\mu_j^0\right)\right) = 0$$

since unlike Lemma 3,

$$Cov\left(T^{1/2-d_1^0}\sum_{t=T_{i-1}+1}^{T_i} F_t\left(0,\theta_{1,i}\right)u_t, T^{d_1^0-1/2}\sum_{t=T_{j-1}+1}^{T_j} F_t\left(0,\theta_{1,i}\right)u_t\right) = 0.$$

Thus, the estimator using the filter (9) is uncorrelated under H_0 which contrasts the one in Theorem 3. ∎

For ℓ vs. $\ell+1$ breaks, Proposition 6 derives the asymptotic distribution of the unconstrained estimators for the i's regime, assuming *one* additional break in this regime. Let $\tau = \hat{T}_{i-1} + \gamma(\hat{T}_i - \hat{T}_{i-1})$ be the additional break point in regime i.

Proposition 6. *Under Assumptions 1-3 for $i = 1, ..., \ell+1$ and under H_0^ℓ :*

$$a)\ \bar{T}\left(d_i^0\right)\left(\hat{\theta}_{i,\tau}-\theta_i^0\right) \Longrightarrow \begin{pmatrix} \frac{\sqrt{6}}{\pi}B^h(\gamma)/\gamma \\ \frac{\sigma\sqrt{1-2d_1^0}\Gamma(1-d_1^0)\hat{W}^h(\gamma)}{\gamma^{1-2d_i^0}} \end{pmatrix}.$$

$$b)\ \bar{T}\left(d_i^0\right)\left(\hat{\theta}_{\tau,i+1}-\theta_i^0\right) \Longrightarrow \begin{pmatrix} \frac{\sqrt{6}}{\pi}B^h(1-\gamma)/(1-\gamma) \\ \frac{\sigma\sqrt{1-2d_i^0}\Gamma(1-d_i^0)\left(\hat{W}^h(1)-\hat{W}^h(\gamma)\right)}{1-\gamma^{1-2d_i^0}} \end{pmatrix}.$$

Proof. **Part a)** The behavior of the denominator of the estimator follows from Lemma 3. First, the l break fractions are superconsistently estimated. We can use arguments similar to the ones in Theorem 3, to show for the numerator

$$\bar{T}\left(d_i^0\right)\sum_{t=T_{i-1}^0+1}^{\tau} u_t\left(\lambda_{i-1}^0,\theta_i^0\right)F_t\left(\lambda_{i-1}^0,\theta_i^0\right) \Rightarrow \begin{pmatrix} \frac{\pi}{\sqrt{6}}B\left(\gamma\left(\lambda_i^0-\lambda_{i-1}^0\right)\right) \\ \frac{\sigma\tilde{W}_{1/2-d_i^0}\left(\gamma\left(\lambda_i^0-\lambda_{i-1}^0\right)\right)}{\sqrt{1-2d_i^0}\Gamma\left(1-d_i^0\right)} + C\left(\lambda_{i-1}^0,\gamma,d_i^0\right) \end{pmatrix}$$

where $C\left(\lambda_{i-1}^0,\gamma,d_i^0\right)$ is discussed in (19). In particular, the convergence of the first component follows from a functional central limit theorem. For the convergence of the second component, we use Marinucci and Robinson (1999) and that (18) converges in distribution to $C\left(\lambda_{i-1}^0,\gamma,d_i^0\right)$. The additional term is a consequence of the too short filter.

Part b) follows similarly. ∎

Proposition 7 analyzes the estimators corresponding to the ones in Propositions 5 in the bootstrap world. $\overset{p}{\Longrightarrow}$ denotes weak convergence in probability as defined in Gine and Zinn (1990).

Proposition 7. *Under Assumptions 1 and 2 and under H_0 or $H_{1,T}$, the estimators $\hat{\theta}^*$ and $\hat{\theta}_{1,i}^*$ converge weakly in probability ($\overset{p}{\Longrightarrow}$) to the same limits as the ones in Propositions 5.*

Proof. The proof follows from combining results about the convergence of partial sums in the bootstrap world to fractional Brownian Motions with the behavior of the estimators in Propositions 5. It remains to show these convergence results. For this, we incorporate into Kapetanios' (2010) analysis, the estimation of the mean but for a process without a short memory component. Since we analyze the behavior of the bootstrap under $H_0/H_{1,T}$, we filter under the assumption of no breaks. In the notation of Kapetanios (2010), we have to show his Theorem 1

$$\tilde{W}^*_{T,1/2-\hat{d}} = \frac{1}{T^{\hat{d}-1/2}} \sum_{t=1}^{[rT]} \pi_{t-1}\left(\hat{d}-1\right) u_t^* \Longrightarrow \tilde{W}_{1/2-d_1^0}(r) \text{ in probability,}$$

where the convergence is in the sense of Giné and Zinn (1990). $\tilde{W}_{1/2-d_1^0}(r)$ is the fractional Brownian Motion of order $1/2-d_1^0$ defined in Proposition 5, and u_t^* is a bootstrap resample of the residuals of the regression under SSR_0. Hence Kapetanios' (2010) first assumption is clearly satisfied. We have to show

$$\begin{aligned}&1)\ E^*|u_t^*|^r < \infty \text{ in probability for some } r > 2.\\ &2)\ \sup_r |\tilde{W}^*_{T,1/2-d^0}(r) - \tilde{W}^*_{T,1/2-\hat{d}}(r)| = o_{p*}(1).\end{aligned}$$

For 1), we have to show that

$$\frac{1}{T}\sum_{t=1}^{T} |\hat{u}_t - \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t|^r = O_p(1)$$

Write

$$\frac{1}{T}\sum_{t=1}^{T} |\hat{u}_t - \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t|^r \le c\left(A_T + D_T + E_T\right)$$

where

$$A_T = \frac{1}{T}\sum_{t=1}^{T}|u_t|^r, D_T = |\frac{1}{T}\sum_{t=1}^{T}u_t|^r \le KA_T \text{ and } E_T = \frac{1}{T}\sum_{t=1}^{T}|\hat{u}_t - u_t|^r.$$

First, as in Park (2002), A_T and D_T are of order $O_p(1)$. Consider E_T

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}|\hat{u}_t - u_t|^r &= \frac{1}{T}\sum_{t=1}^{T}\Bigg|\Delta_t^d\left(\mu_1^0 + T^{d_1^0-1/2}h_\mu\left(\frac{t}{T}\right) - \mu\right) + \\ &\quad \sum_{j=1}^{t-1}\pi_j\left(d - d_1^0 - T^{-1/2}h_d\left(\frac{t}{T}\right)\right)u_{t-j}\Bigg|^r,\end{aligned}$$

where the second term is $o_p(1)$ following from eq. (4.17) in Wright (1995) and the fact that h_d is bounded. Using $\left(\hat{\mu} - \mu_1^0\right) = O(T^{d_1^0-1/2})$ and the boundedness of h_μ, the first term is also of order $o_p(1)$.

For 2), we need to show

$$\max_s \frac{1}{T^{d_1^0-1/2}}\left|\sum_{t=1}^{s}\pi_{t-1}\left(\hat{d}-1\right)u_t^* - \sum_{t=1}^{s}\pi_t\left(d_1^0-1\right)u_t^*\right| = o_{p*}(1)$$

where u_t^* is an *iid* heterogenous process in the bootstrap probability space, drawn with probability $1/T$ from the residuals $\hat{u}_t$. In particular, defining $v_j^* = u_{t-j}^*, j = 1, ..., t$, the proof follows the same steps as the one in Kapetanios (2010).

Similarly, partial sums converge to Brownian Motions. ∎

APPENDIX B: PROOFS

B.1. Proof of Proposition 1

a) We show that the memory estimation is still consistent for $d^0 > 0$, but inconsistent for $d^0 \leq 0$. We analyze heuristically the case of inconsistent estimation of μ with $0 < d^0 < 1/2$. In particular, for $|\hat{\mu} - \mu^0| > C$ the objective function

$$\frac{1}{T}SSR = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left[\left(\mu^0 - \mu\right)\Delta_t^d 1 + \Delta_t^{d-d^0}u_t\right]^2 \tag{26}$$

converges uniformly in $d \in D$ and μ to

$$\sum_{j=1}^{\infty}\pi_j^2(d - d^0).$$

Therefore, the SSR is still minimized at the true parameter d^0 if $0 < d^0 < 1/2$. The asymptotically negligible terms

$$\left(\mu^0 - \mu\right)^2 K_1 T^{-2d} + \frac{1}{T}2\left(\mu^0 - \mu\right)\sum_{t=1}^{T}\pi_t(d-1)\sum_{j=1}^{t-1}\pi_j(d - d^0)u_{t-j}$$

lead to the mentioned finite sample effects which depend on d^0, $\left(\mu^0 - \mu\right)$ and T. Especially, for d^0 close to 0, the bias can be huge leading to a highly upward biased estimator in finite samples. On the support $0 \leq d < 1/2$, the limit of the expression (26) is not continuous due to the additional term $I\left(d = 0\right)\left(\mu^0 - \mu\right)^2$. Clearly, for $|\hat{\mu} - \mu^0| > C$, (26) is in the limit not minimized in $d = 0$. In consequence, the estimator is not consistent for $d^0 = 0$. The same argument is obviously true if we do not estimate μ, just set $\hat{\mu} = 0$.

b)We have to show that

$$\hat{\mu}\left(d\right) - \mu^0 = O_p\left(T^{d^0 - 1/2}\right) \text{ uniformly in } d \in D,$$

by showing convergence of the *fidi* and tightness. For tightness we show in Rachinger (2011) that

$$E|T^{1/2-d^0}\left(\hat{\mu}\left(d_2\right) - \mu^0\right) - T^{1/2-d^0}\left(\hat{\mu}\left(d_1\right) - \mu^0\right)|^2 \leq K|d_2 - d_1|^2. \tag{27}$$

B.2. Proof of Theorem 1

We provide the main steps of the proof and indicate where they differ from the ones of Boldea and Hall (2010). Define

$$d_t\left(\lambda_{k-1}, \theta_k\right) = \hat{u}_t\left(\lambda_{k-1}, \theta_k\right) - u_t, \tag{28}$$

where $\hat{u}_t(\lambda_{k-1}, \theta_k)$ is defined in (4), for $t \in I_j^0 \cap \hat{I}_k$ with $I_j^0 = [T_{j-1}^0 + 1, T_j^0]$ and $\hat{I}_k = [\hat{T}_{k-1} + 1, \hat{T}_k]$ and $k, j = 1, ..., m+1$. $d_t(\lambda_{k-1}, \theta_k)$ and $\hat{u}_t^{(k)}(\lambda_{k-1}, \theta_k)$ depend also on $\{\theta_i^0\}, \{\theta_{i-1}^0, \theta_i^0\}$ and $\{\theta_{i-1}^0, \theta_i^0, \theta_{i+1}^0\}$ in the cases $\lambda_{k-1}^0 < \lambda_{k-1} < t < \lambda_k^0, \lambda_{k-1} < \lambda_{k-1}^0 < t < \lambda_k^0$ and $\lambda_{k-1} < \lambda_{k-1}^0 < \lambda_k^0 < t$ respectively. Boldea and Hall (2010) work with a different expression separating true quantities from estimated ones. In our case, both are fractionally integrated and we work rather with expression (28). First, we focus on the break in T_i^0. For simplicity, we denote $d_t(\lambda_{k-1}, \theta_k)$ and $\hat{u}_t(\lambda_{k-1}, \theta_k)$ as d_t and $\hat{u}_t$. From the CSS estimation we get

$$\sum_{t=1}^{T} \hat{u}_t^2 = \sum_{t=1}^{T} u_t^2 + \sum_{t=1}^{T} d_t^2 - 2\sum_{t=1}^{T} d_t u_t$$

implying that

$$T^{-\delta_i} \sum_{t=1}^{T} d_t^2 + 2T^{-\delta_i} \sum_{t=1}^{T} u_t d_t \leq 0, \tag{29}$$

where $\delta_i = 1$ for a break in T_i^0 in memory and mean or only in memory and $\delta_i = 1 - 2d_i^0$ for a break only in the mean. Denoting $q_T \sim O_p(T^b)$ if $P(|q_T| > T^b) < \bar{\eta}$ for $T \geq T(\bar{\eta})$ for some $b \epsilon R$ and any $\bar{\eta} > 0$ and $q_T \sim O_p^+(T^b)$ if $plim q_T$ is positive, the proof of the consistency works by showing that $T^{-\delta_i} \sum_{t=1}^{T} d_t u_t = o_p(1)$ and $T^{-\delta_i} \sum_{t=1}^{T} d_t^2 = O_p^+(1)$, when the break fraction λ_i is inconsistently estimated. In particular, we use Lemma 1 and 2 for proving Theorem 1. Inequality (29) together with Part (i) of Lemma 2 would imply that $T^{-\delta_i} \sum_{t=1}^{T} d_t^2 = o_p(1)$ which would contradict part (ii) of Lemma 2. In particular, Lemma 2 is also true for an estimator $\hat{\lambda}_i < \lambda_i^0$ and, in consequence, the break fraction is not estimated too low. The same argument applies for $\hat{\lambda}_i > \lambda_i^0$ and we conclude that the break fraction estimator is consistent.

B.3. Proof of Theorem 2

This proof follows closely the proof of Theorem 2 of Boldea and Hall (2010). We consider the case of *three* breaks.We analyze two different cases of changing parameters that require a different analysis:

- case A: a break in memory and mean or a break in memory.
- case B: a break in mean; $d_1^0 = d_2^0 = d_3^0 \geq 0$.

Consistency of the three breaks is already established. Because of consistency we only have to consider the behavior of the break points in

$$V_{\in} = \left\{(T_1, T_2, T_3) : |T_i - T_i^0| \leq \varepsilon T \ (i = 1, 2, 3)\right\}.$$

First, consider case $\hat{T}_2 < T_2^0$. In contrast to Boldea and Hall (2010), here the argument is not symmetric and we have to consider also the case $T_2^0 > T_2$. The proof works basically by showing that the break point is with a very small probability in the set

$$V_{\in}(C) = \left\{(T_1, T_2, T_3) : |T_i - T_i^0| \leq \varepsilon T \ (i = 1, 2, 3); \ \Delta_2 = T_2^0 - T_2 > C\right\}.$$

Hence with large probability $|\hat{T}_2 - T_2^0| < C$. We will show that if $T_2 \in V_{\in}(C)$,

$$P\left\{\min_{V_{\in}(C)} \frac{S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)}{\Delta_2^{\delta}} \leq 0\right\} < \eta, \text{ for } T \geq T(\eta) \tag{30}$$

contradicting the sum of squares minimization and implying that T_2 does not belong to $V_{\in}(C)$. For case A, $\delta = 1$ and, for case B, $\delta = 1 - 2d_2^0$. We show that

$$S_T(T_1, T_2, T_3) - S_T\left(T_1, T_2^0, T_3\right) = (SSR1 - SSR3) - (SSR2 - SSR3)$$

is positive with high probability for large T picking ε and C where

$$\begin{aligned} SSR1 &= S_T(T_1, T_2, T_3), \; SSR2 = S_T\left(T_1, T_2^0, T_3\right) \text{ and} \\ SSR3 &= S_T\left(T_1, T_2, T_2^0, T_3\right). \end{aligned}$$

The behavior of the corresponding estimators is discussed in Lemma 4. We locate the dominating terms in $S_T(T_1, T_2, T_3) - S_T\left(T_1, T_2^0, T_3\right)$ and show that at least some are positive with large probability. Equation (30) is equivalent to

$$\Delta_2^{-\delta}(SSR1 - SSR2) \sim O_p^+(1) \tag{31}$$

We introduce some notation:

$$I_1 = [1, T_1]; I_2 = [T_1 + 1, T_2], I_2^\Delta = [T_2 + 1, T_2^0], \; I_3 = [T_2^0 + 1, T_3], \; I_4 = [T_3 + 1, T].$$

Next,

$$\begin{aligned} \frac{SSR_1 - SSR_3}{\Delta_2^\delta} &= \frac{1}{\Delta_2^\delta}\left[\sum\nolimits_{I_2^\Delta}\left[u_t^2(\theta_3^{**}) - u_t^2\left(\theta_2^\delta\right)\right] + \sum\nolimits_{I_3}\left[u_t^2(\theta_3^{**}) - u_t^2(\theta_3^{***})\right]\right] \\ &= D_1 + D_2. \end{aligned}$$

Since θ_3^{**} estimates θ_3^0 and θ_2^δ estimates θ_2^0, there is a mismatch in D_1, while there is none in D_2 (θ_3^{**} and θ_3^* estimate θ_3^0). In Rachinger (2011), we use Lemmata 5 and 6 to show in a similar way as in Boldea and Hall (2010) that D_1 dominates in the limit D_2. We further show that $\Delta_2^{-\delta}(SSR2 - SSR3) \sim o_p(1)$.

In Theorem 1 and 2, we focus on the break in T_i and assume that all other break fractions are estimated consistently (Theorem 1) and at the rate T (Theorem 2). It suffices to discuss this case since it is the least favorable case for the contradiction that is used for deriving the consistency of the break fraction λ_i.

B.4. Proof of Theorem 3

We first obtain consistency and $\sqrt{T}$-rate convergence of the estimator d_i when it is calculated with estimated rather than true endpoints. Given these results, we establish $T^{1/2-d_i^0}$ rate convergence for the estimator of μ_i. Finally, we show that the estimators using the estimated break points have the same asymptotic distribution as the ones using the true ones.

We start with the asymptotic distribution of the estimators assuming that the break points are the true ones. By the superconsistency of the break fractions, this distribution will correspond to the one when the break points are estimated. First, the consistency of the estimator d_i follows from Lemma 3a). The asymptotic distribution of the estimator follows from Lemma 3a) and b). Because the residuals evaluated at the true parameters and true break fractions $u_t\left(\lambda_{i-1}^0,\theta_i^0\right)$ differ from u_t, the variance of the mean estimator contains the additional term (6). Similarly, the covariance between the estimators μ_i and μ_j

$$D_{ij}^{\mu}\left(\left\{\lambda_k^0,\lambda_{k-1}^0,d_k^0\right\}_{k=i,j}\right)=\frac{\Gamma^2\left(1-d_i^0\right)\left(1-2d_i^0\right)}{\left(\lambda_i^0-\lambda_{i-1}^0\right)^{1-2d_i^0}}\frac{\Gamma^2\left(1-d_j^0\right)\left(1-2d_j^0\right)}{\left(\lambda_j^0-\lambda_{j-1}^0\right)^{1-2d_j^0}}A_{ij}^{\mu} \tag{32}$$

where

$$\begin{aligned} A_{ij}^{\mu} &= \lim_{T\to\infty}T^{-1}\sum_{k=1}^{\left[\lambda_{i-1}^0T\right]}\left(T^{d_i^0}\sum_{t=1}^{\left[\left(\lambda_i^0-\lambda_{i-1}^0\right)T\right]}\pi_{t-1}\left(d_i^0-1\right)\sum_{j=0}^{t}\pi_j\left(d_i^0\right)\pi_{\left[\lambda_{i-1}^0T\right]+t-j-k}\left(-d_i^0\right)\right) \\ &\quad\cdot\left(T^{d_j^0}\sum_{t=1}^{T_j^0-T_{j-1}^0}\pi_{t-1}\left(d_j^0-1\right)\sum_{j=0}^{t}\pi_j\left(d_j^0\right)\pi_{T_{j-1}^0+t-j-k}\left(-d_j^0\right)\right). \end{aligned}$$

Consequently, the estimators μ_i and μ_j are not asymptotically independent.

Next, the proof of consistency of the parameter estimates in the two regimes using the estimated rather than the true break points and the proof that the asymptotic distribution corresponds to the one assuming the true break point follows the same lines as in Boldea and Hall (2010).

B.5. Proof of Theorem 4

For deriving the asymptotic distribution of the test statistic (10), we, first, show for the denominator under the local alternative:

$$\begin{aligned} SSR_k\left(\boldsymbol{\lambda}\right) &= \sum_{i=1}^{k+1}\frac{1}{T}\sum_{t=T_{i-1}+1}^{T_i}\left(\left(\mu_1^0-T^{d_1^0-1/2}h_{\mu}\left(\frac{t}{T}\right)-\mu_i\right)\Delta_t^{d_i}1+\Delta_t^{d_i-d_1^0-\frac{1}{\sqrt{T}}h_d\left(\frac{t}{T}\right)}u_t\right)^2 \\ &= \lambda_1\sum_{k=0}^{\infty}\pi_k^2\left(d_1-d_1^0\right)+\sum_{i=2}^{k+1}\left(\lambda_i-\lambda_{i-1}\right)\sum_{k=0}^{\infty}\pi_k^2\left(d_i-d_1^0\right)+o_p\left(1\right)=\sigma^2+o_p\left(1\right) \end{aligned}$$

where the terms including $\left(\mu_j^0-\mu_i\right)$ are negligible by Lemma 1 and the convergence is a consequence of Lemma 3 a). Next, we discuss the behavior of the numerator. As in Boldea and Hall (2010), we write

$$SSR_0-SSR_k\left(\boldsymbol{\lambda}\right)=\sum_{t=1}^{T}u_t^2(\hat{\theta})-\sum_{i=1}^{k+1}\sum_{t=\lambda_{i-1}T+1}^{\lambda_iT}u_t^2(\hat{\theta}_i)=...=\sum_{i=1}^{k}F_{T,i}^{*}$$

with

$$F_{T,i}^{*}=D^R\left(1,i+1\right)-D^R\left(1,i\right)-D^U\left(i+1,i+1\right) \tag{33}$$

where the index $1,i$ indicates summing over $[1,T_i]$ and i over $[T_{i-1}+1,T_i]$. $D^R\left(1,i\right)=\sum_{1,i}[u_t^2(\hat{\theta}_{1,i})-u_t^2]$ and $D^U\left(i,i\right)=\sum_i[u_t^2(\hat{\theta}_i)-u_t^2]$. We start with the term $D^R\left(1,i\right)$

$$D^R\left(1,i\right)=\sum\nolimits_{1,i}d_t^2(\hat{\theta}_{1,i},\theta_1^0)-2\sum\nolimits_{1,i}u_td_t(\hat{\theta}_{1,i},\theta_1^0)=I^R+II^R$$

As in Boldea and Hall (2010), using a mean value theorem (MVT),

$$\begin{aligned} I^R &= \left[T^{1/2}(\hat{d}_{1,i}-d_1^0)\right]^2 T^{-1}\sum_{1,i} F_{d,t}^2\left(\bar{\theta}_{1,i,t}\right) \\ &\quad + \left[T^{1/2-d_1^0}\left(\hat{\mu}_{1,i}-\mu_1^0\right)\right]^2 T^{-1+2d_1^0}\sum_{1,i} F_{\mu,t}^2\left(\bar{\theta}_{1,i,t}\right) \\ II^R &= 2\left[T^{1/2}(\hat{d}_{1,i}-d_1^0)\right] T^{-1/2}\sum_{1,i} u_t F_{d,t}\left(\bar{\theta}_{1,i,t}\right) \\ &\quad +2\left[T^{1/2-d_1^0}\left(\hat{\mu}_{1,i}-\mu_1^0\right)\right] T^{-1/2+d_1^0}\sum_{1,i} u_t F_{\mu,t}'\left(\bar{\theta}_{1,i,t}\right) \end{aligned}$$

where $\bar{\theta}_{1,i,t}$ lies in the segment line $\hat{\theta}_{1,i}$ and θ_1^0. Also here since $\bar{\theta}_{1,i,t} \xrightarrow{p} \theta_1^0$ for each t and $E\left[F_t(\theta)\, F_t'(\theta)\right]$ has uniform bounds, from Proposition 5 part b) and its proof, we obtain

$$D^R(1,i) \Longrightarrow -\sigma^2\left(\lambda_i^{-1}\left(B^h(\lambda_i)\right)^2+\lambda_i^{-1+2d_1^0}\left(\tilde{W}^h(\lambda_i)\right)^2\right) \tag{34}$$

For the term $D^U(i,i)$ using Proposition 5 and similar arguments as the previous ones, we obtain,

$$\begin{aligned} D^U(i,i) \quad \Longrightarrow \quad & -\sigma^2\left[\left(\lambda_i-\lambda_{i-1}\right)^{-1}\left(B^h(\lambda_i)-B^h(\lambda_{i-1})\right)^2\right. \\ & \left. +\left(\lambda_i^{1-2d_1^0}-\lambda_{i-1}^{1-2d_1^0}\right)^{-1}\left(\tilde{W}^h(\lambda_i)-\tilde{W}^h(\lambda_{i-1})\right)^2\right]. \end{aligned}$$

Finally, combining the two terms and using a continuous mapping theorem (CMT) for the *sup* functional leads to the stated test statistic.

The independence of the estimates of memory and mean, discussed in Theorem 3, implies the additiveness of the test statistic.

B.6. Proof of Theorem 5

First, the estimated break fractions converge to the true ones at rate T, for breaks in the memory H_1^d, the mean H_1^μ and in both $\mathrm{H}_1^{d,\mu}$. Under the alternative, the test statistic (10) diverges since its denominator still converges to σ^2 because break fractions and regime parameters are consistently estimated. If there is at least one break in the memory or in memory and mean, $D^R(1,i)$ is of order $O_p(T)$ and $D^U(i,i)$ is of order $O_p(T^{1-2d_i^0})$ because the mean estimators stop being consistent. Thus, the test statistic diverges at rate T. Equally, we find that, if only the mean is changing, $SSR_0 - SSR_k(\boldsymbol{\lambda}) = O_p(T^{1-2d^0})$ and the test statistic diverges at rate T^{1-2d^0}. If we tested for a break only in the memory or only in the mean, the tests reject under the alternative of a break in the tested and in both parameters. Under the alternative of a break only in the not tested parameter, the tests reject asymptotically with probability α.

B.7. Proof of Proposition 2

Under the hypothesis of one break at λ_1^0, the estimator λ_1 converges at rate T to the break fraction λ_1^0.

Proof of a) Components from Theorem 4 involving the estimation of the mean are negligible. Finally, the components involving the memory behave as in Theorem 4 with the difference that now λ_1 does not have a spurious limit and thus the limit will be a function of the true break fraction. Therefore, the test statistic corresponds to the one of a usual Chow test.

Proof of b) For testing a break in the mean, terms involving the break in the memory are again negligible. Using the filter truncated at the supposed break points, we obtain for the estimator of the mean,

$$\left(\hat{\mu}-\mu^{0}\right)=\frac{\sum_{t=1}^{\lambda_1 T}\left(\Delta_t^{d_1}1\right)\hat{u}_t+\sum_{t=\lambda_1 T+1}^{T}\left(\Delta_{t-\lambda_1 T}^{d_2}1\right)\hat{u}_t}{\sum_{t=1}^{\lambda_1 T}\left(\Delta_t^{d_1}1\right)^2+\sum_{t=\lambda_1 T+1}^{T}\left(\Delta_{t-\lambda_1 T}^{d_2}1\right)^2}.$$

It is easy to show that for $d_1^0 < d_2^0$ in numerator and denominator, the first term dominates and for $d_1^0 > d_2^0$ the second one does. In (33), the first and third term cancel. From the second term of (33), follows the result. For the latter, as mentioned before, $\hat{u}_t$ contains some term similar to the one in Theorem 7 coming from a too short filter causing the increased variance.

B.8. Proof of Theorem 6

Under $\mathrm{H}_0 : m = l$, as in Boldea and Hall (2010), the test statistic can be written as

$$F_T\left(l+1|l\right)=\max_{1\le i\le l}\ \sup_{\tau\in\Delta_{i,\eta}} F_{T,i}\left(l+1|l\right)/\hat{\sigma}_i^2$$

where $F_{T,i}\left(l+1|l\right)=SSR(\hat{T}_{i-1},\hat{T}_i)-SSR(\hat{T}_{i-1},\tau)-SSR(\tau,\hat{T}_i)$ with $SSR(\hat{T}_{i-1},\hat{T}_i)$ being the sum of squared residuals for the segment $[\hat{T}_{i-1},\hat{T}_i]$. Based on Proposition 6, the proof follows using similar arguments to the ones in Theorem 4.

B.9. Proof of Theorem 7

We show that the bootstrap based test (22) has the same asymptotic distribution as the one in Theorem 4. The estimates $\hat{d},\hat{\mu}$ play the role of the true parameter values in Theorem 4. The estimates $\hat{\theta}_{1,i}^{*},\hat{\theta}_i^{*}$ denote the estimates for the bootstrap data $\{y_t^*\}_{t=1}^{T}$. From Proposition 7, these estimators converge weakly in probability to the same limits as the ones in Proposition 5.

For establishing the asymptotic distribution of the test statistic, first we have to show for the denominator that

$$SSR_k^*\left(\boldsymbol{\lambda}\right)/\left(T-\left(k+1\right)p\right)=\sigma^2+o_p^*.$$

In particular,

$$\begin{aligned}SSR_k^*\left(\boldsymbol{\lambda}\right) &= \sum_{i=1}^{k+1}\frac{1}{T}\sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_i T]}\left(\left(\hat{\mu}-\hat{\mu}_i^*\right)\Delta_t^{\hat{d}_i^*}1+\Delta_t^{\hat{d}_i^*-\hat{d}}u_t\right)^2\\ &= \sum_{i=1}^{k+1}\left(\lambda_i-\lambda_{i-1}\right)\sum_{k=0}^{\infty}\pi_k^2\left(\hat{d}_i^*-\hat{d}\right)+o_p^*\left(1\right)=\sigma^2+o_p^*\left(1\right)\end{aligned}$$

To prove the convergence we show that $E^*\left[\frac{1}{T}SSR_1^*\right]=\hat{\sigma}^2$ and $Var^*\left[\frac{1}{T}SSR_1^*\right]=o_p(1)$. For the former,

$$E^*\sum_{i=1}^{k+1}\frac{1}{T}\sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_iT]}\left(\left(\hat{\mu}-\hat{\mu}_i^*\right)\Delta_t^{\hat{d}_i^*}1+\Delta_t^{\hat{d}_i^*-\hat{d}}u_t\right)^2=\hat{\sigma}^2\frac{1}{T}\sum_{i=1}^{k+1}\sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_iT]}\sum_{j=1}^{t-1}\pi_j^2\left(\hat{d}_i^*-\hat{d}\right).$$

For the second term we apply a variant of the Lemma 1, substituting d_1^0 by $\hat{d}$, and a similar argument as the one for the first term. The convergence follows from $T\to\infty$ and the fact that $\hat{d}_1^*$ and $\hat{d}$ converge to d_1^0 and $\hat{\sigma}^2$ converges to σ^2. The behavior of the numerator follows from applying Proposition 7 to the Proof of Theorem 4. Finally, from applying a CMT, we obtain that $\sup_\lambda F_T^*(\lambda,k,p)$ converges weakly in probability to the corresponding limit in Theorem 4 for $h_d=h_\mu=0$.

Proof of c) fixed alternative The test is consistent because the bootstrap test statistic converges to a constant and the original test statistic diverges. For the former, under H_1, the estimators $\hat{d}$ and $\hat{\mu}$ converge to weighted averages of the true parameter values. Applying the test to the newly integrated series, the resulting test statistic has still a bounded limit distribution. Since, from Theorem 5, the test statistic diverges under H_1, the bootstrap test is consistent.

B.10. Proof of Proposition 3

We first show 1). Note that terms including μ are uniformly of order $o_p(1)$. For $i=1$,

$$T^{1/2}\left(d_1(\lambda)-d^0\right)=\frac{T^{-1/2}\sum\limits_{t=1}^{[\lambda T]}\left(\sum\limits_{j=0}^{t-1}\dot{\pi}_j(0)\,u_{t-j}\right)u_t}{T^{-1}\sum\limits_{t=1}^{[\lambda T]}\left(\sum\limits_{j=0}^{t-1}\dot{\pi}_j(0)\,u_{t-j}\right)^2}+o_p(1).$$

For $j=1,2$, N_j denotes the numerator and D_j the denominator of $d_1(\lambda_j)$. Thus,

$$d_1(\lambda_2)-d_1(\lambda_1)=\frac{N_2}{D_2}-\frac{N_1}{D_1}=...=\frac{N_1}{D_1D_2}(D_2-D_1)+\frac{1}{D_2}(N_1-N_2).$$

In consequence, we can show tightness for numerator and denominator separately. For the latter, for showing tightness we need to show that

$$\left\|T^{-1}\sum_{t=[\lambda_1T]+1}^{[\lambda_2T]}\left(\sum_{j=0}^{t-1}\dot{\pi}_j(0)\,u_{t-j}\right)^2\right\|_2^2\le K\left|\lambda_2-\lambda_1\right|^2,$$

From a triangle inequality,

$$\left\|T^{-1}\sum_{t=[\lambda_1T]+1}^{[\lambda_2T]}\left(\sum_{j=0}^{t-1}\dot{\pi}_j(0)\,u_{t-j}\right)^2\right\|_2\le T^{-1}\sum_{t=[\lambda_1T]+1}^{[\lambda_2T]}\left\|\left(\sum_{j=0}^{t-1}\dot{\pi}_j(0)\,u_{t-j}\right)^2\right\|_2\le K\left|\lambda_2-\lambda_1\right|,$$

where the boundedness of the norm follows from previous arguments.

$$T^{-1}\sum_{t=1}^{[\lambda T]}\left(\sum_{j=0}^{t-1}\dot{\pi}_j\left(0\right)\right)^2$$

For the numerator, weak convergence in λ follows from a standard FCLT.

Next using $d_i\left(\lambda\right)-d^0=O_p\left(T^{-1/2}\right)$, we show 2). For $i=1$,

$$T^{1/2-d^0}\left(\mu_1\left(\lambda\right)-\mu^0\right)=\frac{T^{d^0-1/2}\sum_{t=1}^{[\lambda T]}\left(\Delta_t^{d^0}1\right)u_t}{T^{2d^0-1}\sum_{t=1}^{[\lambda T]}\left(\Delta_t^{d^0}1\right)^2}+o_p\left(1\right).$$

Next,

$$\mu_1\left(\lambda_2\right)-\mu_1\left(\lambda_1\right)=\frac{N_1}{D_1D_2}\left(D_2-D_1\right)+\frac{1}{D_2}\left(N_1-N_2\right).$$

Tightness for the denominator follows directly from its deterministic character. For the numerator, we can apply a fractional FCLT (Marinucci and Robinson, 1999) to show that it converges weakly to a fractional Brownian Motion.