Econometric Theory, 2012, Page 1 of 35.
doi:10.1017/S0266466612000357

WALD TESTS FOR DETECTING MULTIPLE STRUCTURAL CHANGES IN PERSISTENCE

MOHITOSH KEJRIWAL
Purdue University

PIERRE PERRON
Boston University

JING ZHOU
Orient Securities Company Limited

This paper considers the problem of testing for multiple structural changes in the persistence of a univariate time series. We propose sup-Wald tests of the null hypothesis that the process has an autoregressive unit root throughout the sample against the alternative hypothesis that the process alternates between stationary and unit root regimes. We derive the limit distributions of the tests under the null and establish their consistency under the relevant alternatives. We further show that the tests are inconsistent when directed against the incorrect alternative, thereby enabling identification of the nature of persistence in the initial regime. We also propose hybrid testing procedures that allow ruling out of stable stationary processes or ones that are subject to only stationary changes under the null, thereby aiding the researcher in interpreting a rejection as emanating from a switch between a unit root and stationary regime. The computation of the test statistics as well as asymptotic critical values is facilitated by the dynamic programming algorithm proposed in Perron and Qu (2006, *Journal of Econometrics* 134, 373–399) which allows imposing within- and cross-regime restrictions on the parameters. Finally, we present Monte Carlo evidence to show that the proposed procedures perform well in finite samples relative to those available in the literature.

1. INTRODUCTION

Issues related to the detection and estimation of structural change in time series models have received a great deal of attention in both the statistics and econometrics literature (see Perron, 2006, for a survey). Substantial advances have been

Perron acknowledges financial support for this work from the National Science Foundation under Grant SES-0649350. The authors are grateful to Robert Taylor (the co-editor) and two anonymous referees for useful comments and suggestions that helped improve the paper. Address correspondence to Mohitosh Kejriwal, Krannert School of Management, Purdue University, 403 West State Street, West Lafayette IN 47907 USA; e-mail: mkejriwa@purdue.edu.

© Cambridge University Press 2012

made to cover models at a level of generality that allows a host of interesting empirical applications. These include models with general stationary regressors, models with trending variables and possible unit roots, cointegrated models, and long memory processes, among others. Also of interest is the interplay between structural changes and unit roots (Perron, 1989). The literature on testing for a change in the persistence of a time series is less extensive and relatively recent. If such a change preserves the stationarity properties of the series in the respective regimes, methods developed in the context of stationary data can still be applied (see Andrews, 1993; Bai and Perron, 1998; 2003). In many cases, however, a process may switch from one with an autoregressive unit root [$I(1)$] to a stationary one [$I(0)$] or vice versa. This has been an issue of substantial empirical interest, especially concerning inflation rate series (e.g., Barsky, 1987; Burdekin and Siklos, 1999), short-term interest rates (e.g., Mankiw, Miron, and Wei, 1987), government budget deficits (e.g., Hakkio and Rush, 1991), and real output (e.g., Delong and Summers, 1988). Kim (2003) shows that standard unit root tests are not consistent against processes displaying a shift from stationarity to nonstationarity and vice versa. Hence, separate methods are needed to distinguish between a process with stable persistence and one that undergoes a shift in persistence over a given period.

Kim (2000), Busetti and Taylor (2004), and Taylor (2005) consider testing the null hypothesis that the series is $I(0)$ throughout the sample versus the alternative that it switches from $I(0)$ to $I(1)$ and vice versa. Harvey, Leybourne, and Taylor (2006) propose test statistics that allow the process to be $I(1)$ or $I(0)$ throughout under the null. The tests are based on partial sums of residuals obtained by regressing the data on a constant or a constant and time trend. Leybourne, Kim, Smith, and Newbold (2003) consider testing the null hypothesis of a stable unit root process versus the same alternatives based on the minimal value of the locally generalized least squares (GLS) detrended augmented Dickey-Fuller (*ADF*) unit root statistic developed in Elliott, Rothenberg, and Stock (1996) over subsamples of the data. They propose different test statistics depending on whether the initial regime is $I(1)$ or $I(0)$. When the direction of the change is unknown, they consider the minimal value of the pair of statistics for each case. Kurozumi (2005) suggests an alternative testing procedure based on the Lagrange multiplier (LM) principle, while Leybourne, Kim, and Taylor (2007a) develop tests of the unit root null based on standardized cumulative sums of squared subsample residuals that do not spuriously reject when the series is a constant $I(0)$ process. Chong (2001) studies the asymptotic properties of the estimated parameters in the first-order autoregressive model with a single break in persistence.

The above tests are designed to detect a single change in persistence and do not allow for multiple changes. Single break tests usually have low power in detecting processes that display multiple shifts in persistence. It is thus useful to develop tests that are valid in the presence of multiple structural changes. In a recent paper, Leybourne, Kim, and Taylor (2007b) develop tests of the unit root null hypothesis based on doubly recursive sequences of *ADF*-type unit root statistics and

associated breakpoint estimators. Their proposed procedure can accommodate processes that exhibit multiple changes in persistence and are valid regardless of the direction of change(s). In particular, they demonstrate the consistency of their tests against such alternatives and show that their procedure can be used to consistently partition the data into its separate $I(0)$ and $I(1)$ regimes. Kang, Kim, and Morley (2009) consider an alternative approach to analyzing multiple regime shifts in U.S. inflation persistence based on an unobserved components model with Markov-switching parameters.

As is evident from this brief review, most tests for changes in persistence are based on either partial sums of the (demeaned or detrended) data or on unit root statistics applied to various data subsamples. In contrast, this paper proposes sup-Wald tests of the null hypothesis that the process is $I(1)$ against the alternative hypothesis that the process alternates between stationary and $I(1)$ regimes. The tests are based on the difference between the sum of squared residuals from the unit root model and those from a model that allows shifts in persistence between stationary and nonstationary regimes. We consider tests for both single and multiple changes in persistence. The limit distributions of the tests are derived under the null, and their consistency is established under the relevant alternatives. We further show that the tests are inconsistent when directed against the incorrect alternative, thereby allowing the researcher to identify the nature of persistence in the initial regime. We also propose hybrid testing procedures that allow ruling out of stable stationary processes or ones that are are subject to only stationary changes under the null, thereby aiding the researcher in interpreting a rejection as emanating from a switch between a unit root and stationary regime. We further discuss how our tests can be used to distinguish between persistence breaks and pure level or trend breaks. The computation of the test statistics as well as asymptotic critical values is facilitated by the dynamic programming algorithm proposed in Perron and Qu (2006), which allows the minimization of the sum of squared residuals under the alternative hypothesis while imposing within- and cross-regime restrictions on the parameters. We also propose estimators for the break dates that can be employed once evidence against a stable persistence parameter is obtained. The performance of the proposed test statistics in small samples is evaluated via an extensive Monte Carlo study.

The paper is organized as follows. Section 2 presents the models, the test statistics, and issues related to the computation of the statistics. Section 3 details the asymptotic properties of the test statistics under the null and alternative hypotheses. Section 4 proposes hybrid testing procedures that allow ruling out processes that are constant $I(0)$ or ones that are subject to only $I(0)$ changes under the null. Section 5 suggests estimators for the locations of the break points that can be applied following evidence against the null hypothesis. Monte Carlo simulations are presented in Section 6 to assess the adequacy of the asymptotic approximations in finite samples. Recommendations for applied work are also included. Section 7 concludes. All technical derivations are in the Appendix.

2. THE MODELS AND TEST STATISTICS

Consider a scalar random variable y_t generated by

$$y_t = c_i + \alpha_i y_{t-1} + u_{it} \tag{1}$$

for $t \in [T_{i-1}+1, T_i]$, $i = 1, \ldots, m+1$, with the convention that $T_0 = 0$ and $T_{m+1} = T$, where T is the sample size. The vector of break fractions is $\lambda = (\lambda_1, \ldots, \lambda_m)$ with $\lambda_i = T_i/T$ for $i = 1, \ldots, m$. Hence, we have m breaks and $m+1$ regimes that increase in length in the same proportion as T increases. The errors $\{u_{it}\}$ are generated by the stationary linear process

$$u_{it} = d_i(L)v_{it},\ d_i(L) = \sum_{s=0}^{\infty} d_{is}L^s, \tag{2}$$

where $\sum_{s=1}^{\infty} s\,|d_{is}| < \infty$. Also, α_i should be understood as standing for the sum of the coefficients in the autoregressive representation for y_t in regime i. We make the following assumptions regarding the innovation process $\{v_{it}\}$ and u_{it} for $i = 1, \ldots, m+1$.

Assumption A1. The process $\{v_{it}\}$ is a martingale difference sequence with $\mathbb{E}(v_{it}^2|v_{it-1}, \ldots) = \sigma^2$, $\mathbb{E}(|v_{it}|^r|v_{it-1}, \ldots) = \kappa_{ir}$ $(r = 3, 4)$, and $\sup_t \mathbb{E}(|v_{it}|^{4+\beta}|v_{it-1}, \ldots) = \kappa_i < \infty$ for some $\beta > 0$.

Assumption A2. All roots of $d_i(L)$ are outside the unit circle.

We consider the following two models depending on whether the initial regime contains a unit root or not: Model 1a: $c_i = 0$, $\alpha_i = 1$ in odd regimes and $|\alpha_i| < 1$ in even regimes; Model 1b: $c_i = 0$, $\alpha_i = 1$ in even regimes and $|\alpha_i| < 1$ in odd regimes. In Model 1a, the process alternates between a unit root and a stationary process with a unit root in the first regime. Model 1b is similar except that the first regime is stationary. To allow for the possibility of trending data, we also consider the process

$$y_t = c_i + b_i t + \alpha_i y_{t-1} + u_{it}.$$

The corresponding models are: Model 2a: $\alpha_i = 1$, $b_i = 0$ in odd regimes and $|\alpha_i| < 1$ in even regimes; Model 2b: $\alpha_i = 1$, $b_i = 0$ in even regimes and $|\alpha_i| < 1$ in odd regimes. We are interested in testing the null hypothesis that y_t is $I(1)$ throughout the sample. For Models 1a and 1b, this implies H_0: $c_i = 0$, $\alpha_i = 1$ for all i. For Models 2a and 2b, the null hypothesis is H_0: $c_i = c$, $b_i = 0$, $\alpha_i = 1$ for all i. In this case, the data generating process (DGP) is denoted by

$$y_t = c + y_{t-1} + u_t, \tag{3}$$

where $u_t = d(L)v_t$, $d(L) = \sum_{s=0}^{\infty} d_s L^s$ with v_t and $d(L)$ satisfying Assumptions A1–A2 and $\sum_{s=1}^{\infty} s|d_s| < \infty$.

It is important to note that under the alternative hypothesis the process generating the data is such that all parameters are allowed to change across regimes. Hence, level shifts and changes in the slope of the trend are allowed, as well as changes in the dynamics and the variance of the errors. We, however, shall not construct test statistics that exploit the possible changes in the dynamics or the variance of the errors. This is because we wish to direct the test against potential changes in the I(0)/I(1) nature of the process to ensure the highest power possible. Also, allowing for breaks in dynamics under the null would lead to limit distributions that depend on the (unknown) number and location of these breaks, thereby making asymptotic inference difficult. A joint test on all parameters would not be particularly informative given the difficulty in interpreting a rejection. As shown in Section 6.2, our test does not have much power against pure changes in short-run dynamics but is powerful when there is a change in both persistence and these dynamics. We nevertheless allow for concurrent changes in level and slope of the trend function, since these often occur simultaneously with a change in persistence and can allow tests with higher power.

We first consider the test statistics for nontrending data, i.e., those based on Models 1a and 1b. Given the fact that the process has an autoregressive representation that can be approximated by an $AR(l_T)$ for some sequence l_T increasing with the sample size, the starting point is to consider the regression

$$\Delta y_t = c_i + (\alpha_i - 1) y_{t-1} + \sum_{j=1}^{l_T} \pi_j \Delta y_{t-j} + v_t^*. \tag{4}$$

In accordance with the discussion above, the coefficients π_j pertaining to the dynamics are not allowed to change across regimes. Also, the tests are based on the constrained and unconstrained sum of squared residuals, which follows a least-squares approach that does not exploit potential changes in the variance of the errors.

We study two types of tests in this section. First, we consider the Wald test that applies when the alternative involves a fixed value $m = k$ of changes. For Models 1a–1b, the test is defined as

$$\begin{aligned} F_{1a}(\lambda, k) &= (T - k - l_T)(SSR_0 - SSR_{1a,k})/[kSSR_{1a,k}] \quad \text{if } k \text{ is even,} \\ F_{1a}(\lambda, k) &= (T - k - 1 - l_T)(SSR_0 - SSR_{1a,k})/[(k+1)SSR_{1a,k}] \quad \text{if } k \text{ is odd,} \end{aligned} \tag{5}$$

$$\begin{aligned} F_{1b}(\lambda, k) &= (T - k - 2 - l_T)(SSR_0 - SSR_{1b,k})/[(k+2)SSR_{1b,k}] \quad \text{if } k \text{ is even,} \\ F_{1b}(\lambda, k) &= (T - k - 1 - l_T)(SSR_0 - SSR_{1b,k})/[(k+1)SSR_{1b,k}] \quad \text{if } k \text{ is odd.} \end{aligned} \tag{6}$$

In (5) and (6), SSR_0 denotes the sum of squared residuals under the null hypothesis, i.e., that obtained from ordinary least squares (OLS) estimation of (4) subject to the restrictions $c_i = 0,\ \alpha_i = 1$ for all i. $SSR_{k,1a}$ denotes the sum of squared residuals obtained from estimating (4) under the restrictions imposed by

Model 1a. Similarly, $SSR_{k,1b}$ denotes the sum of squared residuals obtained from estimating (4) under the restrictions imposed by Model 1b. For some arbitrary small positive number ϵ, we define the set $\Lambda_\epsilon^k = \{\lambda : |\lambda_{i+1} - \lambda_i| \geq \epsilon, \lambda_1 \geq \epsilon, \lambda_k \leq 1-\epsilon\}$. The sup-Wald tests are then defined as sup $F_{1a}(k) = \sup_{\lambda \in \Lambda_\epsilon^k} F_{1a}(\lambda, k)$ and sup $F_{1b}(k) = \sup_{\lambda \in \Lambda_\epsilon^k} F_{1b}(\lambda, k)$. Note that to ensure that the Wald tests are nonnegative, the same number of lags of the first differences of the dependent variable must be used when estimating the models under the null and alternative hypotheses, another reason not to model the changes in the dynamics.

The second type of test is based on the presumption that the nature of persistence in the first regime is unknown, i.e., we do not have any a priori knowledge regarding whether the first regime contains a unit root or not. The tests are given by $W_1(k) = \max[\sup F_{1a}(\lambda, k), \sup F_{1b}(\lambda, k)]$. Finally, in order to accommodate the case with an unknown number of breaks, up to some maximal value A, we consider the statistic $W\max_1 = \max_{1 \leq m \leq A} W_1(m)$. For models 2a and 2b, regression (4) is replaced by

$$\Delta y_t = c_i + b_i t + (\alpha_i - 1) y_{t-1} + \sum_{j=1}^{l_T} \pi_j \Delta y_{t-j} + v_t^*. \tag{7}$$

The Wald tests are defined as

$$\begin{aligned} F_{2a}(\lambda, k) &= (T - 2k - 1 - l_T)(SSR_0^* - SSR_{2a,k})/[(2k) SSR_{2a,k}] \quad \text{if } k \text{ is even,} \\ F_{2a}(\lambda, k) &= (T - 2k - 2 - l_T)(SSR_0^* - SSR_{2a,k})/[(2k+1) SSR_{2a,k}] \quad \text{if } k \text{ is odd,} \end{aligned} \tag{8}$$

$$\begin{aligned} F_{2b}(\lambda, k) &= (T - 2k - 3 - l_T)(SSR_0^* - SSR_{2b,k})/[(2k+2) SSR_{2b,k}] \quad \text{if } k \text{ is even,} \\ F_{2b}(\lambda, k) &= (T - 2k - 2 - l_T)(SSR_0^* - SSR_{2b,k})[(2k+1) SSR_{2b,k}] \quad \text{if } k \text{ is odd.} \end{aligned} \tag{9}$$

In (8) and (9), SSR_0^* denotes the sum of squared residuals under the null hypothesis, i.e., the sum of squared residuals obtained estimating (7) subject to the restrictions $c_i = c,\ b_i = 0,\ \alpha_i = 1$ for all i. Given these tests, the remaining statistics are defined in the same way as for Models 1a and 1b. These are denoted sup $F_{2a}(k)$, sup $F_{2b}(k)$, $W_2(k)$, and $W\max_2$. To compute the sup-Wald test for any particular model, we need to minimize the global sum of squared residuals over the set of permissible break fractions Λ_ϵ^k subject to the restrictions implied by the model. This is accomplished employing the dynamic programming algorithm of Perron and Qu (2006).

3. ASYMPTOTIC RESULTS

We now consider the limiting properties of the proposed statistics. In 3.1 we present the asymptotic distributions of the tests under the null hypothesis that the process is $I(1)$ throughout the sample. The computation of the asymptotic critical

values is discussed in 3.2, and in 3.3 we demonstrate the consistency of the tests under the relevant alternative hypotheses.

3.1. The Null Limiting Distributions

Let $W(.)$ denote a standard Brownian motion on $[0,1]$. Also, let $W^{(j)}(r)$ and $\widetilde{W}^{(j)}(r)$ represent demeaned and detrended Brownian motions, respectively, over $r \in (\lambda_{j-1}, \lambda_j)$ (see the Appendix for detailed expressions). The following theorem states the limit distributions of the tests under the null hypothesis of a unit root. We start with the case where there is no serial correlation and subsequently show that all limit results are valid for the general case.

THEOREM 1. *Suppose that the data are generated by (3) with $u_t = v_t$, where v_t satisfies Assumption A1. Suppose also that the test statistics are constructed based on autoregressions that do not include the lags of first differences of y_t. Then under the null hypothesis H_0: $c_i = 0$, $\alpha_i = 1$ for all i, if k is even, we have*

$$F_{1a}(\lambda, k) \Rightarrow \frac{1}{k}\sum_{i=1}^{k/2}\left[\frac{\left\{\int_{\lambda_{2i-1}}^{\lambda_{2i}} W^{(2i)}(r)dW(r)\right\}^2}{\int_{\lambda_{2i-1}}^{\lambda_{2i}}[W^{(2i)}(r)]^2dr} + \frac{1}{\lambda_{2i}-\lambda_{2i-1}}\{W(\lambda_{2i})-W(\lambda_{2i-1})\}^2\right],$$

$$F_{1b}(\lambda, k) \Rightarrow \frac{1}{k+2}\sum_{i=0}^{k/2}\left[\frac{\left\{\int_{\lambda_{2i}}^{\lambda_{2i+1}} W^{(2i+1)}(r)dW(r)\right\}^2}{\int_{\lambda_{2i}}^{\lambda_{2i+1}}[W^{(2i+1)}(r)]^2dr} + \frac{1}{\lambda_{2i+1}-\lambda_{2i}}\{W(\lambda_{2i+1})-W(\lambda_{2i})\}^2\right].$$

If k is odd,

$$F_{1a}(\lambda, k) \Rightarrow \frac{1}{k+1}\sum_{i=1}^{(k+1)/2}\left[\frac{\left\{\int_{\lambda_{2i-1}}^{\lambda_{2i}} W^{(2i)}(r)dW(r)\right\}^2}{\int_{\lambda_{2i-1}}^{\lambda_{2i}}[W^{(2i)}(r)]^2dr} + \frac{1}{\lambda_{2i}-\lambda_{2i-1}}\{W(\lambda_{2i})-W(\lambda_{2i-1})\}^2\right],$$

$$F_{1b}(\lambda, k) \Rightarrow \frac{1}{k+1}\sum_{i=0}^{(k-1)/2}\left[\frac{\left\{\int_{\lambda_{2i}}^{\lambda_{2i+1}} W^{(2i+1)}(r)dW(r)\right\}^2}{\int_{\lambda_{2i}}^{\lambda_{2i+1}}[W^{(2i+1)}(r)]^2dr} + \frac{1}{\lambda_{2i+1}-\lambda_{2i}}\{W(\lambda_{2i+1})-W(\lambda_{2i})\}^2\right].$$

Under the null hypothesis H_0: $c_i = c$, $b_i = 0$, $\alpha_i = 1$ for all i, if k is even, we have

$$F_{2a}(\lambda,k) \Rightarrow \frac{1}{2k}\left[\begin{array}{l} -\{W(1)\}^2+\sum_{i=0}^{k/2}\left[\frac{1}{\lambda_{2i+1}-\lambda_{2i}}\{W(\lambda_{2i+1})-W(\lambda_{2i})\}^2\right] \\ +\sum_{i=1}^{k/2}\left[\begin{array}{l} \frac{\left\{\int_{\lambda_{2i-1}}^{\lambda_{2i}}\widetilde{W}^{(2i)}(r)dW(r)\right\}^2}{\int_{\lambda_{2i-1}}^{\lambda_{2i}}\left[\widetilde{W}^{(2i)}(r)\right]^2dr}+\frac{1}{\lambda_{2i}-\lambda_{2i-1}}\{W(\lambda_{2i})-W(\lambda_{2i-1})\}^2 \\ +\frac{\left[\int_{\lambda_{2i-1}}^{\lambda_{2i}}\left\{r-(\lambda_{2i}-\lambda_{2i-1})^{-1}\int_{\lambda_{2i-1}}^{\lambda_{2i}}rdr\right\}dW(r)\right]^2}{\int_{\lambda_{2i-1}}^{\lambda_{2i}}\left\{r-(\lambda_{2i}-\lambda_{2i-1})^{-1}\int_{\lambda_{2i-1}}^{\lambda_{2i}}rdr\right\}^2dr} \end{array}\right] \end{array}\right],$$

$$F_{2b}(\lambda,k) \Rightarrow (2k+2)^{-1}\left[\begin{array}{l} -W(1)^2+\sum_{i=1}^{k/2}\left[\frac{1}{\lambda_{2i}-\lambda_{2i-1}}\{W(\lambda_{2i})-W(\lambda_{2i-1})\}^2\right] \\ +\sum_{i=0}^{k/2}\left[\begin{array}{l} \frac{\left\{\int_{\lambda_{2i}}^{\lambda_{2i+1}}\widetilde{W}^{(2i+1)}(r)dW(r)\right\}^2}{\int_{\lambda_{2i}}^{\lambda_{2i+1}}\left[\widetilde{W}^{(2i+1)}(r)\right]^2dr}+\frac{1}{\lambda_{2i+1}-\lambda_{2i}}\{W(\lambda_{2i+1})-W(\lambda_{2i})\}^2 \\ +\frac{\left[\int_{\lambda_{2i}}^{\lambda_{2i+1}}\left\{r-(\lambda_{2i+1}-\lambda_{2i})^{-1}\int_{\lambda_{2i}}^{\lambda_{2i+1}}rdr\right\}dW(r)\right]^2}{\int_{\lambda_{2i}}^{\lambda_{2i+1}}\left\{r-(\lambda_{2i+1}-\lambda_{2i})^{-1}\int_{\lambda_{2i}}^{\lambda_{2i+1}}rdr\right\}^2dr} \end{array}\right] \end{array}\right].$$

If k is odd,

$$F_{2a}(\lambda,k) \Rightarrow \frac{1}{2k+1}\left[\begin{array}{l} -\{W(1)\}^2+\sum_{i=0}^{(k-1)/2}\left[\frac{1}{\lambda_{2i+1}-\lambda_{2i}}\{W(\lambda_{2i+1})-W(\lambda_{2i})\}^2\right] \\ +\sum_{i=1}^{(k+1)/2}\left[\begin{array}{l} \frac{\left\{\int_{\lambda_{2i-1}}^{\lambda_{2i}}\widetilde{W}^{(2i)}(r)dW(r)\right\}^2}{\int_{\lambda_{2i-1}}^{\lambda_{2i}}\left[\widetilde{W}^{(2i)}(r)\right]^2dr}+\frac{1}{\lambda_{2i}-\lambda_{2i-1}}\{W(\lambda_{2i})-W(\lambda_{2i-1})\}^2 \\ +\frac{\left[\int_{\lambda_{2i-1}}^{\lambda_{2i}}\left\{r-(\lambda_{2i}-\lambda_{2i-1})^{-1}\int_{\lambda_{2i-1}}^{\lambda_{2i}}rdr\right\}dW(r)\right]^2}{\int_{\lambda_{2i-1}}^{\lambda_{2i}}\left\{r-(\lambda_{2i}-\lambda_{2i-1})^{-1}\int_{\lambda_{2i-1}}^{\lambda_{2i}}rdr\right\}^2dr} \end{array}\right] \end{array}\right],$$

$$F_{2b}(\lambda,k) \Rightarrow \frac{1}{2k+1}\left[\begin{array}{l} -\{W(1)\}^2+\sum_{i=1}^{(k+1)/2}\left[\frac{1}{\lambda_{2i}-\lambda_{2i-1}}\{W(\lambda_{2i})-W(\lambda_{2i-1})\}^2\right] \\ +\sum_{i=1}^{(k-1)/2}\left[\begin{array}{l} \frac{\left\{\int_{\lambda_{2i}}^{\lambda_{2i+1}}\widetilde{W}^{(2i+1)}(r)dW(r)\right\}^2}{\int_{\lambda_{2i}}^{\lambda_{2i+1}}\left[\widetilde{W}^{(2i+1)}(r)\right]^2dr}+\frac{1}{\lambda_{2i+1}-\lambda_{2i}}\{W(\lambda_{2i+1})-W(\lambda_{2i})\}^2 \\ +\frac{\left[\int_{\lambda_{2i}}^{\lambda_{2i+1}}\left\{r-(\lambda_{2i+1}-\lambda_{2i})^{-1}\int_{\lambda_{2i}}^{\lambda_{2i+1}}rdr\right\}dW(r)\right]^2}{\int_{\lambda_{2i}}^{\lambda_{2i+1}}\left\{r-(\lambda_{2i+1}-\lambda_{2i})^{-1}\int_{\lambda_{2i}}^{\lambda_{2i+1}}rdr\right\}^2dr} \end{array}\right] \end{array}\right].$$

Theorem 1 shows that for all models, the limit distributions of the Wald tests based on a given vector of break fractions $(\lambda_1,\ldots,\lambda_k)$ are pivotal and depend only on functionals of a Wiener process. The limit distributions are different depending on whether the alternative hypothesis specifies that the initial regime has a unit root or is stationary, and are also different for the trending and nontrending cases. The form of the distributions varies according to whether the number of breaks

under the alternative hypothesis is even or odd. With these theoretical results, we can obtain the limit distributions of the proposed tests as a direct consequence of the continuous mapping theorem.

COROLLARY 1. *Denote the limit distribution of the test* $F_j(\lambda, k)$ *by* $F_j^*(\lambda, k)$, $j = 1a, 1b, 2a, 2b$. *Then, under the same null hypothesis as in Theorem 1, we have (a)* $\sup_{\lambda \in \Lambda_\epsilon^k} F_j(\lambda, k) \Rightarrow \sup_{\lambda \in \Lambda_\epsilon^k} F_j^*(\lambda, k)$; *(b)* $W_1(k) \Rightarrow \max[\sup_{\lambda \in \Lambda_\epsilon^k} F_{1a}^*(\lambda, k), \sup_{\lambda \in \Lambda_\epsilon^k} F_{1b}^*(\lambda, k)]$, $W_2(k) \Rightarrow \max[\sup_{\lambda \in \Lambda_\epsilon^k} F_{2a}^*(\lambda, k), \sup_{\lambda \in \Lambda_\epsilon^k} F_{2b}^*(\lambda, k)]$; *(c)* $W\max_1 \Rightarrow \max_{1 \le m \le A}[\max[\sup_{\lambda \in \Lambda_\epsilon^m} F_{1a}^*(\lambda, m), \sup_{\lambda \in \Lambda_\epsilon^m} F_{1b}^*(\lambda, m)]]$, $W\max_2 \Rightarrow \max_{1 \le m \le A}[\max[\sup_{\lambda \in \Lambda_\epsilon^m} F_{2a}^*(\lambda, m), \sup_{\lambda \in \Lambda_\epsilon^m} F_{2b}^*(\lambda, m)]]$.

We now show that the results of Theorem 1 and Corollary 1 remain valid when u_t follows the general linear process (2) with the following assumption about the lag length l_T.

Assumption A3. As $T \to \infty$, the lag length l_T is assumed to satisfy (a) (upper bound condition) $l_T^2 / T \to 0$ and (b) (lower bound condition) $l_T \sum_{j > l_T} \pi_j \to 0$.

Note that the lower bound condition allows for a logarithmic rate of increase for l_T, thereby allowing the use of data-dependent rules such as information criteria to select the lag length (see Ng and Perron, 1995). We now state the result for the general case.

THEOREM 2. *Under Assumptions A1–A3 and the null hypotheses considered in Theorem 1, the test statistics have the same limit distributions as those stated in Theorem 1 and Corollary 1.*

3.2. Asymptotic Critical Values

Given the nonstandard nature of the limit distributions, the critical values are obtained by Monte Carlo simulations. Here again we use Perron and Qu's (2006) dynamic programming algorithm. First, we generate a sample of $T = 500$ observations from a random walk with i.i.d. $N(0, 1)$ errors. We then apply the algorithm to obtain the minimized sum of squared residuals and the corresponding vector of break fractions subject to the relevant restrictions. Next, we simulate a Wiener process using the partial sums of 500 i.i.d. $N(0, 1)$ random variables. Finally, we evaluate the expressions appearing in the limit distributions at the vector of break fractions obtained earlier. This procedure is repeated 5,000 times to obtain the required quantiles of the limit distributions.

Asymptotic critical values are provided in Table 1 with the level of trimming set at $\epsilon = 0.15$. The maximum number of breaks considered is 5. Panel A provides critical values for the nontrending case, while those for the trending case are presented in Panel B. The critical values for Models 1a and 2a are larger than those for Models 1b and 2b, respectively. Note also that the critical values are not monotonically decreasing as k increases. This is due to the fact that the limit

TABLE 1. Asymptotic critical values

	(A) Nontrending case															
	sup $F_{1a}(\lambda, k)$					sup $F_{1b}(\lambda, k)$					$W_1(k)$					
	Number of breaks, k					Number of breaks, k					Number of breaks, k					
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	$W\max_1$
10%	7.94	9.47	7.08	7.04	5.11	5.41	5.64	6.05	5.33	4.84	8.08	9.51	7.28	7.10	5.40	9.86
5%	8.88	10.62	7.73	7.67	5.56	6.39	6.33	6.68	5.84	5.29	8.99	10.62	7.91	7.71	5.79	10.90
2.5%	9.93	11.64	8.33	8.30	5.95	7.28	6.84	7.35	6.31	5.70	10.00	11.64	8.49	8.32	6.21	11.95
1%	11.11	12.72	9.19	9.05	6.46	8.28	7.42	8.04	6.87	6.17	11.21	12.72	9.44	9.05	6.63	13.02

	(B) Trending case															
	sup $F_{2a}(\lambda, k)$					sup $F_{2b}(\lambda, k)$					$W_2(k)$					
	Number of breaks, k					Number of breaks, k					Number of breaks, k					
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	$W\max_2$
10%	7.07	6.90	5.78	5.36	4.27	5.67	5.50	5.24	4.82	4.12	7.28	7.01	5.96	5.48	4.46	7.71
5%	7.84	7.57	6.18	5.77	4.57	6.52	6.02	5.67	5.17	4.39	7.98	7.60	6.36	5.86	4.74	8.43
2.5%	8.49	8.20	6.56	6.14	4.80	7.12	6.43	6.08	5.47	4.69	8.75	8.22	6.77	6.18	4.98	9.18
1%	9.64	9.15	7.23	6.59	5.14	8.07	7.00	6.59	5.82	4.97	9.73	9.18	7.30	6.63	5.34	10.07

distributions are different for the cases with k even or odd. For even or odd values they are, in general, monotonically decreasing as expected.

3.3. Consistency

We now study the properties of the tests under the alternative hypothesis of an unstable persistence parameter. Note, in particular, that under the alternative the dynamics of the process and the variance of the errors are allowed to change along with the level and/or slope of the trend function and the $I(0)/I(1)$ nature of the process. In particular, we demonstrate that in the presence of shifts in persistence of the form considered in this paper, the tests that do not require any information regarding the direction of change are consistent regardless of whether the initial regime is $I(1)$ or $I(0)$, i.e., they reject the null hypothesis with probability one in large samples. We further show that tests that are directed against alternatives in which the initial regime is $I(1)$ [$I(0)$] are inconsistent when the data are generated by alternatives in which the initial regime is $I(0)$ [$I(1)$]. This feature is useful to identify the direction of persistence change. We make the following assumptions.

Assumption A4. The true vector of break fractions, denoted $\lambda^0 = (\lambda_1^0, \ldots, \lambda_m^0)$, is assumed to belong to the set of permissible break fractions, i.e., $\lambda^0 \in \Lambda_\epsilon^m$.

Assumption A3′. As $T \to \infty$, the lag length l_T is assumed to satisfy (a) (upper bound condition) $l_T^6/T \to 0$ and (b) (lower bound condition) $l_T \sum_{j>l_T} \pi_j \to 0$.

Assumption A4 is not very restrictive given that in practice, ϵ can be chosen to be small. Assumption A3′ strengthens the upper bound condition in Assumption A3 to account for the fact that a subset of the regressors in the $I(0)$ regimes (those corresponding to the lagged first differences) is over-differenced. We can then state the following theorem regarding the consistency of the tests under the relevant alternative hypotheses given by Model (2), which allow for changes in the I(1)/I(0) nature of the data as well as changes in the trend function, the dynamics of the process, and the variance of the errors.

THEOREM 3. *Suppose that the data are generated under the alternative hypothesis represented by Model j ($j = 1a, 1b, 2a,$ or $2b$) with m breaks in persistence. Then, under Assumptions A1–A2, A3′, and A4, (a) the tests* $\sup_{\lambda \in \Lambda_\epsilon^m} F_j(\lambda, m)$ *and* $W\max_1$ *are consistent; (b) if the data are generated by Models 1a or 1b, the tests* $W_1(m)$ *and* $W\max_1$ *are consistent, while if the data are generated by Models 2a or 2b, the tests* $W_2(m)$ *and* $W\max_2$ *are consistent; and (c) the test* $\sup_{\lambda \in \Lambda_\epsilon^1} F_{j'}(\lambda, m)$ *is inconsistent, where* $(j, j') = (1a, 1b), (1b, 1a)$.

Parts (a) and (b) of Theorem 3 state that the tests that are directed against the alternatives that represent the true DGP as well as those that do not require any information regarding the direction of change are both consistent. Part (c) states that for models with nontrending data, tests that are directed against the "wrong" alternative are inconsistent, i.e., $O_p(1)$. In Section 6 we show through simulations that these tests have empirical power reasonably close to their nominal size,

thereby enabling the applied researcher to infer the direction of shift from the test outcomes.

4. HYBRID TESTING PROCEDURES

One aspect of the test statistics introduced in Section 2 is that they will reject the null with probability one in large samples even if the process is stable $I(0)$ or one that involves changes in the value of the autoregressive parameter such that the process is still $I(0)$ in each regime, i.e., $I(0)$ preserving changes. In practice, the researcher may be interested in reliably interpreting the test outcome as one emanating from a switch between an $I(1)$ and an $I(0)$ regime. To accommodate such an interpretation, we propose hybrid testing procedures that entail the joint application of our tests with the Bai and Perron (1998) structural change tests designed for a stationary framework as well as the unit root tests proposed by Ng and Perron (2001) with the modification of Perron and Qu (2007) to select the lag length. The number of breaks m is assumed to be known.

The first hybrid procedure is designed to test the null hypothesis that the process is stable $I(1)$ or stable $I(0)$. To this end, we define $BP(m)$ as the Bai-Perron (1998) partial structural change test that jointly tests the stability of the intercept and the autoregressive parameter in (4) while holding fixed the coefficients on the lagged first differences. This test has the correct asymptotic size when the process is constant $I(0)$. We therefore employ the following decision rule labeled the D_m test: "Reject the null if both $W_1(m)$ and $BP(m)$ reject." If the significance level κ is employed for both tests, the asymptotic size of D_m cannot exceed κ, regardless of whether the process is $I(1)$ or $I(0)$ throughout. Further, since $BP(m)$ and $W_1(m)$ are both consistent against processes that involve a switch between an $I(1)$ and an $I(0)$ regime, D_m has unit asymptotic power against such alternatives. Here, the assumption of a known number of breaks can be relaxed using the $W\max_1$ test and the UDmax version of the BP test.

The second hybrid procedure allows the null hypothesis to include the case of $I(0)$ preserving changes in addition to the stable $I(1)/I(0)$ cases. This procedure is useful if the researcher seeks to distinguish between $I(0)$ preserving changes and those that involve at least one switch between an $I(1)$ and an $I(0)$ regime. To facilitate this distinction, we note that with $I(0)$ preserving changes, a unit root test applied on the regime with the largest estimated autoregressive root will reject the null asymptotically, while if an $I(1)$ segment is present, such a test will reject only with probability equal to the nominal significance level in large samples. We therefore recommend using the D_m procedure in conjunction with one of the M^{GLS} tests proposed by Ng and Perron (2001) with the modification of Perron and Qu (2007) to select the lag length, given that these tests avoid the power reversal problem for nonlocal stationary alternatives while maintaining empirical size close to nominal size. The former feature ensures that our hybrid procedure is well sized, while the latter ensures little loss in power. We therefore propose joint application of the D_m procedure and the particular M^{GLS} test on

the regime with the largest estimated autoregressive root, where the regimes are identified by minimizing the unrestricted sum of squared residuals. Specifically, the decision rule, labeled the J_m test is: "Reject the null if D_m rejects and M^{GLS} does not reject." If a significance level κ is used for each of the tests in D_m as well as for M^{GLS}, the asymptotic size of J_m is bounded by κ, while for persistence changes that involve switches between $I(1)$ and $I(0)$ regimes, its asymptotic power is $(1-\kappa)$. The finite sample performance of D_m and J_m will be investigated through simulations in Section 6. Using the J_m test, in large samples one can obtain a complete correct classification into $I(0)$ or $I(1)$ throughout, $I(0)$ changes or $I(1)/I(0)$ changes by letting the size of each test go to zero at a suitable rate.

5. ESTIMATORS FOR THE BREAK DATES

Following evidence against the null hypothesis, it is desirable to determine the location of the break dates. To this end, we propose estimating the break date estimators from global minimization of the sum of squared residuals under the relevant alternative hypothesis. For a model with k breaks, the estimated break dates are thus obtained as $(\tilde{T}_1, \ldots, \tilde{T}_k) = \arg\min_{T_1,\ldots,T_k} SSR_{j,k}(T_1, \ldots, T_k)$ where $SSR_{j,k}(T_1, \ldots, T_k)$ is the sum of squared residuals for Model j ($j = 1a, 1b, 2a, 2b$) evaluated at the partition $\{T_1, \ldots, T_k\}$.[1] When estimating the break dates, we allow the coefficients on the lagged first differences to vary across regimes. The number of lags is also allowed to be regime dependent. The computation of the sum of squared residuals is similar to that discussed in Section 2 except that the cross-regime restrictions on the coefficients governing the short-run dynamics are replaced by within-regime restrictions depending on the number of lags included in a specific regime. The asymptotic properties of these estimators, including their consistency, rate of convergence, and limit distribution, are investigated in a companion paper (Kejriwal and Perron, 2012). Simulations (not reported here) show that the estimators perform very well in small samples in terms of bias and root mean squared error.

6. SIMULATION EXPERIMENTS

In this section we conduct simulation experiments to assess the finite sample performance of the proposed tests as well as to provide a comparison with the tests proposed in Harvey et al. (2006) and Leybourne et al. (2007b). We report results only for the nontrending case, given that qualitatively similar results were obtained for the trending case. In particular, we consider the statistics $W_1(1)$, $W_1(2)$, D_1, D_2, J_1, and J_2. Results for the $W\max_1$ test were found to be similar to those for the W_1 test based on the true number of breaks and hence not reported. The Harvey et al. class of tests is designed to detect a single persistence break and is based on partial sums of the demeaned or detrended data. They recommend using the so-called "m min-modified" and "Sm min-modified" tests

based on extensive simulation experiments. These tests differ in the method used to compute the critical values. Given their similar finite sample performance, we only report results for the "m min-modified" tests. Further, we present results only for the test based on the mean-functional, denoted H, since this was found to outperform the maximum and exponential versions in most of our experiments (as in Harvey et al.) while producing very similar results in others. The Leybourne et al. (2007b) tests allows for multiple changes and are based on a doubly recursive application of the unit root statistic using the local GLS detrending methodology developed in Elliott et al. (1996). More specifically, they propose the test statistic $M = \inf_{\lambda\in(0,1)} \inf_{\tau\in(\lambda,1]} DF_G(\lambda,\tau)$, where $DF_G(\lambda,\tau)$ is the local GLS detrended ADF unit root t-statistic that uses the observations between λT and τT. Both the H and M tests allow the process to be stable $I(1)$ or stable $I(0)$ under the null hypothesis.

We consider cases where the data generating processes (DGPs) involve no break (size), as well as some involving one and two breaks (power). The sample sizes used are $T = 150$, 240. The lag length in the autoregression for our proposed procedures is selected using the Bayesian information criterion (BIC) with the maximum number of lags allowed set at 10. We first obtain the number of lags based on the estimation of the alternative model and then use this number in the estimation of the null model. For the M test we used the Gauss posted by Leybourne et al. (2007b) posted on the Studies in Nonlinear Dynamics and Econometrics website, so that the lag length selection is based on the sequential approach of Ng and Perron (1995), with a maximal lag order of four and a 10% significance level for the t-test on the highest lag. In order to account for the stable $I(0)$ possibility under the null, the rejection frequency of the M-procedure is computed as the proportion of Monte Carlo replications in which the M test rejects, and the corresponding partition selected by the test does not correspond to the full sample. Finally, to compute J_m, we use the MZ_α^{GLS} unit root test of Ng and Perron (2001) with the modification of Perron and Qu (2007) to select the lag length with a maximum of five lags.[2]

In all experiments, $\{e_t\}$ denotes a sequence of i.i.d. $N(0,1)$ variables. The errors $\{u_t\}$ are generated by the autoregressive moving average (ARMA) process $u_t = \rho u_{t-1} + e_t + \theta e_{t-1},\ \ u_0 = 0$. We present results for the following combinations of values of the autoregressive parameter (ρ) and the moving average parameter (θ): (a) $\rho = \theta = 0$; (b) $\rho = 0.5,\ \theta = 0$; (c) $\rho = 0,\ \theta = 0.5$; (d) $\rho = 0,\ \theta = -0.5$; (e) $\rho = 0.3,\ \theta = 0.5$; (f) $\rho = 0.3,\ \theta = -0.5$. The nominal size for all tests is set at 5%. All experiments are based on 1,000 replications.

6.1. The Empirical Size of the Tests

In order to assess the empirical size of the tests, the DGP considered is DGP-0: $y_t = \alpha y_{t-1} + u_t,\ \ y_0 = 0$. The results are presented in Table 2a for $\alpha = 1$ and Table 2b for $\alpha < 1$. For the latter, we report results only for $\rho = \theta = 0$, although the full set of results is available upon request. Consider first the unit root case.

TABLE 2a. Empirical size when the process is constant $I(1)$ (DGP-0, Nominal size = 5%)

(ρ,θ)	(0, 0)		(.5, 0)		(0, .5)		(0, −.5)		(.3, .5)		(.3, −.5)	
Test\T	150	240	150	240	150	240	150	240	150	240	150	240
$W_1(1)$	.05	.06	.07	.06	.08	.08	.13	.10	.07	.07	.15	.13
$W_1(2)$	.04	.05	.03	.05	.05	.07	.12	.07	.06	.06	.10	.13
D_1	.05	.05	.05	.04	.07	.06	.06	.05	.07	.04	.09	.08
D_2	.04	.05	.03	.04	.05	.05	.11	.06	.04	.05	.08	.09
J_1	.05	.05	.04	.04	.07	.05	.06	.05	.07	.04	.09	.08
J_2	.03	.05	.03	.04	.04	.05	.09	.05	.04	.05	.08	.08
M	.17	.13	.15	.11	.23	.17	.90	.83	.25	.17	.45	.41
H	.05	.05	.02	.02	.03	.03	.21	.18	.02	.02	.10	.10

TABLE 2b. Empirical size when the process is constant $I(0)$ (DGP-0, $\rho = \theta = 0$, Nominal size = 5%)

	$\alpha = .5$		$\alpha = .6$		$\alpha = .7$		$\alpha = .8$		$\alpha = .9$	
Test\T	150	240	150	240	150	240	150	240	150	240
$W_1(1)$	.99	1.0	.99	1.0	.87	1.0	.46	.94	.12	.27
$W_1(2)$	.93	1.0	.75	.99	.39	.87	.12	.39	.06	.10
D_1	.04	.04	.05	.05	.06	.06	.05	.06	.05	.04
D_2	.04	.02	.05	.04	.06	.05	.04	.04	.05	.04
J_1	.01	.00	.01	.00	.01	.01	.02	.01	.03	.02
J_2	.01	.01	.02	.00	.02	.01	.01	.01	.02	.02
M	.93	.93	.92	.94	.85	.92	.48	.91	.14	.39
H	.04	.06	.04	.06	.04	.02	.03	.04	.02	.03

When the errors do not contain a negative MA component, all the proposed statistics are adequately sized with the null rejection probabilities never exceeding 10% for either sample size. With a negative MA component, the $W_1(1)$ and $W_1(2)$ tests suffer from important size distortions, which remain prominent even for $T = 240$. As with standard unit root tests, these size problems arise from the downward bias in the persistence parameter estimates under the null hypothesis of a unit root. A useful feature of the D_m and J_m tests is that they remain adequately sized across all values of (ρ, θ). The M test, on the other hand, is seriously oversized irrespective of the nature and extent of serial correlation in the errors. The rejection probability is at least 15% for $T = 150$ and never falls below 10%, even for $T = 240$. These distortions are especially severe with negative MA errors. For instance, with $\rho = 0$, $\theta = -0.5$, and $T = 240$, the empirical size of the M test is 83%. Since the M test is based on the application of unit root tests to data subsamples, the bias in the sum of the autoregressive coefficient estimates is exacerbated, which

in turn contributes to the poor finite sample performance of the test under the null hypothesis. The H test is accurate except when a negative MA component is present. When $\alpha < 1$, the $W_1(1)$, $W_1(2)$, and M tests all overreject the null substantially. These spurious rejections decline as α increases but remain nonnegligible for $\alpha \leq 0.8$. In contrast, the H, D_m, and J_m tests maintain empirical size very close to nominal size for all stationary values of α and both sample sizes.

6.2. The Case with One Break

We now consider the power of the tests with a single break and the following DGPs:

	For $t \leq [T\lambda_1^0]$	For $t \geq [T\lambda_1^0]+1$
DGP-1	$y_t = y_{t-1} + u_t$	$y_t = \alpha y_{t-1} + u_t$
DGP-2	$y_t = \alpha y_{t-1} + u_t$	$y_t = y_{t-1} + u_t$
DGP-3	$y_t = y_{t-1} + \pi_1 \Delta y_{t-1} + e_t$	$y_t = \alpha y_{t-1} + \pi_2 \Delta y_{t-1} + e_t$
DGP-4	$y_t = \alpha y_{t-1} + \pi_1 \Delta y_{t-1} + e_t$	$y_t = y_{t-1} + \pi_2 \Delta y_{t-1} + e_t$
DGP-5	$y_t = y_{t-1} + u_t$	$y_t - y_{[T\lambda_1^0]} = \alpha(y_{t-1} - y_{[T\lambda_1^0]}) + u_t$

DGP-1 and DGP-2 are processes involving a shift in the persistence parameter but no change in the short-run dynamics. DGP-3 and DGP-4 allow for the short-run dynamics to simultaneously change as well. We also examine the power of the tests when the persistence parameter is unity but the short-run dynamics change across regimes, i.e., the data are generated by DGP-3 (or DGP-4) with $\alpha = 1$ but $\pi_1 \neq \pi_2$. DGP-5 is a variant of DGP-1 that is considered in Leybourne et al. (2007b). Such a process is designed to avoid sharp jumps to zero at the break point between the $I(1)$ and $I(0)$ regimes and ensures a joining up of these regimes. We consider three values for the location of the break: $\lambda_1^0 = 0.3, 0.5, 0.7$. We present results for three values of the autoregressive parameter: $\alpha = 0.5, 0.7, 0.8$. Given the extent of size distortions, the powers of $W_1(1)$, $W_1(2)$, and M tests are all size-adjusted. The results are presented in Table 3. We only report results for $\lambda_1^0 = 0.5$ and $T = 240$ (more results are available in the working paper version, including those for $T = 150$). Power does vary with the location of the break: As expected, it is higher when the break occurs early ($\lambda_1^0 = 0.3$) and lower when it occurs late ($\lambda_1^0 = 0.7$) for DGP-1,3,5 and vice versa for DGP-2,4. This is due to the fact that the longer the $I(0)$ segment, the further away the series is from a pure unit root process. Relative to the H test and the proposed tests, however, the M test is much more sensitive to break location. Otherwise, the qualitative features are similar.[3]

Panel (A) of Table 3 provides results for DGP-1. As expected, the power of all the tests decreases as α increases. Power is also lower with serially correlated errors compared to the i.i.d. case, except when the errors contain a negative MA component. The tests are thus subject to a clear size-power trade-off in this latter case. The loss in power from introducing an autoregressive component in the errors is especially significant for the M test, e.g., power falls from 79% to

TABLE 3. Empirical power with one break ($\lambda_1^0 = 0.5$); $T = 240$

	$\alpha = 0.5$					$\alpha = 0.7$					$\alpha = 0.8$				
	W_1	D_1	J_1	M	H	W_1	D_1	J_1	M	H	W_1	D_1	J_1	M	H
(ρ, θ)							(A) DGP-1								
(0, 0)	1.0	1.0	.97	1.0	.93	.99	.94	.90	.79	.70	.84	.79	.76	.37	.38
(.5, 0)	1.0	1.0	.95	.79	.81	.95	.93	.88	.45	.48	.76	.76	.73	.26	.21
(0, .5)	1.0	1.0	.95	.92	.90	.95	.91	.87	.55	.62	.74	.74	.72	.29	.30
(0, −.5)	.99	.99	.92	.99	.97	.93	.91	.85	.87	.87	.78	.76	.73	.48	.66
(.3, .5)	1.0	.99	.99	.79	.85	.94	.91	.91	.47	.52	.73	.72	.72	.26	.24
(.3, −.5)	1.0	1.0	.95	1.0	.96	.97	.95	.86	.88	.81	.83	.83	.69	.47	.52
(ρ, θ)							(B) DGP-2								
(0, 0)	1.0	.99	.95	1.0	.96	.82	.75	.72	.88	.82	.37	.30	.28	.50	.60
(.5, 0)	.93	.93	.89	.85	.90	.52	.54	.51	.57	.68	.24	.23	.20	.34	.43
(0, .5)	.95	.96	.92	.94	.94	.60	.60	.57	.67	.76	.30	.29	.26	.39	.54
(0, −.5)	.93	.94	.88	1.0	.98	.76	.77	.73	.90	.90	.48	.54	.50	.55	.78
(.3, .5)	.91	.91	.87	.84	.92	.50	.50	.46	.55	.71	.24	.22	.18	.35	.47
(.3, −.5)	.98	.99	.93	1.0	.98	.85	.88	.83	.91	.86	.57	.65	.60	.59	.69
(π_1, π_2)							(C) DGP-3								
(0. −.2)	1.0	1.0	.96	.94	.94	.99	.98	.91	.73	.66	.90	.86	.80	.51	.37
(−.3, − .5)	1.0	1.0	.96	.79	.91	.97	.95	.90	.48	.61	.78	.76	.70	.35	.35
(π_1, π_2)							(D) DGP-4								
(0. −.2)	.91	.90	.89	1.0	.95	.54	.53	.64	.90	.77	.31	.31	.26	.57	.55
(−.3, −.5)	.74	.71	.68	.79	.93	.21	.20	.21	.45	.69	.09	.07	.06	.26	.48
(ρ, θ)							(E) DGP-5								
(0, 0)	1.0	1.0	.94	1.0	.96	.98	.93	.87	.91	.84	.81	.71	.67	.53	.59
(.5, 0)	1.0	.99	.93	.88	.91	.91	.85	.78	.59	.70	.66	.63	.59	.36	.42
(0, .5)	1.0	1.0	.93	.94	.95	.89	.85	.78	.70	.79	.65	.61	.58	.41	.53
(0, −.5)	.98	.98	.89	1.0	.98	.89	.91	.83	.92	.91	.74	.71	.65	.60	.78
(.3, .5)	.99	.99	.92	.85	.93	.87	.78	.74	.58	.73	.61	.56	.52	.35	.46
(.3, −.5)	1.0	1.0	.92	1.0	.97	.94	.94	.85	.95	.89	.79	.78	.71	.64	.68

Note: In all cases, W_1 stands for the statistic $W_1(1)$.

45% as ρ increases from 0 to 0.5 when $\alpha = 0.7$. In comparison, the power of the proposed tests is much more robust to the extent of error serial correlation. Moreover, there is only a mild loss in power from using the D_1 and J_1 tests compared to the less robust $W_1(1)$. This property is important in applications where the researcher does not want to take a stand on the nature of the process under the null

TABLE 4. Empirical power (DGP-3, $\alpha = 1,\ \rho = \theta = 0,\ \lambda_1^0 = 0.5$, nominal size = 5%); $T = 240$

	$W_1(1)$	$W_1(2)$	D_1	D_2	J_1	J_2	M	H
$\pi_1 = 0,\ \pi_2 = -.2$	.09	.08	.08	.05	.07	.05	.19	.06
$\pi_1 = -.3,\ \pi_2 = -.5$	.08	.07	.07	.04	.07	.04	.30	.09

hypothesis. The use of the proposed tests appears to be advantageous relative to the M and H tests in terms of detecting an $I(1)$-$I(0)$ shift.

The results for DGP-2 are reported in Panel (B) of Table 3. The H test dominates in this case, while the rejection probabilities of the M and $W_1(1)$ tests are broadly similar, except when the errors contain a pure negative MA component, in which case the M test rejects the null more often. Again, the D_1 and J_1 tests retain power close to $W_1(1)$. For DGP-3 and DGP-4, the results are presented in Panels (C) and (D). Again, the proposed tests are generally superior to the others for DGP-3, while the H test performs favorably for DGP-4. Finally, the rejection frequencies for DGP-5 reported in Panel (E) indicate that, relative to DGP-1, the M and H tests now have higher power, while the proposed tests have lower power, though the latter tests still exhibit the highest power except when the errors are driven by a negative MA component, where the H test rejects the null more often.

It is of interest to assess the power of the tests when the short-run dynamics are allowed to change while the process remains $I(1)$ throughout. Table 4 reports the rejection frequencies for DGP-3 when $\alpha = 1,\ \pi_1 \neq \pi_2,\ \lambda_1^0 = 0.5$, and $\rho = \theta = 0$. In contrast to the other tests, the rejection frequencies of the D_m and J_m tests do not exceed 10% in any of the cases. The rejection frequencies of the M test increase sharply relative to the case where the DGP does not involve a shift in short-run dynamics. Hence, our tests are more robust to potential changes in the dynamics only so that a rejection by our tests can be more reliably interpreted as coming from a change in persistence as opposed to a shift in short-run dynamics.

6.3. The Case With Two Breaks

With two breaks in persistence, we report results for locations of the breaks at $(\lambda_1^0, \lambda_2^0) = (0.3, 0.6)$, while similar results were obtained using the coordinates $(0.3, 0.7)$ and $(0.4, 0.7)$. The DGPs considered are:

	For $t \leq [T\lambda_1^0]$	For $[T\lambda_1^0]+1 \leq t \leq [T\lambda_2^0]$	For $t \geq [T\lambda_2^0]+1$
DGP-6	$y_t = y_{t-1} + u_t$	$y_t = \alpha y_{t-1} + u_t$	$y_t = y_{t-1} + u_t$
DGP-7	$y_t = \alpha y_{t-1} + u_t$	$y_t = y_{t-1} + u_t$	$y_t = \alpha y_{t-1} + u_t$
DGP-8	$y_t = y_{t-1} + \pi_1 \Delta y_{t-1} + e_t$	$y_t = \alpha y_{t-1} + \pi_2 \Delta y_{t-1} + e_t$	$y_t = y_{t-1} + e_t$
DGP-9	$y_t = \alpha y_{t-1} + \pi_1 \Delta y_{t-1} + e_t$	$y_t = y_{t-1} + \pi_2 \Delta y_{t-1} + e_t$	$y_t = \alpha y_{t-1} + e_t$
DGP-10	$y_t = y_{t-1} + u_t$	$y_t - y_{[T\lambda_1^0]} = \alpha(y_{t-1} - y_{[T\lambda_1^0]}) + u_t$	$y_t = y_{t-1} + u_t$

TABLE 5. Empirical power with two breaks ($\lambda_1^0 = 0.3,\ \lambda_2^0 = 0.6$); $T = 240$

	$\alpha = 0.5$					$\alpha = 0.7$					$\alpha = 0.8$				
	W_2	D_2	J_2	M	H	W_2	D_2	J_2	M	H	W_2	D_2	J_2	M	H
(ρ, θ)							(A) DGP-6								
(0, 0)	.96	.94	.94	.90	.08	.71	.63	.63	.39	.06	.41	.38	.38	.20	.07
(.5, 0)	.92	.91	.87	.43	.05	.68	.64	.61	.22	.03	.46	.43	.42	.14	.04
(0, .5)	.91	.92	.90	.61	.07	.63	.61	.59	.28	.04	.40	.42	.41	.15	.06
(0, −.5)	.83	.85	.83	.92	.19	.54	.57	.56	.49	.15	.32	.40	.39	.22	.16
(.3, .5)	.89	.89	.87	.48	.06	.63	.60	.58	.26	.03	.42	.41	.40	.14	.05
(.3, −.5)	.95	.96	.94	.94	.12	.71	.70	.69	.48	.09	.43	.50	.49	.20	.10
(ρ, θ)							(B) DGP-7								
(0, 0)	1.0	.96	.91	1.0	.84	.81	.80	.77	.75	.55	.50	.49	.51	.40	.28
(.5, 0)	.94	.94	.89	.73	.68	.71	.70	.67	.45	.33	.51	.50	.49	.28	.15
(0, .5)	.90	.92	.86	.91	.78	.64	.68	.63	.57	.47	.43	.49	.45	.30	.22
(0, −.5)	.97	.97	.73	1.0	.89	.78	.75	.60	.82	.71	.48	.45	.42	.45	.54
(.3, .5)	.87	.94	.84	.78	.72	.64	.73	.63	.50	.38	.43	.53	.43	.29	.18
(.3, −.5)	.99	.99	.86	1.0	.87	.89	.90	.73	.82	.63	.58	.78	.55	.46	.41
(π_1, π_2)							(C) DGP-8								
(0, −.2)	1.0	.98	.96	.74	.07	.88	.81	.81	.49	.06	.66	.57	.56	.29	.06
(−.3, −.5)	1.0	1.0	.96	.46	.10	.85	.84	.84	.32	.10	.64	.63	.58	.23	.09
(π_1, π_2)							(D) DGP-9								
(0, −.2)	.98	.91	.87	1.0	.76	.75	.70	.66	.79	.48	.42	.44	.41	.49	.24
(−.3, −.5)	.75	.68	.61	.99	.69	.50	.42	.40	.84	.42	.31	.27	.26	.54	.22
(ρ, θ)							(E) DGP-10								
(0, 0)	.95	.95	.89	.96	.07	.58	.55	.51	.56	.06	.26	.28	.26	.30	.06
(.5, 0)	.77	.78	.74	.55	.04	.41	.41	.41	.32	.03	.21	.25	.24	.21	.04
(0, .5)	.81	.87	.81	.73	.05	.42	.49	.44	.40	.04	.22	.30	.28	.23	.05
(0, −.5)	.74	.81	.76	.97	.17	.45	.54	.49	.65	.16	.26	.41	.28	.33	.17
(.3, .5)	.68	.77	.70	.59	.04	.34	.39	.37	.36	.04	.17	.25	.22	.23	.04
(.3, −.5)	.92	.95	.90	.98	.11	.65	.75	.65	.65	.10	.41	.60	.41	.32	.10

Note: In all cases, W_2 stands for the statistic $W_1(2)$.

The results are presented in Table 5. First, consider the power of the various tests when the data are generated by DGP-6 and DGP-7 (Panels (A)–(B)). For DGP-6, the proposed tests are clearly preferred to the M and H tests, with the H test exhibiting very little power even with a large sample size. In unreported

simulations, we found that the power of all tests (except the H test) is higher for $\lambda_1^0 = 0.3, \lambda_2^0 = 0.7$ compared to the other two location pairs. This is not unexpected since power should depend positively on the length of the $I(0)$ segment in the data. For DGP-7, our tests again outperform the others except in the case with pure negative MA errors, although the discrepancy in this latter case is not substantial. The performance of the M test was again found to be quite sensitive to the location of the breaks for both DGP-6 and DGP-7. Interestingly, the H test has much higher power against DGP-7 relative to DGP-6, which when combined with the results in Table 3 indicates that this test is more effective at detecting deviations from the null when the initial regime is $I(0)$. For DGP-9 (Panel (D)), the rejection frequencies of the tests are close to those in the absence of regime-specific short-run dynamics. Surprisingly though, in the case of DGP-8 (Panel (C)), the proposed tests are more powerful relative to the case with no change in the short-run dynamics, even though the tests are based against the alternative that these dynamics remain unchanged across regimes. Finally, the conclusions based on power results for DGP-10 (Panel (E)) are qualitatively similar to those discussed for DGP-5.

6.4. Identifying the Initial Regime

As discussed in Section 3.3, the proposed tests can be used to distinguish between processes with an initial $I(1)$ regime and those with an initial $I(0)$ regime. Here we evaluate the empirical power of single and double break tests that are directed against the incorrect alternative, for instance, when the data involve an $I(1)$–$I(0)$ change but the researcher applies a test directed against the $I(0)$–$I(1)$ alternative. To save space, we only present results for DGPs 1, 2, 6, and 7 for the case $\rho = \theta = 0$. For the single break case, the results are reported in Panels (A) and (B) of Table 6, while those for two breaks are reported in Panels (C) and (D) of the same table. The results indicate that when the initial regime is $I(0)$ in the true DGP (DGPs 2 and 7), the rejection frequencies are well controlled irrespective of the number and locations of breaks as well as the sample size. Even when the initial regime is $I(1)$, the rejection frequencies in most cases are within 10%; the exceptions are when the break occurs early in the single change case and when $(\lambda_1, \lambda_2) = (0.4, 0.7)$ in the two breaks case. An important feature of these results is that the rejection frequencies do not display any tendency to increase with the sample size, thereby confirming that the tests are indeed inconsistent when directed against incorrect alternatives.

6.5. Summary and Practical Recommendations

In summary, the simulation results reveal that the D_m, J_m, and H tests have much better size control in finite samples relative to the M test. The latter test has a substantial probability of overrejection regardless of the degree of serial correlation in the errors and whether the process is $I(1)$ or $I(0)$. In most cases the suggested

TABLE 6. Empirical power against incorrect alternatives, $\rho = \theta = 0$

	$\alpha = 0.5$		$\alpha = 0.7$		$\alpha = 0.8$		$\alpha = 0.9$	
$\lambda^0 \setminus T$	150	240	150	240	150	240	150	240
	(A) DGP-1, sup $F_{1b}(1)$							
0.3	.17	.17	.09	.13	.04	.07	.01	.01
0.5	.02	.02	.01	.01	.01	.01	.00	.00
0.7	.00	.00	.00	.00	.00	.00	.00	.00
	(B) DGP-2, sup $F_{1a}(1)$							
0.3	.01	.01	.01	.01	.01	.01	.01	.01
0.5	.02	.01	.02	.01	.02	.01	.01	.01
0.7	.09	.09	.04	.06	.02	.03	.01	.02
	(C) DGP-6, sup $F_{1b}(2)$							
(0.3,0.6)	.03	.03	.01	.01	.01	.01	.01	.01
(0.3,0.7)	.05	.06	.02	.02	.01	.01	.01	.01
(0.4,0.7)	.15	.18	.09	.13	.06	.09	.02	.03
	(D) DGP-7, sup $F_{1a}(2)$							
(0.3,0.6)	.09	.12	.04	.05	.03	.04	.01	.02
(0.3,0.7)	.04	.03	.02	.02	.02	.02	.01	.02
(0.4,0.7)	.04	.04	.01	.02	.01	.01	.01	.01

statistics are also shown to have superior performance in terms of rejecting the null when the alternatives of interest drive the DGP. The power performance of the H test is quite sensitive to whether the initial regime is $I(1)$ or $I(0)$, with power being much higher in the latter case. This feature appears especially relevant in the presence of multiple breaks, in which case the H test has very little power when the initial regime is $I(1)$. Hence, combining the size and power results in the previous section, the D_m and J_m tests appear to constitute a very useful addition to the existing battery of procedures designed to detect shifts in persistence.

In practice, the researcher may be interested not only in determining if the process is governed by a stable persistence parameter, but also in distinguishing between shifts that preserve the $I(0)$ nature of the process in each segment and those that are characterized by switches between $I(1)$ and $I(0)$ regimes. In what follows we show that the use of the J_m test allows one to successfully discriminate between these possibilities, while existing procedures are not suited for the same. In particular, we consider the following DGP-S: $y_t = e_t$ if $t \leq [T\lambda_1^0]$ and $y_t = \alpha y_{t-1} + e_t$ if $t \geq [T\lambda_1^0] + 1$, where $y_0 = 0$, $\lambda_1^0 = 0.5$ and $e_t \sim iid\ N(0,1)$. The rejection probabilities of the J_1, M, and H tests for a range of stationary values of α are reported in Table 7. The results show that the M test almost always

TABLE 7. Null rejection probabilities for an $I(0)$-$I(0)$ change (DGP-S, $\alpha_1 = 0$, $\alpha_2 = \alpha$, $\rho = \theta = 0$, $\lambda_1^0 = 0.5$)

	$\alpha = .5$		$\alpha = .6$		$\alpha = .7$		$\alpha = .8$		$\alpha = .9$	
Test \T	150	240	150	240	150	240	150	240	150	240
J_1	.06	.07	.15	.06	.15	.07	.21	.08	.55	.23
M	.96	.97	.98	.99	.99	1.0	1.0	1.0	1.0	1.0
H	.22	.24	.27	.35	.44	.51	.68	.75	.87	.95

rejects, regardless of the sample size and the break magnitude. The H and J_1 tests are much more sensitive to the magnitude of the change, rejecting the null more frequently as the break becomes larger. Among the latter two tests, however, the J_1 test is much more immune to the value of α, the likelihood of rejection being substantial only when $\alpha = 0.9$ and $T = 150$. This experiment thus clearly illustrates the usefulness of the recommended tests in identifying the nature of the persistence shifts responsible for instabilities in the process generating the data.

It remains to discuss how to disentangle a rejection of the proposed statistics as coming from a change in persistence and not only a change in the trend function. In the trendless case where the process is $I(0)$ with pure level shifts, the use of the J_m tests again provides a reliable safeguard, since its rejection frequencies are controlled owing to the fact that the unit root test on the regime with the largest estimated autoregressive unit root rejects with probability one in large samples (given the consistency of the estimated breakpoints). Consider next the case where the process is $I(0)$ with breaks in the slope of the trend function. Then, our tests will have power but so would unit root tests allowing for a change in the trend function; see Kim and Perron (2009). If there are changes both in persistence and in the slope of the trend, then the latter would not reject (see Kim, 2000). So our test can be used in conjunction with those of Kim and Perron to make sure that a change in persistence is indeed present and not only a change in the trend function. Finally, consider the case in which the process is $I(1)$ across two segments but with a change in trend. The following procedure can be used to distinguish such a process from a persistence change process. We first detrend the data using a regression of the data on a time trend and a slope dummy (where the break date is chosen by minimizing the sum of squared residuals). We then apply our persistence change tests to the detrended data. The problem is that the limits of the resulting statistics under the unit root null depend on the true trend break date. But we can use the critical values corresponding to Models 2a or 2b (as the case may be) as a benchmark. If there is only a pure trend break, these tests should not reject the null, while if there is an accompanying change in persistence, one of the tests (for Model 2a or 2b) would reject. To examine the finite sample performance of the detrended test statistics in the single break case, we consider

TABLE 8. Empirical size and power of $W_1(1)$ and $W_2(1)$ (DGP-T, $\mu_0 = \beta_0 = 0$, $\tau^0 = 0.5$) (For J_1: $\beta_1 = 0,\ \mu_1 = 5$ and for $W_2(1)$: $\mu_1 = 0,\ \beta_1 = 0.5$)

	$T=150$		$T=240$		$T=150$		$T=240$		$T=150$		$T=240$	
	J_1	W_2	J_1	W_2	J_1	W_2	J_1	W_2	J_1	W_2	J_1	W_2
$\alpha_1 = \alpha_2 = 1$	.04	.09	.05	.08								
	$\lambda_1^0 = 0.3$				$\lambda_1^0 = 0.5$				$\lambda_1^0 = 0.7$			
$\alpha_1 = 1,\ \alpha_2 = 0.5$	.71	.66	.78	.68	.93	.95	.96	.99	.86	.39	.97	.45
$\alpha_1 = 1,\ \alpha_2 = 0.7$	.37	.45	.67	.52	.70	.72	.92	.92	.55	.27	.80	.30
$\alpha_1 = 0.5,\ \alpha_2 = 1$	.33	.53	.76	.82	.78	.85	.95	.97	.71	.95	.96	1.0
$\alpha_1 = 0.7,\ \alpha_2 = 1$	.11	.22	.32	.44	.28	.47	.78	.80	.23	.56	.72	.90

the following DGP-T:

$$y_t = \mu_0 + \mu_1 I(t > [T\tau_1^0]) + \beta_0 t + \beta_1 (t - [T\tau_1^0]) I(t > [T\tau_1^0]) + y_t^*,$$

where $y_t^* = \alpha_1 y_{t-1}^* + e_t$ if $t \leq [\lambda_1^0 T]$ and $y_t^* = \alpha_2 y_{t-1}^* + e_t$ if $t \geq [\lambda_1^0 T] + 1$ with $y_0^* = 0,\ e_t \sim iid\, N(0,1)$. Note that we allow $\tau_1^0 \neq \lambda_1^0$. In the simulations, we fix $\tau_1^0 = 0.5$ and set $\lambda_1^0 = 0.3, 0.5, 0.7$. We consider the case of a pure level shift ($\mu_0 = \beta_0 = \beta_1 = 0,\ \mu_1 = 5$) in which case the test J_1 is employed, as well as the case of a trend break ($\mu_0 = \mu_1 = \beta_0 = 0,\ \beta_1 = 0.5$) in which case the test $W_2(1)$ is employed. The results are reported in Table 8. Note that in both the pure level shift and trend break cases, the size remains adequate, never exceeding 10%. Power is generally highest when the trend break date coincides with the persistence break date. An exception is the trend break case where the persistence change is from an $I(0)$ to an $I(1)$ process: Here, power is highest when the the trend break precedes the persistence break. In the general case, which allows for the possibility that the number of trend beaks can be different from the number of persistence breaks, we can potentially adopt the following procedure. In a first step, the number of trend breaks can be estimated using the sequential procedure developed by Kejriwal and Perron (2010) that is robust to whether the errors are $I(1)$ or $I(0)$. This estimate can subsequently be used to detrend the data and apply the $W\max_2$ test in the second step. While such a procedure is likely to be computationally intensive, it has the advantage of being agnostic to the types of breaks. Investigation of the asymptotic and finite sample properties of such a procedure is left as an important avenue for future research.

7. CONCLUSION

This paper has presented issues related to testing for multiple structural changes in the persistence of a univariate time series. In contrast to the existing literature,

which has primarily focused on subsample unit root tests and tests based on partial sums of residuals, we propose sup-Wald tests based on the difference between the sum of squared residuals under the null hypothesis of a unit root and that under the alternative hypothesis that the process displays changes in persistence over the sample. Our simulation experiments demonstrate that these tests have adequate finite sample properties. One important issue that we have not addressed is how to select the number of breaks. Indeed, we have assumed that the number of breaks is known a priori or less than some known upper bound. Bai and Perron (1998) propose a sequential strategy based on repeated application of the single break test in the context of stationary regression models. Such a strategy, however, does not directly extend to our framework, given that the process is stationary in only some regimes but has a unit root in others. Developing methods that would allow the consistent estimation of the number of breaks in this framework is an important avenue for future research. Finally, it is important to address the issue of the estimation of the break dates and develop a method to form confidence intervals. These and other issues are the object of ongoing research.

NOTES

1. Such an estimate was proposed by Chong (2001) for an AR(1) model with a single shift in persistence, although his estimation procedure did not impose the unit root restriction in the relevant regime.
2. The size and power properties using other versions of the M^{GLS} test were very similar.
3. The full set of results is available upon request.

REFERENCES

Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.

Bai, J. & P. Perron (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47–78.

Bai, J. & P. Perron (2003) Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18, 1–22.

Barsky, R.B. (1987) The Fisher hypothesis and the forecastibility and persistence of inflation. *Journal of Monetary Economics* 19, 3–24.

Berk, K.N. (1974) Consistent autoregressive spectral estimates. *Annals of Statistics* 2, 489–502.

Burdekin, R.C.K. & P.L. Siklos (1999) Exchange rate regimes and shifts in inflation persistence: Does nothing else matter. *Journal of Money, Credit and Banking* 31, 235–247.

Busetti, F. & A.M.R. Taylor (2004) Tests of stationarity against a change in persistence. *Journal of Econometrics* 123, 33–66.

Chang, M.C. (1989) Testing for Overdifferencing. Ph.D. dissertation, North Carolina State University.

Chang, M.C. & D.A. Dickey (1994) Recognizing overdifferenced time series. *Journal of Time Series Analysis* 15, 1–18.

Chong, T.T.L. (2001) Structural change in AR(1) models. *Econometric Theory* 17, 87–155.

DeLong, J.B. & L.H. Summers (1988) How does macroeconomic policy affect output? *Brookings Papers on Economic Activity* 2, 433–494.

Elliott, G., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.

Hakkio, C.S. & M. Rush (1991) Is the budget deficit too large? *Economic Inquiry* 29, 429–445.
Harvey, D.I., S.J. Leybourne, & A.M.R. Taylor (2006) Modified tests for a change in persistence. *Journal of Econometrics* 134, 441–469.
Kang, K.H., C.J. Kim, & J. Morley (2009) Changes in U.S. inflation persistence. *Studies in Nonlinear Dynamics & Econometrics* vol. 13(4), article 1.
Kejriwal, M. & P. Perron (2010) A sequential procedure to determine the number of breaks in trend with an integrated or stationary noise component. *Journal of Time Series Analysis* 31, 305–328.
Kejriwal, M. & P. Perron (2012) Estimating a Structural Change in Persistence. Manuscript in preparation, Boston University.
Kim, D. & P. Perron (2009) Unit root tests allowing for a break in the trend function under both the null and alternative hypotheses. *Journal of Econometrics* 148, 1–13.
Kim, J.Y. (2000) Detection of change in persistence of a linear time series. *Journal of Econometrics* 54, 159–178.
Kim, J.Y. (2003) Inference on segmented cointegration. *Econometric Theory* 19, 620–639.
Kurozumi, E. (2005) Detection of structural change in the long-run persistence in a univariate time series. *Oxford Bulletin of Economics and Statistics* 67, 181–206.
Leybourne, S.J., T. Kim, V. Smith, & P. Newbold (2003) Tests for a change in persistence against the null of difference-stationarity. *Econometrics Journal* 6, 291–311.
Leybourne, S.J., T. Kim, & A.M.R. Taylor (2007a) CUSUM of squares-based tests for a change in persistence. *Journal of Time Series Analysis* 28, 408–433.
Leybourne, S.J., T. Kim, & A.M.R. Taylor (2007b) Detecting multiple changes in persistence. *Studies in Nonlinear Dynamics & Econometrics* vol. 11(3), article 2.
Lütkepohl, H. & P. Saikkonen (1999) Order selection in testing for the cointegrating rank of a VAR process. In R.F. Engle & H. White (eds.), *Cointegration, Causality and Forecasting*, pp. 168–99. Oxford University Press.
Mankiw, N.G., J.A. Miron, & D.N. Weil (1987) The adjustment of expectations to change in regime: A study of the founding of the Federal Reserve. *American Economic Review* 77, 358–374.
Ng, S. & P. Perron (1995) Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90, 268–281.
Ng, S. & P. Perron (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.
Perron, P. (1989) The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57, 1361–1401.
Perron, P. (2006) Dealing with structural breaks. In K. Patterson & T.C. Mills (eds.), *Palgrave Handbook of Econometrics*, pp. 278–352. Palgrave Macmillan.
Perron, P. & Z. Qu (2006) Estimating restricted structural change models. *Journal of Econometrics* 134, 373–399.
Perron, P. & Z. Qu (2007) A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. *Economics Letters* 94, 12–19.
Taylor, A.M.R. (2005) Fluctuation tests for a change in persistence. *Oxford Bulletin of Economics and Statistics* 67, 207–230.

APPENDIX

As a matter of notation, throughout, we use the matrix norm $||B||_1 = \sup_{\|x\|\leq 1} ||Bx||$, with $\|.\|$ the standard euclidean norm. Note that $||B||_1$ equals the square root of the largest eigenvalue of $B'B$ and that $||Bx|| \leq ||B||_1||x||$. Also, we use the usual norm $||B||^2 = \text{tr}(B'B)$, such that $||B||_1^2 \leq ||B||^2$. Note that for any conformable matrices B_1 and B_2, we have $||B_1B_2|| \leq ||B_1||||B_2||_1$. Next, we define $\bar{z}_j = (T_j - T_{j-1})^{-1}\sum_{t=T_{j-1}+1}^{T_j} z_t$ and $\bar{z}_{j,-1} = (T_j - T_{j-1})^{-1}\sum_{t=T_{j-1}+1}^{T_j} z_{t-1}$. Finally, we define the following regime-wise

demeaned and detrended Brownian motions: $W^{(j)}(r) = W(r) - (\lambda_j - \lambda_{j-1})^{-1} \int_{\lambda_{j-1}}^{\lambda_j} W(r)dr$, and

$$\widetilde{W}^{(j)}(r) = W^{(j)}(r) - \left[\frac{\int_{\lambda_{j-1}}^{\lambda_j} rW^{(j)}(r)dr}{\int_{\lambda_{j-1}}^{\lambda_j} \left\{ r - \left(\lambda_j - \lambda_{j-1}\right)^{-1} \int_{\lambda_{j-1}}^{\lambda_j} rdr \right\}^2 dr} \right] \times \left[r - \left(\lambda_j - \lambda_{j-1}\right)^{-1} \int_{\lambda_{j-1}}^{\lambda_j} rdr \right],$$

where $W(.)$ denotes a standard Brownian motion on [0, 1]. We first state a lemma about the weak convergence of various sample moments whose proof is standard and thus omitted.

LEMMA A.1. *If $\{w_t\}$ is generated as $w_t = w_{t-1} + v_t$, where v_t satisfies Assumption A1, the following weak convergence results hold (for $i = 1, \ldots, k+1$): (a) $T^{-3/2} \sum_{t=1}^{[T\lambda_i]} w_t \Rightarrow \sigma \int_0^{\lambda_i} W(r)dr$; (b) $T^{-3/2} \sum_{t=1}^{[T\lambda_i]} w_t^2 \Rightarrow \sigma^2 \int_0^{\lambda_i} W(r)^2 dr$; (c) $T^{-1} \sum_{t=1}^{[T\lambda_i]} w_{t-1} v_t \Rightarrow \sigma^2 \int_0^{\lambda_i} W(r)dW(r)$.*

Proof of Theorem 1. We shall prove the theorem for Models 1a and 2a. The proofs for the other models are similar and hence omitted. For Model 1a, We have $y_t = c_i + \alpha_i y_{t-1} + u_t$, $t = T_{i-1}+1, \ldots, T_i$ for $i = 1, \ldots, k+1$ with $\alpha_i = 1$, $c_i = 0$ in odd regimes and $|\alpha_i| < 1$, c_i unrestricted in even regimes. Under the null hypothesis of a unit root throughout the sample, the sum of squared residuals is $SSR_0 = \sum_{t=1}^{T} (y_t - y_{t-1})^2 = \sum_{t=1}^{T} u_t^2$. If k is even, the sum of squares residuals under the alternative hypothesis is

$$SSR_{1a,k} = \sum_{i=1}^{k/2} \left[\sum_{t=T_{2i-1}+1}^{T_{2i}} \left\{ y_t - \bar{y}_{2i} - \hat{\alpha}_{2i} (y_{t-1} - \bar{y}_{2i,-1}) \right\}^2 \right] + \sum_{i=0}^{k/2} \sum_{t=T_{2i}+1}^{T_{2i+1}} u_t^2, \qquad \textbf{(A.1)}$$

where, for $i = 1, \ldots, k/2$, $\hat{\alpha}_{2i} = \sum_{t=T_{2i-1}+1}^{T_{2i}} (y_t - \bar{y}_{2i})(y_{t-1} - \bar{y}_{2i,-1}) / \sum_{t=T_{2i-1}+1}^{T_{2i}} (y_{t-1} - \bar{y}_{2i,-1})^2$. Note that, under the null, $y_t = y_{t-1} + u_t$, which implies, $\bar{y}_{2i} = \bar{y}_{2i,-1} + \bar{u}_{2i}$. Substituting in the expression for $\hat{\alpha}_{2i}$ and using Lemma A.1, we have

$$T\left(\hat{\alpha}_{2i} - 1\right) = \frac{T^{-1} \sum_{t=T_{2i-1}+1}^{T_{2i}} \left(y_{t-1} - \bar{y}_{2i,-1}\right) u_t}{T^{-2} \sum_{t=T_{2i-1}+1}^{T_{2i}} \left(y_{t-1} - \bar{y}_{2i,-1}\right)^2} \Rightarrow \frac{\int_{\lambda_{2i-1}}^{\lambda_{2i}} W^{(2i)}(r) dW(r)}{\int_{\lambda_{2i-1}}^{\lambda_{2i}} \left[W^{(2i)}(r) \right]^2 dr}.$$

From (A.1), we thus have, under the null hypothesis,

$$\begin{aligned} SSR_{1a,k} &= \sum_{i=1}^{k/2} \left[\frac{-\left\{ \sum_{t=T_{2i-1}+1}^{T_{2i}} \left(y_{t-1} - \bar{y}_{2i,-1}\right) u_t \right\}^2}{\sum_{t=T_{2i-1}+1}^{T_{2i}} \left(y_{t-1} - \bar{y}_{2i,-1}\right)^2} + \sum_{t=T_{2i-1}+1}^{T_{2i}} (u_t - \bar{u}_{2i})^2 \right] + \sum_{i=0}^{k/2} \sum_{t=T_{2i}+1}^{T_{2i+1}} u_t^2 \\ &= \sum_{i=1}^{k/2} \left[\frac{-\left\{ \sum_{t=T_{2i-1}+1}^{T_{2i}} \left(y_{t-1} - \bar{y}_{2i,-1}\right) u_t \right\}^2}{\sum_{t=T_{2i-1}+1}^{T_{2i}} \left(y_{t-1} - \bar{y}_{2i,-1}\right)^2} - \frac{T}{T_{2i} - T_{2i-1}} \left\{ T^{-1/2} \sum_{t=T_{2i-1}+1}^{T_{2i}} u_t \right\}^2 \right] \\ &\quad + \sum_{t=1}^{T} u_t^2, \end{aligned}$$

so that

$$SSR_0 - SSR_{1a,k} = \sum_{i=1}^{k/2}\left[\frac{\left\{\sum_{t=T_{2i-1}+1}^{T_{2i}}\left(y_{t-1}-\bar{y}_{2i,-1}\right)u_t\right\}^2}{\sum_{t=T_{2i-1}+1}^{T_{2i}}\left(y_{t-1}-\bar{y}_{2i,-1}\right)^2}+\frac{T}{T_{2i}-T_{2i-1}}\right.$$
$$\left.\times\left\{T^{-1/2}\sum_{t=T_{2i-1}+1}^{T_{2i}}u_t\right\}^2\right]$$
$$\Rightarrow \sigma^2\sum_{i=1}^{k/2}\left[\frac{\left\{\int_{\lambda_{2i-1}}^{\lambda_{2i}}W^{(2i)}(r)dW(r)\right\}^2}{\int_{\lambda_{2i-1}}^{\lambda_{2i}}\left[W^{(2i)}(r)\right]^2dr}+\frac{1}{\lambda_{2i}-\lambda_{2i-1}}\right.$$
$$\left.\times\left\{W\left(\lambda_{2i}\right)-W\left(\lambda_{2i-1}\right)\right\}^2\right].$$

It is easy to show that $T^{-1}SSR_{1a,k} = T^{-1}\sum_{t=1}^{T}u_t^2 + o_p(1) \xrightarrow{p} \sigma^2$, so that

$$kF_{1a}(\lambda,k) \Rightarrow \sum_{i=1}^{k/2}\left[\frac{\left\{\int_{\lambda_{2i-1}}^{\lambda_{2i}}W^{(2i)}(r)dW(r)\right\}^2}{\int_{\lambda_{2i-1}}^{\lambda_{2i}}\left[W^{(2i)}(r)\right]^2dr}+\frac{1}{\lambda_{2i}-\lambda_{2i-1}}\left\{W\left(\lambda_{2i}\right)-W\left(\lambda_{2i-1}\right)\right\}^2\right].$$

If k is odd,

$$SSR_{1a,k} = \sum_{i=0}^{(k-1)/2}\sum_{t=T_{2i}}^{T_{2i+1}}u_t^2 + \sum_{i=1}^{(k+1)/2}\left[\sum_{t=T_{2i-1}+1}^{T_{2i}}\left\{y_t-\bar{y}_{2i}-\hat{\alpha}_{2i}(y_{t-1}-\bar{y}_{2i,-1}\right\}^2\right]$$

and similar derivations show that

$$(k+1)F_{1a}(\lambda,k) \Rightarrow \sum_{i=1}^{(k+1)/2}\left[\frac{\left\{\int_{\lambda_{2i-1}}^{\lambda_{2i}}W^{(2i)}(r)dW(r)\right\}^2}{\int_{\lambda_{2i-1}}^{\lambda_{2i}}\left[W^{(2i)}(r)\right]^2dr}+\frac{1}{\lambda_{2i}-\lambda_{2i-1}}\right.$$
$$\left.\times\left\{W(\lambda_{2i})-W(\lambda_{2i-1})\right\}^2\right].$$

For Model 2a**,** we have $y_t = c_i + b_i t + \alpha_i y_{t-1} + u_t$, $t = T_{i-1}+1,\ldots,T_i$, with $\alpha_i = 1$, $b_i = 0$, c_i unrestricted in odd regimes and $|\alpha_i| < 1$, b_i, c_i unrestricted in even regimes. Under the null, $y_t = c + y_{t-1} + u_t$. For this model, we have $SSR_0^* = \sum_{t=1}^{T}[y_t - y_{t-1} - T^{-1}\sum_{t=1}^{T}(y_t - y_{t-1})]^2 = \sum_{t=1}^{T}(u_t - \bar{u})^2$. Again, consider first the case with k even. For $t \in [T_{2i-1}+1, T_{2i}]$, define

$$\tilde{y}_t = y_t - \bar{y}_{2i} - \frac{\sum_{t=T_{2i-1}+1}^{T_{2i}} \left(y_t - \bar{y}_{2i}\right)\left(t - \bar{t}_{2i}\right)}{\sum_{t=T_{2i-1}+1}^{T_{2i}} \left(t - \bar{t}_{2i}\right)^2}\left(t - \bar{t}_{2i}\right)$$

$$\tilde{y}_{t-1} = y_{t-1} - \bar{y}_{2i,-1} - \frac{\sum_{t=T_{2i-1}+1}^{T_{2i}} \left(y_{t-1} - \bar{y}_{2i,-1}\right)\left(t - \bar{t}_{2i}\right)}{\sum_{t=T_{2i-1}+1}^{T_{2i}} \left(t - \bar{t}_{2i}\right)^2}\left(t - \bar{t}_{2i}\right).$$

Then, under the null hypothesis, we can write

$$\tilde{y}_t = \tilde{y}_{t-1} + u_t - \bar{u}_{2i} - \frac{\sum_{t=T_{2i-1}+1}^{T_{2i}} \left(t - \bar{t}_{2i}\right) u_t}{\sum_{t=T_{2i-1}+1}^{T_{2i}} \left(t - \bar{t}_{2i}\right)^2}\left(t - \bar{t}_{2i}\right). \tag{A.2}$$

We have

$$\begin{aligned} SSR_{2a,k} = &\sum_{i=1}^{k/2} \sum_{t=T_{2i-1}+1}^{T_{2i}} \left\{\tilde{y}_t - \tilde{\alpha}_{2i}\tilde{y}_{t-1}\right\}^2 \\ &+ \sum_{i=0}^{k/2} \sum_{t=T_{2i}+1}^{T_{2i+1}} \left\{ y_t - y_{t-1} - \frac{1}{T_{2i+1} - T_{2i}} \sum_{t=T_{2i}+1}^{T_{2i+1}} \left(y_t - y_{t-1}\right) \right\}^2, \end{aligned} \tag{A.3}$$

where $\tilde{\alpha}_{2i} = \sum_{t=T_{2i-1}+1}^{T_{2i}} \tilde{y}_t \tilde{y}_{t-1} / \sum_{t=T_{2i-1}+1}^{T_{2i}} \tilde{y}_{t-1}^2$. Then using (A.2), we can express (A.3) as

$$\begin{aligned} SSR_{2a,k} = \sum_{i=1}^{k/2} \Bigg[&\frac{-\left\{\sum_{t=T_{2i-1}+1}^{T_{2i}} \tilde{y}_{t-1} u_t\right\}^2}{\sum_{t=T_{2i-1}+1}^{T_{2i}} \tilde{y}_{t-1}^2} + \sum_{t=T_{2i-1}+1}^{T_{2i}} (u_t - \bar{u}_{2i})^2 \\ &- \frac{\left\{\sum_{t=T_{2i-1}+1}^{T_{2i}} \left(t - \bar{t}_{2i}\right) u_t\right\}^2}{\sum_{t=T_{2i-1}+1}^{T_{2i}} \left(t - \bar{t}_{2i}\right)^2} \Bigg] + \sum_{i=0}^{k/2} \sum_{t=T_{2i}+1}^{T_{2i+1}} \left(u_t - \bar{u}_{2i+1}\right)^2. \end{aligned}$$

We thus get

$$\begin{aligned} SSR_0^* - SSR_{2a,k} = &- \left(T^{-1/2} \sum_{t=1}^{T} u_t \right)^2 + \sum_{i=0}^{k/2} \left[\frac{T}{T_{2i+1} - T_{2i}} \left(T^{-1/2} \sum_{t=T_{2i}+1}^{T_{2i+1}} u_t \right)^2 \right] \\ &+ \sum_{i=1}^{k/2} \Bigg[\frac{\left\{\sum_{t=T_{2i-1}+1}^{T_{2i}} \tilde{y}_{t-1} u_t\right\}^2}{\sum_{t=T_{2i-1}+1}^{T_{2i}} \tilde{y}_{t-1}^2} + \frac{T}{T_{2i} - T_{2i-1}} \left\{ T^{-1/2} \sum_{t=T_{2i-1}+1}^{T_{2i}} u_t \right\}^2 \\ &+ \frac{\left\{\sum_{t=T_{2i-1}+1}^{T_{2i}} \left(t - \bar{t}_{2i}\right) u_t\right\}^2}{\sum_{t=T_{2i-1}+1}^{T_{2i}} \left(t - \bar{t}_{2i}\right)^2} \Bigg], \end{aligned}$$

which yields

$$2kF_{2a}(\lambda,k) \Rightarrow -\{W(1)\}^2 + \sum_{i=0}^{k/2}\left[\frac{1}{\lambda_{2i+1}-\lambda_{2i}}\left\{W(\lambda_{2i+1})-W(\lambda_{2i})\right\}^2\right]$$
$$+\sum_{i=1}^{k/2}\left[\begin{array}{l}\dfrac{\left\{\int_{\lambda_{2i-1}}^{\lambda_{2i}}\widetilde{W}^{(2i)}(r)dW(r)\right\}^2}{\int_{\lambda_{2i-1}}^{\lambda_{2i}}\left[\widetilde{W}^{(2i)}(r)\right]^2 dr}+\frac{1}{\lambda_{2i}-\lambda_{2i-1}}\left\{W(\lambda_{2i})-W(\lambda_{2i-1})\right\}^2 \\ +\dfrac{\left[\int_{\lambda_{2i-1}}^{\lambda_{2i}}\left\{r-(\lambda_{2i}-\lambda_{2i-1})^{-1}\int_{\lambda_{2i-1}}^{\lambda_{2i}}rdr\right\}dW(r)\right]^2}{\int_{\lambda_{2i-1}}^{\lambda_{2i}}\left\{r-(\lambda_{2i}-\lambda_{2i-1})^{-1}\int_{\lambda_{2i-1}}^{\lambda_{2i}}rdr\right\}^2 dr}\end{array}\right].$$

If k is odd,

$$SSR_{2a,k} = \sum_{i=0}^{(k-1)/2}\sum_{t=T_{2i}+1}^{T_{2i+1}}\left\{y_t - y_{t-1} - \frac{1}{T_{2i+1}-T_{2i}}\sum_{t=T_{2i}+1}^{T_{2i+1}}(y_t-y_{t-1})\right\}^2$$
$$+\sum_{i=1}^{(k+1)/2}\sum_{t=T_{2i-1}+1}^{T_{2i}}\left\{\tilde{y}_t-\tilde{\alpha}_{2i}\tilde{y}_{t-1}\right\}^2$$

and similar derivations yield the result stated in Theorem 1. Given these limits, the results of Theorem 1 follow from an application of the continuous mapping theorem. ■

For the proof of Theorem 2, we consider Model 1a when k is even; the proof is similar for the other cases. The autoregression in the ith regime ($i = 1,\ldots,k/2$) is

$$\Delta y_t = c_{2i} + (\alpha_{2i}-1)y_{t-1} + \sum_{j=1}^{l_T}\pi_j\Delta y_{t-j} + v_t^*, \tag{A.4}$$

with $v_t^* = e_t + v_t$, and $e_t = \sum_{j>l_T}\pi_j\Delta y_{t-j}$. Let $\eta_t' = (\Delta y_{t-1},\ldots,\Delta y_{t-l_T})$, $\eta = (\eta_1,\ldots,\eta_T)'$, $\Pi = (\pi_1,\ldots,\pi_{l_T})'$, $V^* = (v_1^*,\ldots,v_T^*)' = V + \mathcal{E}$ with $V = (v_1,\ldots,v_T)'$ and $\mathcal{E} = (e_1,\ldots e_T)'$. We can write (A.4) as $\Delta y_t = c_i + (\alpha_i - 1)y_{t-1} + \eta_t'\Pi + v_t^*$ with $\alpha_i = 1$, $c_i = 0$ in odd regimes and $|\alpha_i| < 1$, c_i unrestricted in even regimes. For $j = 1,\ldots,k+1$, we denote $\Delta Y_j = (\Delta y_{T_{j-1}+1},\ldots,\Delta y_{T_j})'$, $\eta_j^* = (\eta_{T_{j-1}+1},\ldots,\eta_{T_j})'$, $\mathcal{E}_j = (e_{T_{j-1}+1},\ldots,e_{T_j})'$, $V_j = (v_{T_{j-1}+1},\ldots,v_{T_j})'$, and $V_j^* = (v_{T_{j-1}+1}^*,\ldots,v_{T_j}^*)'$. For $i = 1,\ldots,k/2$, let $\hat{\gamma}_{2i} = (\hat{c}_{2i},\hat{\alpha}_{2i}-1)'$ and $Z_{2i} = (z_{T_{2i-1}+1},\ldots,z_{T_{2i}})'$ where $z_t = (1, y_{t-1})'$ for $t = T_{2i-1}+1,\ldots,T_{2i}$. Define the (2×2) diagonal matrix $D_T = \mathrm{diag}(T^{-1/2}, T^{-1})$. The proof of Theorem 2 is based on the following lemma.

LEMMA A.2. *Assume y_t is generated as $y_t = y_{t-1} + u_t$. Under Assumptions A1–A3, we have (a) $||(\eta'\eta)^{-1}||_1 = O_p(T^{-1})$; (b) (i) $||D_T Z_{2i}'\eta_{2i}^*|| = O_p(l_T^{1/2})$ and (ii) $||D_T Z_{2i}'\mathcal{E}_{2i}|| = o_p(l_T^{-1})$, for $i = 1,\ldots,k/2$; (c) $||\eta'V|| = O_p(T^{1/2}l_T^{1/2})$; (d) $||\eta'\mathcal{E}|| = o_p(Tl_T^{-1/2})$; (e) $||\mathcal{E}'\mathcal{E}|| = o_p(T)$; (f) $||\mathcal{E}'V|| = o_p(T)$; (g) $||\eta'V^*|| = o_p(Tl_T^{-1/2})$; (h) $||[\eta'\eta - \sum_{i=1}^{k/2}\left\{\eta_{2i}^{*\prime}Z_{2i}(Z_{2i}'Z_{2i})^{-1}Z_{2i}'\eta_{2i}^*\right\}]^{-1}||_1 = O_p(T^{-1})$.*

Proof of Lemma A.2. (a) Let $\Gamma_l^* = (\Gamma_{i-j})_{i,j=1}^{l_T}$, where $\Gamma_h = \mathbb{E}(u_t u_{t-h})$. From Berk (1974, Lem. 3), it follows that $||(T^{-1}\eta'\eta)^{-1} - (\Gamma_l^*)^{-1}||_1 = O_p(T^{-1/2}l_T)$. Since $\|(\Gamma_l^*)^{-1}\|_1 = O(1)$ uniformly in l_T for sequences such that $T^{-1/2}l_T \to 0$, the result follows from the fact that $|||(T^{-1}\eta'\eta)_1^{-1} - \|(\Gamma_l^*)^{-1}\|||_1| \le ||(T^{-1}\eta'\eta)^{-1} - (\Gamma_l^*)^{-1}||_1 = o_p(1)$. (b) For (i), the result follows since each element of $D_T Z_{2i}' \eta_{2i}^*$ is $O_p(1)$ and the number of elements is of order $O(l_T)$. For (ii), the result follows from Lemma A.2(a) of Lütkepohl and Saikkonen (1999). (c) The elements of $T^{-1/2}\eta' V$ are each $O_p(1)$ (since each element of η_t and v_t is uncorrelated), and the result follows since the number of elements is of order $O(l_T)$. (d) We have

$$\begin{aligned}
\mathbb{E}\left\|T^{-1}\eta'\mathcal{E}\right\| &\le T^{-1}\sum_{t=1}^{T}\mathbb{E}(\|e_t\eta_t\|) \le \left\{\mathbb{E}\left(\|\eta_t\|^2\right)\mathbb{E}(e_t^2)\right\}^{1/2} \\
&= C_2 l_T^{1/2}\left\{\mathbb{E}\left(\sum_{j>l_T}\Delta y_{t-j}\pi_j\right)^2\right\}^{1/2} \le C_2 l_T^{1/2}\left\{\sum_{i>l_T}\sum_{j>l_T}\left|\Gamma_{i-j}\right|\left|\pi_i\right|\left|\pi_j\right|\right\}^{1/2} \\
&\le C_3 l_T^{1/2}\sum_{j>l_T}\left|\pi_j\right| = o\left(l_T^{-1/2}\right),
\end{aligned}$$

using the fact that $\left|\Gamma_{i-j}\right|$ is uniformly bounded by the stationarity of u_t. (e) We have

$$\begin{aligned}
\mathbb{E}\left\|T^{-1}\mathcal{E}'\mathcal{E}\right\| &= T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(e_t^2\right) = T^{-1}\sum_{t=1}^{T}\sum_{i>l_T}\sum_{a>l_T}\pi_i\mathbb{E}\left(\Delta y_{t-i}\Delta y_{t-a}\right)\pi_a \\
&\le T^{-1}\sum_{i>l_T}\sum_{a>l_T}\sum_{t=1}^{T}|\pi_i|\left|\Gamma_{a-i}\right||\pi_a| \le o\left(l_T^{-2}\right) = o(1),
\end{aligned}$$

where we again use the fact that $\left|\Gamma_j\right|$ is bounded uniformly in j. (f) We have $T^{-1}\sum_{t=1}^{T} v_t e_t = T^{-1}\sum_{i>l_T}\Pi_i'\sum_{t=1}^{T}\Delta y_{t-i}v_t$, so that

$$\left\|T^{-1}\sum_{t=1}^{T}v_t e_t\right\| \le T^{-1}\sum_{i>l_T}\|\pi_i\|\left\|\sum_{t=1}^{T}\Delta y_{t-i}v_t\right\| = o_p\left(l_T^{-1}T^{-1/2}\right) = o_p(1),$$

where we used the fact that $T^{-1/2}\sum_{t=1}^{T}\Delta y_{t-i}v_t = O_p(1)$. (g) Since $V^* = V + \mathcal{E}$, $||\eta'V^*|| \le ||\eta'V|| + ||\eta'\mathcal{E}|| = O_p(T^{1/2}l_T^{1/2}) + o_p(Tl_T^{-1/2}) = o_p(Tl_T^{-1/2})$. (h) Let

$$q = \left\|\left[T^{-1}\eta'\eta - T^{-1}\sum_{i=1}^{k/2}\left\{\eta_{2i}^{*\prime}Z_{2i}\left(Z_{2i}'Z_{2i}\right)^{-1}Z_{2i}'\eta_{2i}^*\right\}\right]^{-1} - \left(\Gamma_l^*\right)^{-1}\right\|_1$$

and $Q = ||T^{-1}\eta'\eta - T^{-1}\sum_{i=1}^{k/2}\left\{\eta_{2i}^{*\prime}Z_{2i}(Z_{2i}'Z_{2i})^{-1}Z_{2i}'\eta_{2i}^*\right\} - \Gamma_l^*||_1$.

Then $q \leq \{q + \|(\Gamma_l^*)^{-1}\|_1\} Q \|(\Gamma_l^*)^{-1}\|_1$ or $q \leq \|(\Gamma_l^*)^{-1}\|_1^2 Q/[1 - Q\|(\Gamma_l^*)^{-1}\|_1]$. Also,

$$\begin{aligned} Q &\leq \left\| T^{-1}\eta'\eta - \Gamma_l^* \right\|_1 + \left\| T^{-1} \sum_{i=1}^{k/2} \left\{ \eta_{2i}^{*\prime} Z_{2i} (Z_{2i}' Z_{2i})^{-1} Z_{2i}' \eta_{2i}^* \right\} \right\| \\ &= \left\| T^{-1}\eta'\eta - \Gamma_l^* \right\|_1 + T^{-1} \sum_{i=1}^{k/2} \left\{ \left\| \eta_{2i}^{*\prime} Z_{2i} D_T \right\| \left\| (D_T Z_{2i}' Z_{2i} D_T)^{-1} \right\| \left\| D_T Z_{2i}' \eta_{2i}^* \right\| \right\} \\ &= O_p\left(l_T/T^{1/2}\right) + T^{-1} O_p\left(l_T^{1/2}\right) O_p(1)\, O_p\left(l_T^{1/2}\right) = O_p\left(l_T/T^{1/2}\right). \end{aligned}$$

Since $\|(\Gamma_l^*)^{-1}\|_1 = O_p(1)$, we get $q = O_p(l_T/T^{1/2})$, and thus

$$\begin{aligned} &\left| \left\| \left[T^{-1}\eta'\eta - T^{-1} \sum_{i=1}^{k/2} \left\{ \eta_{2i}^{*\prime} Z_{2i} (Z_{2i}' Z_{2i})^{-1} Z_{2i}' \eta_{2i}^* \right\} \right]^{-1} \right\|_1 - \left\| (\Gamma_l^*)^{-1} \right\|_1 \right| \\ &\quad = O_p\left(l_T/T^{1/2}\right) = o_p(1), \end{aligned}$$

so that $\left\| \left[T^{-1}\eta'\eta - T^{-1} \sum_{i=1}^{k/2} \left\{ \eta_{2i}^{*\prime} Z_{2i} (Z_{2i}' Z_{2i})^{-1} Z_{2i}' \eta_{2i}^* \right\} \right]^{-1} \right\|_1 = O_p(1)$ and the result follows. ■

Proof of Theorem 2 (Model 1a and *k* even). For $i = 1, \ldots, k+1$, we denote the vector of residuals in the jth regime under the null and alternative hypotheses by $\tilde{V}_i^*$ and $\hat{V}_i^*$, respectively. Then we have

$$\begin{aligned} \tilde{V}_i^* &= \Delta Y_i - \eta_i^* \tilde{\Pi}, \quad \text{for } i = 1, \ldots, k+1 \\ \hat{V}_{2i}^* &= \Delta Y_{2i} - \eta_{2i}^* \hat{\Pi} - Z_{2i} \hat{\gamma}_{2i}, \quad \text{for } i = 1, \ldots, k/2 \\ \hat{V}_{2i+1}^* &= \Delta Y_{2i+1} - \eta_{2i+1}^* \hat{\Pi}, \quad \text{for } i = 0, \ldots, k/2, \end{aligned} \tag{A.5}$$

where $\tilde{\Pi} - \Pi = \left(\eta'\eta\right)^{-1} \eta' V^*$ under H_0. Also, $\hat{\Pi}$ and $\hat{\gamma}_{2i}$ satisfy the first-order conditions

$$Z_{2i}' \hat{V}_{2i}^* = 0, \quad \text{for } i = 1, \ldots, k/2 \tag{A.6}$$

$$\sum_{i=1}^{k/2} \eta_{2i}^* \hat{V}_{2i}^* + \sum_{i=0}^{k/2} \eta_{2i+1}^* \hat{V}_{2i+1}^* = 0. \tag{A.7}$$

Under H_0, from (A.7), we have $\hat{\Pi} - \Pi = (\eta'\eta)^{-1}(\eta' V^* - \sum_{i=1}^{k/2} \eta_{2i}^{*\prime} Z_{2i} \hat{\gamma}_{2i})$. Next, from (A.6),

$$D_T^{-1} \hat{\gamma}_{2i} = \left(D_T Z_{2i}' Z_{2i} D_T\right)^{-1} \left[D_T Z_{2i}' \eta_{2i}^* (\Pi - \hat{\Pi}) + D_T Z_{2i}' \mathcal{E}_{2i} + D_T Z_{2i}' V_{2i} \right] \tag{A.8}$$

for $i = 1, \ldots, k/2$. Solving for $(\hat{\Pi} - \Pi)$ we get

$$\begin{aligned} \hat{\Pi} - \Pi = &\left[\eta'\eta - \sum_{i=1}^{k/2} \left\{ \eta_{2i}^{*\prime} Z_{2i} \left(Z_{2i}' Z_{2i}\right)^{-1} Z_{2i}' \eta_{2i}^* \right\} \right]^{-1} \\ &\times \left[\eta' V^* - \sum_{i=1}^{k/2} \left\{ \eta_{2i}^{*\prime} Z_{2i} \left(Z_{2i}' Z_{2i}\right)^{-1} Z_{2i}' V_{2i}^* \right\} \right]. \end{aligned} \tag{A.9}$$

Using Lemma A.2 (b,g,h), we get $||\hat{\Pi} - \Pi|| = o_p(l_T^{-1/2})$. Then, using Lemma A.2(b),

$$\left\|\left(D_T Z_{2i}' Z_{2i} D_T\right)^{-1} D_T Z_{2i}' \eta_{2i}^* \left(\Pi - \hat{\Pi}\right)\right\| \leq \left\|\left(D_T Z_{2i}' Z_{2i} D_T\right)^{-1}\right\| \left\|D_T Z_{2i}' \eta_{2i}^*\right\| \left\|\left(\Pi - \hat{\Pi}\right)\right\|$$
$$= O_p(1).O_p\left(l_T^{1/2}\right) o_p\left(l_T^{-1/2}\right) = o_p(1).$$

Also, $\left\|D_T Z_{2i}' \mathcal{E}_{2i}\right\| = o_p(l_T^{-1})$. Using this in (A.8), we have

$$D_T^{-1}\hat{\gamma}_{2i} = \left(D_T Z_{2i}' Z_{2i} D_T\right)^{-1} D_T Z_{2i}' V_{2i} + o_p(1). \tag{A.10}$$

Further, we get $\hat{\Pi} - \tilde{\Pi} = -(\eta'\eta)^{-1} \sum_{i=1}^{k/2} \left\{\eta_{2i}^{*\prime} Z_{2i} \hat{\gamma}_{2i}\right\}$ so that

$$\left\|\hat{\Pi} - \tilde{\Pi}\right\| \leq \left\|(\eta'\eta)^{-1}\right\|_1 \left\|\sum_{i=1}^{k/2} \left\{\left(\eta_{2i}^{*\prime} Z_{2i} D_T\right)\left(D_T^{-1}\hat{\gamma}_{2i}\right)\right\}\right\|$$
$$\leq \left\|(\eta'\eta)^{-1}\right\|_1 \sum_{i=1}^{k/2} \left\|\eta_{2i}^{*\prime} Z_{2i} D_T\right\| \left\|D_T^{-1}\hat{\gamma}_{2i}\right\| = O_p\left(l_T^{1/2} T^{-1}\right). \tag{A.11}$$

We can write, from (A.5), for $i = 1, \ldots, k/2$, $\tilde{V}_{2i}^* = \hat{V}_{2i}^* + Z_{2i}\hat{\gamma}_{2i} + \eta_{2i}^*(\hat{\Pi} - \tilde{\Pi})$, and for $i = 0, \ldots, k/2$, $\tilde{V}_{2i+1}^* = \hat{V}_{2i+1}^* + \eta_{2i+1}^*(\hat{\Pi} - \tilde{\Pi})$. Thus the numerator of the F statistic can be written as

$$SSR_0 - SSR_{1a,k} = \sum_{i=1}^{k/2} \left\{\tilde{V}_{2i}^{*\prime}\tilde{V}_{2i}^* - \hat{V}_{2i}^{*\prime}\hat{V}_{2i}^*\right\} + \sum_{i=0}^{k/2} \left\{\tilde{V}_{2i+1}^{*\prime}\tilde{V}_{2i+1}^* - \hat{V}_{2i+1}^{*\prime}\hat{V}_{2i+1}^*\right\}$$
$$= \sum_{i=1}^{k/2} \left(D_T^{-1}\hat{\gamma}_{2i}\right)' \left(D_T Z_{2i}' Z_{2i} D_T\right) D_T^{-1}\hat{\gamma}_{2i}$$
$$+ \left(\hat{\Pi} - \tilde{\Pi}\right)' \sum_{i=1}^{k/2} \left(\eta_{2i}^{*\prime} Z_{2i} D_T\right)\left(D_T^{-1}\hat{\gamma}_{2i}\right). \tag{A.12}$$

Next,

$$\left\|\left(\hat{\Pi} - \tilde{\Pi}\right)' \sum_{i=1}^{k/2} \left(\eta_{2i}^{*\prime} Z_{2i} D_T\right)\left(D_T^{-1}\hat{\gamma}_{2i}\right)\right\| \leq \left\|\hat{\Pi} - \tilde{\Pi}\right\| \sum_{i=1}^{k/2} \left\|\left(\eta_{2i}^{*\prime} Z_{2i} D_T\right)\right\| \left\|\left(D_T^{-1}\hat{\gamma}_{2i}\right)\right\|$$
$$= O_p\left(l_T^{1/2} T^{-1}\right).O_p\left(l_T^{1/2}\right).O_p(1)$$
$$= O_p\left(l_T T^{-1}\right) = o_p(1).$$

Then, using (A.10) in (A.12), we have

$$SSR_0 - SSR_{1a,k} = \sum_{i=1}^{k/2} \left\{V_{2i}' Z_{2i} D_T (D_T Z_{2i}' Z_{2i} D_T)^{-1} D_T Z_{2i}' V_{2i}\right\} + o_p(1). \tag{A.13}$$

Under H_0, we have the Beveridge-Nelson decomposition, $y_t = d(1)w_t + \bar{u}_0 - \bar{u}_t$, where $w_t = \sum_{j=1}^{t} v_j$, $\bar{u}_t = \sum_{s=0}^{\infty} \bar{d}_s v_{t-s}$, $\bar{d}_s = \sum_{i=s+1}^{\infty} d_i$. Note that $(\bar{u}_t)$ is stochastically of smaller order of magnitude than (w_t). Then for $r \in (0, 1]$, we have $T^{-2}\sum_{t=1}^{[Tr]} y_t^2 =$

$d(1)^2 T^{-2} \sum_{t=1}^{[Tr]} w_t^2 + o_p(1)$ and $T^{-1} \sum_{t=1}^{[T\lambda_i]} y_{t-1} v_t = d(1) T^{-1} \sum_{t=1}^{[Tr]} w_{t-1} v_t + o_p(1)$. Using these results in (A.13),

$$SSR_0 - SSR_{1a,k} \Rightarrow \sigma^2 \sum_{i=1}^{k/2} \left[\frac{\left\{ \int_{\lambda_{2i-1}}^{\lambda_{2i}} W^{(2i)}(r) dW(r) \right\}^2}{\int_{\lambda_{2i-1}}^{\lambda_{2i}} \left[W^{(2i)}(r) \right]^2 dr} + \frac{1}{\lambda_{2i} - \lambda_{2i-1}} \{ W(\lambda_{2i}) - W(\lambda_{2i-1}) \}^2 \right].$$

Using the fact that $T^{-1} SSR_{1a,k} \xrightarrow{p} \sigma^2$, the result follows. ■

Proof of Theorem 3. For part (a), we prove the result for Model 1a and k even. To show that the test is consistent, we will show that for $\lambda^0 = (\lambda_1^0, \ldots, \lambda_k^0)$, the true break fractions, the statistic $F_{1a}(\lambda^0, k)$ diverges. To see this, first note that we can express the vector of residuals computed under the null and alternative as, respectively,

$$\begin{aligned} \tilde{V}^* &= M_\eta \Delta Y \\ \hat{V}^* &= M_\eta \hat{V}^* = M_\eta \Delta Y - M_\eta \bar{Z}^0 \hat{\gamma} = \tilde{V}^* - M_\eta \bar{Z}^0 \hat{\gamma}, \end{aligned} \tag{A.14}$$

where $M_\eta = I_T - \eta(\eta'\eta)^{-1}\eta'$, $\hat{\gamma}$ and γ^0 are the estimated and true values under the alternative and $\bar{Z}^0$ is the diagonal partition of $Z = (z_1, \ldots, z_T)'$ at the true break dates $(\lambda_1^0, \ldots, \lambda_k^0)$ (see Bai and Perron, 1998). From (A.14), we can write

$$\tilde{V}^{*\prime} \tilde{V}^* - \hat{V}^{*\prime} \hat{V}^* = \hat{\gamma}' \bar{Z}^{0\prime} M_\eta \bar{Z}^0 \hat{\gamma} + 2 \hat{V}^{*\prime} M_\eta \bar{Z}^0 \hat{\gamma} = \hat{\gamma}' \bar{Z}^{0\prime} M_\eta \bar{Z}^0 \hat{\gamma}, \tag{A.15}$$

where the second term is zero by the first-order conditions (A.6) and (A.7). Define the $[2(k+1) \times 2(k+1)]$ matrix $D_{1T} = \mathrm{diag}(D_T, T^{-1/2} I_2, D_T, T^{-1/2} I_2, \ldots, D_T)$. Then we have $D_{1T} \bar{Z}^{0\prime} M_\eta \bar{Z}^0 D_{1T} = O_p(1)$. Next, note that $D_{1T}^{-1} \hat{\gamma} = T^{1/2}(\mathbf{0}, \hat{\gamma}_2', \mathbf{0}, \hat{\gamma}_4', \ldots, \mathbf{0})'$ since $\hat{\gamma}_{2i+1}' = 0$ for $i = 0, 1, \ldots, k/2$. Hence, we only need to focus on the behavior of $T^{1/2} \hat{\gamma}_{2i}$ for $i = 1, 2, \ldots, k/2$. Combining conditions (A.6) and (A.7), we get

$$\begin{aligned} \hat{\gamma}_{2i} &= \gamma_{2i}^0 + \left(Z_{2i}' Z_{2i} \right)^{-1} Z_{2i}' \eta_{2i}^* \left(\Pi_{2i}^0 - \hat{\Pi} \right) + \left(Z_{2i}' Z_{2i} \right)^{-1} Z_{2i}' V_{2i}^*, \\ \hat{\Pi} &= \left[\eta'\eta - \sum_{i=1}^{k/2} \left\{ \eta_{2i}^{*\prime} Z_{2i} \left(Z_{2i}' Z_{2i} \right)^{-1} Z_{2i}' \eta_{2i}^* \right\} \right]^{-1} \\ &\quad \times \left[\eta' \Delta Y - \sum_{i=1}^{k/2} \left\{ \eta_{2i}^{*\prime} Z_{2i} \left(Z_{2i}' Z_{2i} \right)^{-1} Z_{2i}' \Delta Y_{2i} \right\} \right]. \end{aligned} \tag{A.16}$$

It is easy to show that $(Z_{2i}' Z_{2i})^{-1} = O_p(T^{-1})$ and $Z_{2i}' V_{2i}^* = O_p(T^{1/2})$ given that regime $2i$ (for $i = 1, 2, \ldots, k/2$) is an $I(0)$ regime. Now, using results in Chang (1989) and Chang and Dickey (1994) assuming the condition $l_T^6 / T = o_p(1)$ holds, we have $||\eta' \Delta Y|| = O_p(T l_T^{1/2})$, $||Z_{2i}' \Delta Y_{2i}|| = O_p(T)$, $||\eta_{2i}^{*\prime} Z_{2i}|| = O_p(T)$, and $||[\eta'\eta - \sum_{i=1}^{k/2} \{\eta_{2i}^{*\prime} Z_{2i} (Z_{2i}' Z_{2i})^{-1} Z_{2i}' \eta_{2i}^*\}]^{-1}||_1 = O_p(l_T^2 T^{-1})$. Substituting in (A.16), we get $T^{1/2} \hat{\gamma}_{2i} = O_p(T^{1/2} l_T^{5/2})$ and hence $D_{1T}^{-1} \hat{\gamma} = O_p(T^{1/2} l_T^{5/2})$. Then, from (A.15), we have

$$SSR_0 - SSR_{1a,k} = \left(D_{1T}^{-1} \hat{\gamma} \right)' \left(D_{1T} \bar{Z}^{0\prime} M_\eta \bar{Z}^0 D_{1T} \right) D_{1T}^{-1} \hat{\gamma} = O_p \left(T l_T^5 \right). \tag{A.17}$$

Next, the denominator of $F_{1a}(\lambda, k)$ is

$$T^{-1}SSR_{1a,k} = T^{-1}\hat{V}^{*\prime}\hat{V}^{*} = T^{-1}\Delta Y' M_{\eta}\Delta Y - 2T^{-1}\Delta Y' M_{\eta}\bar{Z}^{0}\hat{\gamma} + T^{-1}\hat{\gamma}'\bar{Z}^{0\prime}M_{\eta}\bar{Z}^{0}\hat{\gamma}$$
$$= O_p(1) - O_p(1) + O_p\left(l_T^5\right) = O_p\left(l_T^5\right). \quad \textbf{(A.18)}$$

From (A.17) and (A.18), we therefore have $F_{1a}(\lambda^0, k) = O_p(T)$. This proves (a). Part (b) follows directly from (a) and the definition of the tests. For part (c), we focus on the simple AR(1) model with a single break for simplicity of exposition. We also abstract from short-run dynamics in the regression model so that the regressors included are only a constant and the lagged dependent variable. The proof for the more general model essentially follows the same steps although it is much more tedious and thus omitted. We assume that the true DGP is given by Model 1a and study the limit of $F_{1b}(\lambda, 1)$ for $\lambda \leq \lambda^0$ and $\lambda > \lambda^0$. We show that $F_{1b}(\lambda, 1) = O_p(1)$, uniformly over λ. First, consider the case where $\lambda \leq \lambda_0$. We have

$$SSR_0 = \sum_{t=1}^{T}\left(y_t - y_{t-1}\right)^2 = \sum_{t=1}^{[T\lambda]}\left(y_t - y_{t-1}\right)^2 + \sum_{t=[T\lambda]+1}^{T}\left(y_t - y_{t-1}\right)^2,$$
$$SSR_{1b,1} = \sum_{t=1}^{[T\lambda]}\left[y_t - \bar{y}_1 - \hat{\alpha}_1\left(y_{t-1} - \bar{y}_{1,-1}\right)\right]^2 + \sum_{t=[T\lambda]+1}^{T}\left(y_t - y_{t-1}\right)^2,$$
$$SSR_0 - SSR_{1b,1} = \sum_{t=1}^{[T\lambda]}\left(y_t - y_{t-1}\right)^2 - \sum_{t=1}^{[T\lambda]}\left[y_t - \bar{y}_1 - \hat{\alpha}_1\left(y_{t-1} - \bar{y}_{1,-1}\right)\right]^2$$
$$= [T\lambda]\bar{u}_1^2 - \left(1 - \hat{\alpha}_1\right)^2\sum_{t=1}^{[T\lambda]}\left(y_{t-1} - \bar{y}_{1,-1}\right)^2$$
$$-2\left(1 - \hat{\alpha}_1\right)\sum_{t=1}^{[T\lambda]}\left(y_{t-1} - \bar{y}_{1,-1}\right)u_t.$$

Using the facts that $\bar{u}_1 = O_p(T^{-1/2})$, $1 - \hat{\alpha}_1 = O_p(T^{-1})$, $\sum_{t=1}^{[T\lambda]}(y_{t-1} - \bar{y}_{1,-1})^2 = O_p(T^2)$, and $\sum_{t=1}^{[T\lambda]}(y_{t-1} - \bar{y}_{1,-1})u_t = O_p(T^{-1})$, we get $SSR_0 - SSR_{1b,1} = O_p(1)$. Based on similar arguments, we have $SSR_{1b,1} = O_p(T)$ so that $F_{1b}(\lambda, 1) = O_p(1)$ for $\lambda \leq \lambda_0$. For $\lambda > \lambda^0$,

$$SSR_0 - SSR_{1b,1} = \sum_{t=1}^{[T\lambda^0]}\left(y_t - y_{t-1}\right)^2 + \sum_{t=[T\lambda^0]+1}^{[T\lambda]}\left[c_2 + (\alpha_2 - 1)y_{t-1} + u_t\right]^2$$
$$- \sum_{t=1}^{[T\lambda^0]}\left[\left(y_t - \hat{\alpha}_1 y_{t-1}\right) - \left(\bar{y}_1 - \hat{\alpha}_1\bar{y}_{1,-1}\right)\right]^2$$
$$- \sum_{t=[T\lambda^0]+1}^{[T\lambda]}\left[\left(y_t - \hat{\alpha}_1 y_{t-1}\right) - \left(\bar{y}_1 - \hat{\alpha}_1\bar{y}_{1,-1}\right)\right]^2$$
$$= \sum_{t=1}^{[T\lambda^0]}{u_t}^2 + c_2^2 T\left(\lambda - \lambda^0\right)$$

$$
\begin{aligned}
&+(\alpha_2-1)^2 \sum_{t=[T\lambda^0]+1}^{[T\lambda]} y_{t-1}^2 + \sum_{t=[T\lambda^0]+1}^{[T\lambda]} u_t^2 + 2c_2(\alpha_2-1) \sum_{t=[T\lambda^0]+1}^{[T\lambda]} y_{t-1} \\
&+2c_2 \sum_{t=[T\lambda^0]+1}^{[T\lambda]} u_t + 2(\alpha_2-1) \sum_{t=[T\lambda^0]+1}^{[T\lambda]} y_{t-1}u_t \\
&- \sum_{t=1}^{[T\lambda^0]} \left[\left(1-\hat{\alpha}_1\right) y_{t-1} + u_t - \left(\bar{y}_1 - \hat{\alpha}_1 \bar{y}_{1,-1}\right)\right]^2 \\
&- \sum_{t=[T\lambda^0]+1}^{[T\lambda]} \left[c_2 + \left(\alpha_2 - \hat{\alpha}_1\right) y_{t-1} + u_t - \left(\bar{y}_1 - \hat{\alpha}_1 \bar{y}_{1,-1}\right)\right]^2 \\
= &(\alpha_2-1)^2 \sum_{t=[T\lambda^0]+1}^{[T\lambda]} y_{t-1}^2 - \left(\alpha_2 - \hat{\alpha}_1\right)^2 \\
&\times \sum_{t=[T\lambda^0]+1}^{[T\lambda]} y_{t-1}^2 + 2c_2\left(\hat{\alpha}_1 - 1\right) \sum_{t=[T\lambda^0]+1}^{[T\lambda]} y_{t-1} \\
&+2(\hat{\alpha}_1-1) \sum_{t=[T\lambda^0]+1}^{[T\lambda]} y_{t-1}u_t + 2\left(\bar{y}_1 - \hat{\alpha}_1 \bar{y}_{1,-1}\right) \\
&\times \left(\sum_{t=1}^{[T\lambda^0]} u_t + \sum_{t=[T\lambda^0]+1}^{[T\lambda]} u_t\right) - \left(1-\hat{\alpha}_1\right)^2 \sum_{t=1}^{[T\lambda^0]} y_{t-1}^2 \\
&-2\left(1-\hat{\alpha}_1\right) \sum_{t=1}^{[T\lambda^0]} y_{t-1}u_t + 2\left(1-\hat{\alpha}_1\right)\left(\bar{y}_1 - \hat{\alpha}_1 \bar{y}_{1,-1}\right) \\
&\times \sum_{t=1}^{[T\lambda^0]} y_{t-1} - [T\lambda]\left(\bar{y}_1 - \hat{\alpha}_1 \bar{y}_{1,-1}\right)^2 \\
&+2\left(\bar{y}_1 - \hat{\alpha}_1 \bar{y}_{1,-1}\right) T\left(\lambda - \lambda^0\right)\left(\alpha_2 - \hat{\alpha}_1\right) \left(-c_2\left(\hat{\alpha}_1 - \alpha_2\right)^{-1}\right. \\
&\left. + \left[T\left(\lambda - \lambda^0\right)\right]^{-1} \sum_{t=[T\lambda^0]+1}^{[T\lambda]} y_{t-1}\right). \qquad \textbf{(A.19)}
\end{aligned}
$$

Now we have $\sum_{t=1}^{[T\lambda^0]} y_{t-1}^2 = O_p(T^2)$, $\hat{\alpha}_1 = 1 + O_p(T^{-1})$, $\bar{y}_1 - \hat{\alpha}_1\bar{y}_{1,-1} = (1-\hat{\alpha}_1)\bar{y}_{1,-1} + O_p(T^{-1}) = O_p(T^{-1/2})$, $\sum_{t=[T\lambda^0]+1}^{[T\lambda]} y_{t-1}^2 = [T(\lambda-\lambda^0)]O_p(1)$, $\sum_{t=[T\lambda^0]+1}^{[T\lambda]} y_{t-1}u_t = [T(\lambda-\lambda^0)]^{1/2}O_p(1)$, $\sum_{t=[T\lambda^0]+1}^{[T\lambda]} u_t = [T(\lambda-\lambda^0)]^{1/2}O_p(1)$. Note that the last term in brackets can be expressed as $2(\bar{y}_1 - \hat{\alpha}_1\bar{y}_{1,-1})(\alpha_2-\hat{\alpha}_1)\sum_{t=[T\lambda^0]+1}^{[T\lambda]}(y_{t-1} - c_2/(\hat{\alpha}_1-\alpha_2)) = O_p(T^{-1/2})O_p(1)O_p(T^{1/2}) = O_p(1)$. Substituting these orders in (A.19), each of the terms in brackets is $O_p(1)$. Again, similar arguments can be used to show that $SSR_{1b,1} = O_p(T)$ and thus $F_{1b}(\lambda,1) = O_p(1)$. ■