Purdue University Economics Department Working Paper No# **1303**

A Robust Sequential Procedure for Estimating the Number of Structural Changes in Persistence*

Mohitosh Kejriwal†

Purdue University

December 9, 2017

Abstract

This paper proposes a new procedure for estimating the number of structural changes in the persistence of a univariate time series. In contrast to the extant literature that primarily assumes (regime-wise) stationarity, our framework also allows the underlying stochastic process to switch between stationary [$I(0)$] and unit root [$I(1)$] regimes. We develop a sequential testing approach based on the simultaneous application of two Wald-type tests for structural change, one of which assumes the process is $I(0)$-stable under the null hypothesis while the other assumes the stable $I(1)$ model as the null hypothesis. This feature allows the procedure to maintain correct asymptotic size regardless of whether the regimes are $I(0)$ or $I(1)$. We also propose a novel procedure for distinguishing processes with pure level and/or trend shifts from those that are also subject to concurrent shifts in persistence. The large sample properties of the recommended procedures are derived and the relevant asymptotic critical values tabulated. Our Monte Carlo experiments demonstrate that the advocated approach compares favorably relative to the commonly employed approach based on $I(0)$ sequential testing, especially when the data contain an $I(1)$ segment.

Keywords: multiple structural changes, unit root, stationary, sequential procedure, Wald tests

JEL Classification: C22

*I am grateful to Anil Bera, Jihyung Lee and participants at the UIUC Econometrics Seminar Series for comments and discussion. Any errors are my own.

†Krannert School of Management, Purdue University, 403 West State Street, West Lafayette IN 47907 (mkejriwa@purdue.edu).

1 Introduction

The twin problems of testing for and estimation of structural changes in time series models have generated a vast literature in both econometrics and statistics (see Perron, 2006, for a survey). While early approaches to these issues were primarily based on the assumption of independently and identically distributed data (e.g., Page, 1955; Quandt, 1960), substantial advances over the past two decades have made possible the development of methods that are applicable to a wide range of data generating processes. A notable feature of such developments is that they allow different types of dependence structures, e.g., models with stationary or mixing processes, long memory, unit roots, cointegration and deterministic trends. The empirical relevance of structural change in the context of modeling economic series has also been well documented (e.g., Stock and Watson, 1996).

A particularly important issue that has received considerable attention has been the intricate interplay between structural change and unit roots. Perron (1989) argued that it is difficult to distinguish a unit root process from one that is subject to infrequent changes in its trend function but is otherwise stationary within regimes specified by the break dates. In particular, he showed that unit root tests directed against the (trend) stationary alternative (e.g., Dickey and Fuller, 1979; Phillips and Perron, 1988) are biased in favor of the unit root model if the true data generating process is stationary around a broken deterministic trend. Similarly, most structural change detection procedures that do not account for the presence of unit roots would tend to favor the alternative hypothesis of structural change if a unit root is indeed present but the model parameters are constant (Perron, 2006). When testing for the presence of one of these features, therefore, it is a prudent approach to allow for the possibility of the other.

Determining the number of structural changes is a crucial component of empirical analysis from the viewpoint of model selection. Conditional on the estimated number of breaks, the economic significance of the differences between regime-specific parameters can be assessed thereby providing evidence regarding the relative degree of instability associated with each of the changes. An early contribution in this regard is by Yao (1988) who proposed choosing the number of breaks by minimizing the Bayesian Information Criterion (BIC). Weak consistency of the estimator was established for a sequence of independent and identically distributed Gaussian random variables. Liu et al. (1997) suggest a modified BIC based on a more severe penalty to cover the non-Gaussian case. Bai and Perron (1998, BP henceforth) propose a sequential testing procedure in a general regression framework based on the sup-Wald test for

structural change (Andrews, 1993) that involves successively applying the test to evaluate the null hypothesis of, say, l changes against the alternative of $l+1$ changes starting with $l = 0$.[1] The estimate of the number of breaks thus obtained is shown to be consistent provided the significance level of the test decreases to zero at an appropriate rate as the sample size increases. Monte Carlo evidence presented in Bai and Perron (2006) show that the sequential testing approach dominates information criteria-based selection when the data generating process includes at least one break. Altissimo and Corradi (2003) recommend an alternative sequential procedure that consists in testing for a single break based on the maximum of the absolute value of the partial sums of demeaned data. They derive the asymptotic critical value of the proposed test that forms the basis for their decision rule for rejecting the null hypothesis of no change at each step. For the case of multiple changes in mean, the method is shown to yield a strongly consistent estimate of the number of breaks.

An assumption common to all the aforementioned methods of break selection is one of short memory or stationarity [referred to as $I(0)$, henceforth]. In the univariate context, this implies that the process is either $I(0)$ over the full sample (in the no break case) or $I(0)$ within regimes specified by the break dates. While convenient in theory, it rules out the possibility of a unit root [referred to as $I(1)$, henceforth] in a subsample of the data or over the whole sample. The importance of allowing for such nonstationarity can again be traced back to the argument in Perron (1989) so that ignoring the possible presence of a unit root can generate spurious breaks thereby resulting in an inconsistent estimate of the number of breaks. For instance, Kejriwal and Perron (2010a) show that break selection procedures that assume cointegration among a set of variables tend to select the maximum number of breaks allowed when the regression is spurious (see also Perron, 2006).

A plethora of procedures now exist for testing changes in persistence that allow for the possibility of a unit root under the null and/or alternative hypotheses. 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 et al. (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 et al. (2003) consider testing the null hypothesis of a stable unit root process versus the

[1] A similar approach was proposed by Kejriwal and Perron (2010a) for estimating the number of changes in single-equation cointegrated models and by Kejriwal and Perron (2010b) for determining the number of trend breaks with stationary or unit root innovations.

same alternatives based on the minimal value of the locally GLS detrended augmented Dickey-Fuller (ADF) unit root statistic developed in Elliott et al. (1996) over sub-samples of the data. They propose different test statistics depending on whether the initial regime is $I(1)$ or $I(0)$. Kurozumi (2005) suggests an alternative testing procedure based on the Lagrange Multiplier (LM) principle while Leybourne et al. (2007a) develop tests of the unit root null based on standardized cumulative sums of squared sub-sample residuals that do not spuriously reject when the series is a constant $I(0)$ process. The foregoing procedures are all designed to test for a single break in persistence. As shown in Bai and Perron (2006), single break tests can suffer from serious power deficiencies when the alternative hypothesis involves multiple breaks. Leybourne et al. (2007b) develop tests of the unit root null hypothesis that can accommodate multiple changes in persistence under the alternative and is capable of consistently partitioning the data into its separate $I(0)$ and $I(1)$ regimes. Their procedure is based on doubly-recursive sequences of ADF-type statistics and associated estimators of the break dates. Monte Carlo evidence presented in Kejriwal et al. (2013, KPZ henceforth), however, shows the procedure to be subject to considerable size distortions when the process is stable $I(0)$ or when the process is stable $I(1)$ with serially correlated errors. KPZ propose procedures for detecting multiple persistence breaks based on sup-Wald tests when the process can be either $I(1)$ or $I(0)$ under the null hypothesis although their analysis assumes the number of breaks to be known a priori. Based on the preceding discussion, it appears relevant from a practical perspective to develop a procedure for break selection that allows the process to be either $I(1)$ or $I(0)$ in the stable case and also remains valid regardless of whether the structural changes preserve the $I(0)$ nature of the process in each regime or involve switches between $I(0)$ and $I(1)$ regimes.

This paper proposes a new sequential procedure for estimating the number of breaks in the persistence of a univariate time series that is robust to the presence of a unit root over the full sample or in any (asymptotically non-negligible) subsample of the data. The procedure is based on simultaneous application of two Wald-type tests for structural change, one of which assumes the process is $I(0)$-stable under the null hypothesis while the other assumes the stable $I(1)$ model as the null hypothesis. In particular, we use the BP test for the former and the sup-F test proposed in KPZ for the latter. Using the intersection of the two critical regions as the relevant critical region enables the procedure to maintain correct asymptotic size regardless of whether the regimes are $I(0)$ or $I(1)$. The procedure starts by testing the null hypothesis of no structural change against the alternative of an unknown number of changes (subject to an upper bound). Upon a rejection, the single break test is applied

successively to segments of the data determined by estimating the break dates obtained from minimizing the global sum of squared residuals. At each step, the null hypothesis of, say, l breaks against the alternative of $l+1$ breaks is rejected if the maximum of the single break tests computed over the $(l+1)$ segments is significant. The procedure continues until a non-rejection occurs and the estimated number of breaks is determined as the number of rejections obtained until then.

Our theoretical results are two-fold. First, we establish that the BP test has incorrect asymptotic size under the null hypothesis of a unit root suggesting that the BP procedure employed in isolation is likely to over-estimate the number of breaks if a unit root is present. Second, the large sample properties of our procedure are derived in the general framework that does not restrict the process to be $I(0)$ within regimes but allows for $I(1)$ non-stationarities. Both trending and non-trending cases are analyzed. The relevant asymptotic critical values are shown to be obtained from the appropriate quantiles of the single break limit distributions and tabulated for a range of trimming values. The limit distribution of the BP test is derived for the trending case given that the asymptotics in BP were obtained under the assumption of $I(0)$ regressors. We also discuss how our procedure can be employed to address the important practical issue of distinguishing processes with pure level and/or trend breaks from those that are characterized by concurrent shifts in persistence as well. Our Monte Carlo experiments demonstrate that the advocated approach compares favorably relative to the commonly employed BP approach based on $I(0)$ sequential testing, especially when the data contain an $I(1)$ segment.

The rest of the paper is organized as follows. Section 2 lays out the persistence change model and the associated assumptions. The limit distribution of the BP test under the null hypothesis of a unit root is derived in section 3. Section 4 develops the proposed sequential testing procedure and its large sample properties. Section 5 considers a modification of the analysis to allow for deterministic trends. Section 6 suggests procedures for distinguishing processes with pure level and/or trend breaks from those that are subject to breaks in persistence as well. The extension to serially correlated errors is considered in section 7. Details regarding the computation of the asymptotic critical values are provided in section 8. Section 9 presents Monte Carlo evidence to assess the adequacy of the asymptotic approximations and evaluate the merits of the proposed approach relative to $I(0)$-sequential testing. Section 10 concludes. All proofs are included in a Technical Appendix.

As a matter of notation, we let $\xrightarrow{p}$ denote convergence in probability, $\xrightarrow{d}$ convergence in distribution, and "$\Rightarrow$" weak convergence of the associated probability measures. Let $B_1(.)$ and

$B_2(.)$ denote standard independent Brownian motions on $[0,1]$ and $B(.) = [B_1(.), B_2(.)]'$. Further, let $\widetilde{B}_j^{(i)}(.)$ represent $B_j(.)$ demeaned over $[\lambda_{i-1}, \lambda_i]$, i.e., $\widetilde{B}_j^{(i)}(r) = B_j(r) - (\lambda_i - \lambda_{i-1})^{-1} \int_{\lambda_{i-1}}^{\lambda_i} B_j,\ r \in [\lambda_{i-1}, \lambda_i]$. The Brownian motions demeaned over the full sample are denoted as $\widetilde{B}_j(.) = B_j(.) - \int_0^1 B_j$. Finally, for brevity of presentation, all integrals of the form $\int_a^b g(r)dr$ are expressed as $\int_a^b g$.

2 The Persistence Change Model

Consider a univariate time series y_t generated as

$$\left.\begin{array}{c} y_t = \mu_i + u_t \\ u_t = u_{T_{i-1}^0} + h_t \\ h_t = \alpha_i h_{t-1} + e_t \\ h_{T_{i-1}^0} = 0 \end{array}\right\} \; t = T_{i-1}^0 + 1, T_{i-1}^0 + 2, ..., T_i^0;\ i = 1, ..., m+1 \tag{1}$$

with the convention that $T_0^0 = 0$ and $T_{m+1}^0 = T$, where T is the sample size. The process is therefore subject to m breaks or $m+1$ regimes with break dates $(T_1, ..., T_m)$. Both the break dates and the number of breaks are assumed to be unknown. The same data generating process was considered by Leybourne et al. (2007b) and is designed to ensure that the successive $I(1)$ and $I(0)$ regimes join up at the breakpoints thereby avoiding the problem of spurious jumps to zero in u_t. We make the following assumption on the break dates and the noise component e_t:

Assumption A1: $T_i^0 = [T\lambda_i^0]$, where $0 < \lambda_1^0 < ... < \lambda_m^0 < 1$.

Assumption A2: The process $\{e_t\}$ is a martingale difference sequence with respect to $\{\mathcal{F}_t\}$, where $\mathcal{F}_t = \sigma\text{-}field\{e_t, t \leq s\}$ with $E(e_t^2|\mathcal{F}_{t-1}) = \sigma^2$ and $\sup_t\ E(|e_t|^{4+\beta}\,|\mathcal{F}_{t-1}) < \infty$ for some $\beta > 0$.

Assumption A1 allows the development of the asymptotic theory by requiring the breakpoints to be asymptotically distinct. Each segment is assumed to increase proportionately with the sample size. This requirement is standard in the structural change literature (see, e.g., Bai and Perron, 1998; 2003a). Assumption 2 rules out serial correlation in the innovation sequence and requires conditional homoskedasticity. The case with serial correlation where e_t follows a general linear process will be considered in section 7. The assumption of conditional homoskedasticity, albeit restrictive, facilitates the development of nuisance-parameter free asymptotic distributions. The assumption is relaxed in a companion paper

(Kejriwal and Yu, 2017) where a wild bootstrap approach is employed to account for the heteroskedastic nature of the error process.

From (1), we can write

$$y_t = c_i + \alpha_i y_{t-1} + e_t \tag{2}$$

where $c_i = (\mu_i - \mu_{i-1} + y_{T^0_{i-1}})(1 - \alpha_i)$. KPZ consider tests of the null hypothesis $H_0^{(1)}$: $\alpha_i = 1$ for all i. [We use the notation $H_0^{(a)}$ to denote the $I(a)$ null hypothesis, $a = 0, 1$]. Note that under $H_0^{(1)}$, $c_i = 0$ for all i so that the time series follows a stable unit root process. Under the alternative hypothesis of unstable persistence, the following two models are considered depending on whether the initial regime contains a unit root or not:

Model 1a: $\alpha_i = 1$ in odd regimes and $|\alpha_i| < 1$ in even regimes.

Model 1b: $\alpha_i = 1$ 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. We denote the corresponding alternative hypotheses as $H_{1a,k}^{(1)}$ and $H_{1b,k}^{(1)}$, respectively.

KPZ consider a variety of tests of $H_0^{(1)}$. First, 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

$$\begin{aligned} F_{1a}(\lambda, k) &= (T-k)(SSR_0^{(1)} - SSR_{1a,k})/[kSSR_{1a,k}^{(1)}] \text{ if } k \text{ is even} \\ F_{1a}(\lambda, k) &= (T-k-1)(SSR_0^{(1)} - SSR_{1a,k})/[(k+1)SSR_{1a,k}^{(1)}] \text{ if } k \text{ is odd} \end{aligned} \tag{3}$$

$$\begin{aligned} F_{1b}(\lambda, k) &= (T-k-2)(SSR_0^{(1)} - SSR_{1b,k}^{(1)})/[(k+2)SSR_{1b,k}^{(1)}] \text{ if } k \text{ is even} \\ F_{1b}(\lambda, k) &= (T-k-1)(SSR_0^{(1)} - SSR_{1b,k}^{(1)})/[(k+1)SSR_{1b,k}^{(1)}] \text{ if } k \text{ is odd} \end{aligned} \tag{4}$$

In (3) and (4), $SSR_0^{(1)}$ denotes the sum of squared residuals under $H_0^{(1)}$ while $SSR_{1a,k}^{(1)}$ and $SSR_{1b,k}^{(1)}$ denote, respectively, the sum of squared residuals obtained from estimating (2) under the restrictions imposed by Model 1a and 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 $F_{1a}(k) = \sup_{\lambda \in \Lambda_\epsilon^k} F_{1a}(\lambda, k)$ and $F_{1b}(k) = \sup_{\lambda \in \Lambda_\epsilon^k} F_{1b}(\lambda, k)$.

The second type of tests 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 is $I(0)$ or $I(1)$. The test is given by $W_1(k) = \max[F_{1a}(k), F_{1b}(k)]$. Finally,

in order to accommodate the case with an unknown number of breaks, up to some maximal value A, we consider the statistic $Wmax_1 = \max_{1\leq k\leq A} W_1(k)$.

While the foregoing tests are based on the $I(1)$ null hypothesis, the stable $I(0)$ null can be tested by employing the BP procedure. Specifically, consider testing $H_0^{(0)}$: $c_i = c,\ \alpha_i = \alpha$, for all i with $|\alpha| < 1$ in the model

$$y_t = c_i + \alpha_i y_{t-1} + e_t \tag{5}$$

with $c_i = \mu_i(1-\alpha_i)$. The relevant alternative hypothesis within the BP framework is $H_{1,k}^{(0)}$: $\alpha_1 \neq \alpha_2 \neq ... \neq \alpha_{k+1}$, $|\alpha_i| < 1$ for all i. The time series is thus regimewise-$I(0)$ under $H_{1,k}^{(0)}$. The BP test for a fixed number $m = k$ changes is given by

$$G_1(\lambda, k) = [T - 2(k+1)](SSR_0^{(0)} - SSR_{1,k}^{(0)})/[kSSR_{1,k}^{(0)}] \tag{6}$$

In (6), $SSR_0^{(0)}$ denotes the sum of squared residuals under $H_0^{(0)}$ while $SSR_{1,k}^{(0)}$ denotes the sum of squared residuals obtained from *unrestricted* OLS estimation of (5). The BP test is then defined as $G_1(k) = \sup_{\lambda\in\Lambda_\epsilon^k} G_1(\lambda, k)$. When the number of breaks is unknown, the relevant test statistic is $UDmax_1 = \max_{1\leq k\leq A} G_1(k)$.

Remark 1 *The limiting distributions of the KPZ tests are pivotal under $H_0^{(1)}$ while those of the BP tests are pivotal under $H_0^{(0)}$, except for the trimming choice ϵ. The asymptotic critical values for both sets of tests can be obtained through Monte Carlo simulation. With $m = k$, the tests $W_1(k)$, $Wmax_1$, $G_1(k)$, $UDmax_1$ are each consistent under $H_{1a,k}^{(1)}$, $H_{1b,k}^{(1)}$ and $H_{1,k}^{(0)}$[see KPZ for details].*

Remark 2 *To compute the KPZ tests, 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). Similarly, the computation of the BP tests is based on the unrestricted minimization of the global sum of squared residuals using the dynamic programming algorithm developed in Bai and Perron (1998, 2003a).*

3 The Bai-Perron Test under the Unit Root Null Hypothesis

If assumptions A1 and A2 hold, the limit distribution of $G_1(k)$ and $UDmax_1$ under the stable $I(0)$ null hypothesis $H_0^{(0)}$ follows from Proposition 6 in BP:

$$\begin{aligned} G_1(k) &\Rightarrow (1/k) \sup_{\lambda\in\Lambda_\epsilon^k} \sum_{i=1}^{k} \frac{[\lambda_i B(\lambda_{i+1}) - \lambda_{i+1}B(\lambda_i)]'[\lambda_i B(\lambda_{i+1}) - \lambda_{i+1}B(\lambda_i)]}{\lambda_{i+1}\lambda_i(\lambda_{i+1} - \lambda_i)} \overset{def}{=} G_1^*(k) \\ UDmax_1 &\Rightarrow \max_{1\leq k\leq A} G_1^*(k) \end{aligned} \tag{7}$$

For $k = 1$, we recover the limit distribution derived in Andrews (1993):

$$G_1(1) \Rightarrow \sup_{\lambda \in \Lambda_\epsilon^1} \frac{[\lambda B(1) - B(\lambda)]'[\lambda B(1) - B(\lambda)]}{\lambda(1-\lambda)} \tag{8}$$

The limit in (8) represents the supremum of the square of a tied-down Bessel process of order 2. The asymptotic distribution of the BP tests are therefore independent of any nuisance parameters except the trimming choice ϵ. The relevant critical values for a range of trimming choices can be obtained from Bai and Perron (2003b).

The following proposition states the limit distribution of the BP test under the stable $I(1)$ null hypothesis:

Proposition 1 *Suppose Assumptions A1-A2 hold. Under* $H_0^{(1)}$, *we have*

$$\begin{aligned} G_1(k) &\Rightarrow (1/k) \sup_{\lambda \in \Lambda_\epsilon^k} \left\{ \begin{array}{c} \sum_{i=1}^{k} \frac{[\lambda_i B_1(\lambda_{i+1}) - \lambda_{i+1} B_1(\lambda_i)]^2}{\lambda_{i+1}\lambda_i(\lambda_{i+1}-\lambda_i)} \\ -\frac{\left(\int_0^1 \widetilde{B}_1 dB_1\right)^2}{\int_0^1 \widetilde{B}_1^2} + \sum_{i=1}^{k+1} \frac{\left[\int_{\lambda_{i-1}}^{\lambda_i} \widetilde{B}_1^{(i)} dB_1\right]^2}{\int_{\lambda_{i-1}}^{\lambda_i} \left[\widetilde{B}_1^{(i)}\right]^2} \end{array} \right\} \stackrel{def}{=} \widetilde{G}_1(k) \\ UDmax_1 &\Rightarrow \max_{1 \le k \le A} \widetilde{G}_1(k) \end{aligned} \tag{9}$$

The limit distribution in (9) is different from the one in (7). While the first component in (9) corresponds to the intercept term that is common to both distributions, the second and third terms in (9) now represent the contribution from the $I(1)$ component. The second component is the squared limit of the Dickey-Fuller (1979) unit root test over the full sample when a constant is included in the regression. For the third component, the i-th term inside the summation represents the squared limit of the Dickey-Fuller unit root test when computed over the i-th regime. This result indicates that under the $I(1)$ null hypothesis the asymptotic level of the BP test is different from the corresponding nominal level.

The degree of size distortions in large samples can be quantified by computing the asymptotic size through Monte Carlo simulation. Specifically, the Brownian motion $B_1(.)$ is approximated by partial sums of i.i.d. standard normal random variables with step size equal to 1000. Based on 10,000 Monte Carlo replications, the asymptotic size for a given significance level is computed as the fraction of replications in which $\widetilde{G}_1(k)$ exceeds the corresponding BP critical value. For $k = 1$, this value was found to be 67%, 52% and 22% for nominal levels of 10%, 5% and 1% respectively. The 90th, 95th and 99th percentiles of $\widetilde{G}_1(1)$ were obtained as 18.00, 20.25 and 24.49, respectively. The corresponding percentiles of $G_1^*(1)$ are 9.81, 11.47 and 15.37, respectively (Bai and Perron, 2003b). Since the BP procedure for break selection is based on sequential application of the single break BP test, the

considerable size distortions in the unit root case are likely to manifest themselves through overestimation of the number of breaks in finite samples. This is true not only when the process is stable $I(1)$ but also when any asymptotically negligible subsample of the data are $I(1)$. We confirm that this is indeed the case in the Monte Carlo experiments conducted in section 9.

Remark 3 *Under the local to unity parameterization* $\alpha_T = 1 + c/T,\ c < 0$, *the Weiner process* $B_1(.)$ *in* (9) *is replaced by the Ornstein-Uhlenbeck process* $J_c(.)$ *generated in continuous time by the stochastic differential equation* $dJ_c(r) = cJ_c(r) + dB_1(r)$, *with initial condition* $J_c(0) = 0$ *(Phillips and Perron, 1988). This result can explain the size distortions associated with the BP procedure when* α_T *is close to but not equal to one (see section 9).*

4 A Robust Sequential Procedure

The analysis in section 3 shows that since the BP procedure for break selection is based on the stable $I(0)$ null hypothesis, it is not robust to the presence of a unit root. On the other hand, the KPZ tests described in section 2 are based on the stable-$I(1)$ null hypothesis and diverge to positive infinity when the process is stable-$I(0)$. In order to control asymptotic size when the process is either $I(1)$ or $I(0)$ under the null hypothesis, KPZ propose simultaneous application of their test and the BP test. In other words, they recommend choosing the intersection of the two critical regions as the relevant critical region for their robust procedure. Denote by H_0 the null hypothesis that the process is either $I(1)$-stable or $I(0)$-stable. For a fixed number of breaks k, KPZ recommend employing the following decision rule:

$$\text{“Reject } H_0 \text{ if both } W_1(k) \text{ and } G_1(k) \text{ reject”} \tag{10}$$

When the number of breaks is unknown, the relevant decision rule takes the form

$$\text{“Reject } H_0 \text{ if both } Wmax_1 \text{ and } UDmax_1 \text{ reject”} \tag{11}$$

Consider using decision rule (10) with a given nominal level η for each of $W_1(k)$ and $G_1(k)$. Since $W_1(k)$ and $G_1(k)$ are correctly sized under $H_0^{(1)}$ and $H_0^{(0)}$, respectively, it follows that the decision rule (10) has asymptotic size at most η under H_0. A similar argument holds for (11). Define the following test statistics:

$$\begin{aligned} H_1(k,\eta) &= \min\left[W_1(k), \frac{cv_{w,k}(\eta)}{cv_{g,k}(\eta)}G_1(k)\right] \\ Hmax_1(\eta) &= \min\left[Wmax_1, \frac{cv_{w,max}(\eta)}{cv_{g,max}(\eta)}UDmax_1\right] \end{aligned}$$

where, at level η, $cv_{w,k}(\eta)$, $cv_{g,k}(\eta), cv_{w,max}(\eta)$, $cv_{g,max}(\eta)$ are the critical values of the tests $W_1(k)$, $G_1(k)$, $Wmax_1$ and $UDmax_1$, respectively. It is straightforward to verify that the decision rule (10) is equivalent to the decision rule

$$\text{"Reject } H_0 \text{ if } H_1(k,\eta) > cv_{w,k}(\eta)\text{"} \tag{12}$$

while the decision rule (11) is equivalent to

$$\text{"Reject } H_0 \text{ if } Hmax_1(\eta) > cv_{w,max}(\eta)\text{"} \tag{13}$$

We now develop a sequential test of the null hypothesis of l breaks against the alternative hypothesis of $(l+1)$ breaks in equation (2). First, we obtain the estimates of the break dates $\hat{T}_1, ..., \hat{T}_l$ as global minimizers of the sum of squared residuals from the *unrestricted* model with l breaks estimated by least squares:

$$(\hat{T}_1, ..., \hat{T}_l) = \text{argmin}_{(T_1,...,T_l)} SSR_{1,l}^{(0)}$$

This can be achieved using the dynamic programming algorithm proposed by Bai and Perron (2003a). Next, we test for the presence of an additional break in each of the $(l+1)$ segments obtained using the estimated partition $(\hat{T}_1, ..., \hat{T}_l)$. Define the quantities

$$\begin{aligned} \eta_{l+1} &= 1-(1-\eta)^{1/(l+1)} \\ H_1^{(i)}(1,\eta_{l+1}) &= \min\left[W_1^{(i)}(1), \frac{cv_{w,1}(\eta_{l+1})}{cv_{g,1}(\eta_{l+1})} G_1^{(i)}(1)\right] \\ H_1(l+1|l) &= \max_{1\le i\le l+1}\{H_1^{(i)}(1,\eta_{l+1})\} \end{aligned}$$

where $W_1^{(i)}(1)$ denotes the $W_1(1)$ test computed using data in the (estimated) regime i, i.e., $[\hat{T}_{i-1}+1, \hat{T}_i]$ and $G_1^{(i)}(1)$ denotes the $G_1(1)$ test computed using the observations in $[\hat{T}_{i-1}+1, \hat{T}_i]$. We conclude in favor of a model with $(l+1)$ breaks if

$$H_1(l+1|l) > cv_{w,1}(\eta_{l+1}) \tag{14}$$

The test thus amounts to the computation of $(l+1)$ tests of the null hypothesis of no change versus the alternative hypothesis of a single change and assessing whether their maximum is sufficiently large. The threshold value $cv_{w,1}(\eta_{l+1})$ is the $(1-\eta_{l+1})$-th quantile of the limit distribution of $W_1(1)$. The following result shows that the decision rule (14) has asymptotic size at most η under the null hypothesis of l breaks:

Theorem 1 *Suppose Assumptions A1-A2 hold. Under the null hypothesis that the true number of breaks is l, we have* $\lim_{T\to\infty} P(H_1(l+1|l) > cv_{w,1}(\eta_{l+1})) \leq \eta$.

The test based on $H_1(l+1|l)$ can be used to provide an estimate of the number of breaks $\hat{m}$ in the following way:

1. First, apply the decision rule (13) that tests the null hypothesis H_0 against an unknown number of breaks. If a non-rejection is obtained, set $\hat{m} = 0$ and the procedure stops. Otherwise, go to step 2.

2. Upon a rejection in step 1, use the rule (14) with $l = 1$ to determine if there is more than one break. This process is repeated by increasing l sequentially until the test fails to reject the null hypothesis of no additional structural breaks.

3. The estimate $\hat{m}$ is then obtained as the number of rejections.

The following result shows that the proposed algorithm guarantees that the probability of selecting the true number of breaks is at least $(1-\eta)$ in large samples:

Theorem 2 *Suppose Assumptions A1-A2 hold. Let m be the true number of breaks and $\hat{m}$ be the number of breaks obtained using the proposed sequential procedure based on the test statistics $Hmax_1$ and $H_1(l+1|l)$ $[l \geq 1]$ applied with nominal level η. Then $lim_{T\to\infty}$ $P(\hat{m} = m) \geq 1-\eta$.*

The sequential procedure can be made consistent by allowing the significance level of the tests $Hmax_1$ and $H_1(l+1|l)$ to decrease to zero at a suitable rate as the sample size increases. It can be shown that if the true number of breaks is at least $l+1$, $Hmax_1$ and $H_1(l+1|l)$ diverge at rate $O_p(T)$. Thus, if the critical values $cv_{w,1},\ cv_{g,1},\ cv_{w,max},\ cv_{g,max}$ are allowed to increase at rate $O_p(T^{1-\varepsilon})$, $0 < \varepsilon < 1$, the size of the tests converges to zero at T increases while ensuring their consistency under the alternative. We thus have the following corollary whose proof is similar to that of Hosoya (1988) and is thus omitted:

Corollary 1 *Let m be the true number of breaks and $\hat{m}$ be the number of breaks obtained using the sequential procedure based on the test statistic $H_1(l+1|l)$ applied with significance level η_T. Consider a sequence of critical values $cv_{w,1} = c_1 T^{1-\varepsilon},\ cv_{g,1} = c_2 T^{1-\varepsilon},\ cv_{w,max} = c_{max} T^{1-\varepsilon}$ ($c_1,\ c_2,\ c_{max}$ are positive constants) so that η_T converges to zero while ensuring that $H_1(l+1|l)$ remains consistent. Then, under Assumptions A1-A2, $P(\hat{m} = m) \to 1$ as $T \to \infty$.*

Remark 4 *A purely sequential approach would involve testing for* $l = 0$ *versus* $l = 1$ *based on the* $H_1(1|0)$ *statistic in the first step. However, as is well known, if the number of breaks assumed in the construction of the test statistic is smaller than the true number of breaks under the alternative hypothesis, the test can be subject to serious non-monotonic power problems [see Bai and Perron, 2006; Perron, 2006]. Using the rule* (13) *in the first step provides a safeguard against this issue.*

Remark 5 *Theorem 2 and Corollary 1 hold not only under the alternative hypotheses* $H^{(1)}_{1a,m}$, $H^{(1)}_{1b,m}$ *and* $H^{(0)}_{1,m}$ *but also under more general alternatives where the process involves a mix of* $I(1)$ *and* $I(0)$ *regimes with possibly adjacent* $I(0)$ *regimes.*

Remark 6 *While Theorem 2 suggests that the probability of break selection can be guaranteed to exceed any pre-assigned value by choosing a sufficiently small* η*, it must be borne in mind that it is a large sample result that uses the fact that the test* $H_1(l+1|l)$ *is consistent under the alternative hypothesis of at least* $(l+1)$ *breaks. In finite samples, however, the power of the test depends on the significance level used and using too small a level can lead to underestimating the true break number.*

5 Deterministic Trends

This section discusses how the sequential procedure proposed in section 4 can be extended to allow for the presence of deterministic trends. We consider an extension of (1) that includes the possibility of m breaks in the deterministic trend:

$$\left.\begin{array}{c} y_t = \mu_0 + \beta_0 t + \sum_{j=1}^{m} \mu_j DU_{jt} + \sum_{j=1}^{m} \beta_j DT_{jt} + u_t \\ u_t = u_{T^0_{i-1}} + h_t \\ h_t = \alpha_i h_{t-1} + e_t \\ h_{T^0_{i-1}} = 0 \end{array}\right\} \begin{array}{c} t = T^0_{i-1}+1, T^0_{i-1}+2, ..., T^0_i; \\ i = 1, ..., m+1 \end{array} \quad (15)$$

where $DU_{jt} = I(t > T^0_j)$, $DT_{jt} = I(t > T^0_j)(t - T^0_j)$; $j = 1, ..., m$. The data generating process (15) can be expressed as

$$y_t = c_i + b_i t + \alpha_i y_{t-1} + e_t \quad (16)$$

with

$$\begin{aligned} c_i &= (1-\alpha_i)\left\{\mu_0 + \sum_{j=1}^{i-1}(\mu_j - T_j^0) + y_{T_{i-1}^0} - \mu_{i-1}\right\} + \alpha_i\left\{\beta_0 + \sum_{j=1}^{i-1}\beta_j\right\} \\ b_i &= (1-\alpha_i)(\beta_0 + \sum_{j=1}^{i-1}\beta_j) \end{aligned} \tag{17}$$

The difference between (2) and (16) is the presence of the deterministic trend in the latter and that the intercept c_i is now a function of the trend parameters appearing in (15).

KPZ propose tests of the null hypothesis $\widetilde{H}_0^{(1)}$: $c_i = c,\ \alpha_i = 1$ for all i in (16). Note that under $\widetilde{H}_0^{(1)}$, $b_i = 0$ for all i so that the process follows a stable unit root process with possible drift. As in the trendless case, KPZ consider two models under the alternative hypothesis depending on whether the initial regime is trend or difference stationary:

Model 2a: $\alpha_i = 1$ in odd regimes and $|\alpha_i| < 1$ in even regimes.

Model 2b: $\alpha_i = 1$ in even regimes and $|\alpha_i| < 1$ in odd regimes.

In accordance with the notation in section 2, the test statistics in the trending case are denoted by $F_{2a}(\lambda, k),\ F_{2b}(\lambda, k),\ W_2(k)$ and $Wmax_2$. The limit distributions of the tests under $\widetilde{H}_0^{(1)}$ are shown to be different from those in the trendless case (see KPZ for details).

We now turn to testing the null of a stable trend stationary process, i.e., $\widetilde{H}_0^{(0)}$: $c_i = c,\ b_i = b,\ \alpha_i = \alpha$ for all i where $|\alpha| < 1$ in the model

$$y_t = c_i + b_i t + \alpha_i y_{t-1} + e_t \tag{18}$$

with $c_i = (1-\alpha_i)\left\{\mu_0 + \sum_{j=1}^{i-1}(\mu_j - T_j^0)\right\} + \alpha_i\left\{\beta_0 + \sum_{j=1}^{i-1}\beta_j\right\}$ and b_i defined as in (17). The test statistic for a fixed number $m = k$ changes is based on

$$G_2(\lambda, k) = [T - 3(k+1)](\widetilde{SSR}_0^{(0)} - SSR_{2,k}^{(0)})/[kSSR_{2,k}^{(0)}] \tag{19}$$

In (19), $\widetilde{SSR}_0^{(0)}$ denotes the sum of squared residuals under $\widetilde{H}_0^{(0)}$, i.e., that obtained from OLS estimation of (18) subject to the restrictions $c_i = c,\ b_i = b,\ \alpha_i = \alpha$ for all i. The quantity $SSR_{2,k}^{(0)}$ denotes the sum of squared residuals obtained from *unrestricted* OLS estimation of (16). The test statistic is then defined as $G_2(k) = \sup_{\lambda \in \Lambda_\epsilon^k} G_2(\lambda, k)$. When the number of breaks is unknown, the relevant test statistic is $UDmax_2 = \max_{1 \le k \le A} G_2(k)$.

The limit distributions of $G_2(k)$ and $UDmax_2$ are the not the same as those of $G_1(k)$ and $UDmax_1$ and therefore the BP critical values are not valid in the trending case. The reason

is that the BP asymptotics are based on the assumption of $I(0)$ regressors and hence do not allow for deterministic trends. We therefore derive the limit distributions of $G_2(k)$ and $UDmax_2$ under $\widetilde{H}_0^{(0)}$ in order to obtain the appropriate critical values. The result is stated in the following theorem:

Theorem 3 *Suppose Assumptions A1-A2 hold. Let* $F(r) = (1, r)'$, $r \in [0, 1]$. *Under* $\widetilde{H}_0^{(0)}$, *we have*

$$\begin{aligned} G_2(k) \;\Rightarrow\; & \sup_{\lambda \in \Lambda_\epsilon^k} \left\{ \begin{array}{c} \sum_{i=1}^{k} \frac{[\lambda_i B_2(\lambda_{i+1}) - \lambda_{i+1} B_2(\lambda_i)]^2}{\lambda_{i+1}\lambda_i(\lambda_{i+1}-\lambda_i)} \\ - \left(\int_0^1 F(r) dB_1(r)\right)' \left(\int_0^1 F(r)F(r)'\right)^{-1} \left(\int_0^1 F(r) dB_1(r)\right) \\ + \sum_{i=1}^{k+1} \left(\int_{\lambda_{i-1}}^{\lambda_i} F(r) dB_1(r)\right)' \left(\int_{\lambda_{i-1}}^{\lambda_i} F(r)F(r)'\right)^{-1} \left(\int_{\lambda_{i-1}}^{\lambda_i} F(r) dB_1(r)\right) \end{array} \right\} \\ & \stackrel{def}{=} G_2^*(k) \\ UDmax_2 \;\Rightarrow\; & \max_{1 \le k \le A} G_2^*(k) \end{aligned}$$

Remark 7 *The limit* $G_2^*(k)$ *depends on the two independent Brownian motions* $B_1(.)$ *and* $B_2(.)$ *as in the trendless case. However, while* $B_2(.)$ *enters the limit in the same form as in the BP distribution, allowing for a deterministic trend changes the way in which* $B_1(.)$ *enters the limit. The second and third terms in the expression for* $G_2^*(k)$ *reflect the difference from the* $I(0)$ *framework in BP through the dependence on the vector* $F(r)$.

Remark 8 *While the limit distribution stated in Theorem 3 is different from that in BP, asymptotic critical values can be obtained through Monte Carlo simulation for a given trimming choice* ϵ. *Section 8 discusses the details involved in the computation of the critical values.*

Remark 9 *The sequential procedure described in section 4 can be applied in the trending case with* $G_1(1)$, $W_1(1)$, $UDmax_1$ *and* $Wmax_1$ *replaced by* $G_2(1)$, $W_2(1)$, $UDmax_2$ *and* $Wmax_2$, *respectively. Theorems 1 and 2 continue to hold for this modified sequential procedure.*

6 Structural Breaks in Level and Trend

This section addresses the empirically important issue of distinguishing between processes with pure level shifts and/or trend breaks from those where these breaks are accompanied by concurrent shifts in persistence. In the trendless case, the issue of interest is that of

disentangling an $I(0)$ process with pure level shifts but a stable persistence parameter from one that is subject to shifts in both level and persistence. This case is discussed in section 6.1. In section 6.2, we consider the problem of discriminating between a process with pure trend breaks and one that is driven by breaks in persistence as well. Unlike the former case, the latter case needs to account for the possibility that the process can be either $I(0)$-stable or $I(1)$-stable around a broken deterministic trend.

6.1 Breaks in Level

The KPZ and BP tests on which the sequential approach advocated in section 4 is based are consistent against processes with pure shifts in level but a stable $I(0)$ persistence parameter. The approach cannot therefore be directly used to distinguish between processes characterized by level shifts only and those that are characterized by concurrent shifts in level and persistence. Our proposed procedure for distinguishing between these two possibilities is related to the approach taken in Hsu and Kuan (2001) who consider the problem of disentangling a slope change from a level shift in a time trend model with stable $I(0)$ errors. They suggest a two-step procedure in which the joint null of stability in both coefficients is tested and, conditional on a rejection, the break date is estimated by minimizing the sum of squared residuals and used in a second step to test the stability of the intercept or trend coefficient, while allowing the other coefficient to change.

Our approach to disentangling the two alternatives of interest is based on the fact that the estimated number of breaks obtained from applying the sequential procedure in section 4 is still consistent (Corollary 1) even if the process is only subject to shifts in level. Further, the estimated breakpoints $\hat{\lambda} = (\hat{\lambda}_1, ..., \hat{\lambda}_{\hat{m}})'$ obtained by minimizing the global sum of squared residuals from the *unrestricted* model (2) are T-consistent for the corresponding true breakpoints (Bai, 1994; Bai and Perron, 1998) regardless of whether the level shifts are accompanied by concurrent shifts in persistence. We can then test the stability of the persistence parameter based on a model that allows the intercept to be regime-specific under both the null and alternative hypotheses. Specifically, consider the standard Wald statistic for testing $\alpha_i = \alpha$ for all i in the model

$$y_t = c_i + \alpha_i y_{t-1} + e_t, \quad t \in [\hat{T}_{i-1} + 1, \hat{T}_i]; \; i = 1, ..., m+1 \tag{20}$$

Denote the resulting statistic at $W^*(m)$. The following result establishes the limit distribution of $W^*(m)$ under the null hypothesis that the process has a stable $I(0)$ persistence parameter with m shifts in level:

Theorem 4 *Suppose Assumptions A1-A2 hold. Under the null hypothesis* $\alpha_i = \alpha$ *for all* i *where* $|\alpha| < 1$, *we have*

$$W^*(m) \xrightarrow{d} \chi^2(m)$$

We therefore recommend the following three-step procedure. First, we determine the number of breaks ($\hat{m}$) applying the robust sequential procedure proposed in section 4. Second, conditional on $\hat{m}$ and the corresponding estimates of the breakpoints $(\hat{T}_1, \hat{T}_2, ..., \hat{T}_{\hat{m}})$ obtained from the unrestricted model (2), we compute the Wald statistic for testing the stable $I(0)$ null hypothesis (i.e., constancy of α_i over all i in (20)) while allowing the intercept to vary across the $(\hat{m}+1)$ regimes. Third, the null hypothesis of stable $I(0)$ persistence is rejected if the Wald statistic is significant at the specified level where the critical value is obtained from the $\chi^2(\hat{m})$ distribution. Otherwise, the null is not rejected and we conclude in favor of a model with pure level shifts.

6.2 Breaks in Trend

A rejection obtained by the test statistics presented in section 5 cannot be directly interpreted as one emanating from a change in persistence unless the trend function in (15) is stable, i.e., $\mu_j = \beta_j = 0$ for all $j \geq 1$. The reason is that the statistics are *joint* tests of the null hypothesis that the trend and persistence parameters are stable so that they are not only consistent against processes characterized by concurrent breaks in trend and persistence but also against processes that are subject to breaks in trend only. The problem of distinguishing among these two types of processes is complicated by the fact that one needs to account for the possibility that with pure trend breaks, the process can be either trend stationary or difference stationary, i.e., the persistence parameter can be either $I(1)$-stable or $I(0)$-stable.

Our approach to discriminating between the two alternatives at hand is based on recognizing that Theorem 2 remains valid for the sequential procedure (modified according to section 5) even when the process is subject to trend breaks only, given the joint nature of the null hypotheses. Further, the estimates of the breakpoints obtained by global minimization of the sum of squared residuals in the *unrestricted* model (18) that allows concurrent trend and persistence breaks are T-consistent regardless of whether the trend breaks are accompanied by shifts in persistence in the true data generating process. In the pure trend break case, this result follows from the results in Perron and Zhu (2005) while in the concurrent case, it follows from the results in Chong (2001) and Kejriwal and Perron (2012). Based on the estimated break number and the corresponding breakpoints, the null hypothesis of constant

persistence can be evaluated by testing for structural change in the persistence parameter while allowing the trend parameters to change across regimes. Under the null hypothesis of stable $I(0)$ persistence and m trend breaks, standard Wald tests have chi-squared limit distributions with m degrees of freedom while under the null hypothesis of stable $I(1)$ persistence, the limit distribution is non-standard (see Theorem 5 below). The null hypothesis of constant persistence [$I(1)$ or $I(0)$] can then be tested based on the intersection of the critical regions relevant for testing the $I(0)$ and $I(1)$ null hypotheses, respectively.

For testing $I(0)$-stability, we consider the standard Wald statistic for testing $\alpha_i = \alpha$ for all i in the model

$$y_t = c_i + b_i t + \alpha_i y_{t-1} + e_t, \quad t \in [\hat{T}_{i-1} + 1, \hat{T}_i], \ i = 1, ..., m+1 \tag{21}$$

Denote the resulting statistic as $\widetilde{W}_0(m)$. For testing $I(1)$-stability, we compute the Wald statistic based on the difference between the restricted and unrestricted sum of squared residuals where the former is obtained by estimating the restricted model

$$\Delta y_t = c_i + e_t, \quad t \in [\hat{T}_{i-1} + 1, \hat{T}_i], \ i = 1, ..., m+1 \tag{22}$$

while the latter is obtained by estimating the unrestricted model (21). We denote this statistic as $\widetilde{W}_1(m)$. The following theorem states the limit distributions of $\widetilde{W}_0(m)$ and $\widetilde{W}_1(m)$ under the respective null hypotheses:

Theorem 5 *Suppose that Assumptions A1-A2 hold.*
(a) Under the null hypothesis of m trend breaks and $\alpha_i = \alpha$ for all i where $|\alpha| < 1$, we have

$$\widetilde{W}_0(m) \xrightarrow{d} \chi^2(m)$$

(b) Let $F(r) = (1, r)'$. Under the null hypothesis of m trend breaks and $\alpha_i = 1$ for all i, we have

$$\widetilde{W}_1(m) \Rightarrow \sum_{i=1}^{m+1} \left\{ \begin{array}{c} \left(\int_{\lambda_{i-1}^0}^{\lambda_i^0} \left[\breve{B}_1^{(i)} \right]^2 \right)^{-1} \left(\int_{\lambda_{i-1}^0 1}^{\lambda_i^0} \breve{B}_1^{(i)} dB_1 \right)^2 + \\ \left(\int_{\lambda_{i-1}^0}^{\lambda_i^0} F(r) dB_1(r) \right)' \left(\int_{\lambda_{i-1}^0}^{\lambda_i^0} F(r) F(r)' \right)^{-1} \left(\int_{\lambda_{i-1}^0}^{\lambda_i^0} F(r) dB_1(r) \right) \\ - \frac{\left[B_1(\lambda_i^0) - B_1(\lambda_{i-1}^0) \right]^2}{\lambda_i^0 - \lambda_{i-1}^0} \end{array} \right\} \tag{23}$$

where $\breve{B}_1^{(i)}(r)$ are the residuals from a continuous-time regression of $\left\{ B_1(r) - B_1(\lambda_{i-1}^0) \right\}$ onto the space spanned by $F(r)$ for $r \in [\lambda_{i-1}^0, \lambda_i^0]$, $i = 1, ..., m+1$.

While the statistic $\widetilde{W}_0(m)$ has a standard chi-squared limiting distribution, $\widetilde{W}_1(m)$ is not asymptotically pivotal and its limit depends on the vector of true unknown fractions $\lambda^0 = (\lambda_1^0, ..., \lambda_m^0)'$. Asymptotic critical values cannot therefore be tabulated for use in practice and must be obtained on a case-by-case basis.

We suggest two alternative procedures for approximating the critical values of $\widetilde{W}_1(m)$. The first involves approximating the limit in (23) via Monte Carlo simulation by replacing λ^0 with the estimated break fractions $\hat{\lambda} = (\hat{\lambda}_1, ..., \hat{\lambda}_{\hat{m}})'$ obtained from the unrestricted model (21) and approximating the Weiner process $B_1(.)$ by partial sums of i.i.d. standard normal deviates. The second is a residual bootstrap approach that is based on resampling the residuals obtained by the estimating the model (22) and generating bootstrap samples on the time series again from (22) using the estimated c_i. The statistic $\widetilde{W}_1(m)$ is computed for each bootstrap sample and the critical values can be approximated using the quantiles of the bootstrap distribution of $\widetilde{W}_1(m)$. It is important that the bootstrap samples are drawn from a data generating process that imposes the null hypothesis of stable $I(1)$ persistence so that the resulting bootstrap distribution provides an accurate approximation to the finite sample null distribution. We employ the first procedure for the Monte Carlo experiments in section 9. Investigating the relative merits of the two approaches is outside the scope of the present paper and left for future research.

Based on the above discussion, we suggest the following three-step procedure for disentangling pure trend break processes from those that involve simultaneous breaks in trend and persistence. First, we determine the number of breaks ($\hat{m}$) applying the robust sequential procedure developed in section 4 appropriately modified for the trending case as discussed in section 5. Second, conditional on the estimated number of breaks and the corresponding estimates of the breakpoints obtained from the unrestricted model (21), we compute the Wald statistics for testing the stable $I(0)$ null hypothesis (i.e., constancy of α_i over all i in (21)) and the stable $I(1)$ null hypothesis (i.e., $\alpha_i = 1$ over all i in (21)). The critical value of the test statistic in the $I(0)$ case is obtained from, say, the 90th-percentile of the χ^2 distribution with $\hat{m}$ degrees of freedom while the value in the $I(1)$ case is obtained via one of the two aforementioned procedures. Third, the null hypothesis of constant persistence [$I(1)$ or $I(0)$] is rejected if both statistics are significant at the specified significance level. Otherwise, the null is not rejected and we conclude in favor of a model with pure trend breaks.

7 Serially Correlated Errors

We now provide an extension of model (1) that relaxes the AR(1) assumption on u_t to a general linear process driven by martingale difference innovations. Specifically, we make the following assumption:

Assumption A3: The errors h_t are generated as

$$\left.\begin{array}{c} A_i(L)h_t = e_t,\ A_i(L) = 1 - \sum_{s=1}^{\infty} a_{is}L^s \\ h_t = 0,\ t \leq T_{i-1}^0 \end{array}\right\} \begin{array}{c} t = T_{i-1}^0 + 1, T_{i-1}^0 + 2, ..., T_i^0; \\ i = 1, ..., m+1 \end{array} \tag{24}$$

where $\sum_{s=1}^{\infty} |a_{is}| < \infty$ for all i and e_t satisfies Assumption A2.

We also assume the following condition on the roots of $A_i(L)$:

Assumption A4: The autoregressive polynomial $A_i(L)$ has at most one real-valued root on the unit circle and all others strictly outside the unit circle.

From (24), we can write

$$h_t = \alpha_i h_{t-1} + \sum_{j=1}^{\infty} \pi_{ij} \Delta h_{t-j} + e_t$$

where $\alpha_i = \sum_{s=1}^{\infty} a_{is},\ \pi_{ij} = -\sum_{s=j+1}^{\infty} a_{is}$. In the trendless case, $y_t = \mu_i + u_t$ which leads to the test regression

$$\Delta y_t = c_i + (\alpha_i - 1)y_{t-1} + \sum_{j=1}^{l_T} \pi_j \Delta y_{t-j} + e_t^* \tag{25}$$

for some sequence l_T increasing with the sample size (precise conditions on l_T will be specified later). In the trending case, we augment (25) with a deterministic time trend regressor. Note that we do not allow the coefficients of the first differences (representing the short-run dynamics) to change across regimes. This is because, as stressed in KPZ, we wish to direct the tests 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 $I(1)$ null would lead to limit distributions that depend on the (unknown) number and location of these breaks thereby making asymptotic inference difficult (Under the $I(0)$ null, the BP limit distributions remain the same whether or not the dynamics change; see BP). Simulation evidence presented in KPZ illustrates that the test statistics do not have much power against pure changes in short-run dynamics but are 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. The test statistics in the serially correlated case are constructed in a similar way as in section 4 except that the relevant test regression is now (25) instead of (2). For the sake of brevity, we do not reproduce the expressions for the test statistics here as they are detailed in KPZ.

Finally, we make the following assumption on the rate at which the lag length l_T increases with the sample size that facilitates the derivation of the limit distributions in the serially correlated case:

Assumption A5: 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 6 *Suppose Assumptions A1-A5 hold. Then for the KPZ and BP test statistics computed from* (25)*, Theorems 1-5 continue to hold.*

The proof of the theorem is omitted as it follows directly from the fact that all limit results for the test statistics (under the null and alternative hypotheses considered in the AR(1) case) remain valid in the general case as long as the statistics are computed from (25) and the lag length satisfies Assumption A5 [see Theorems 2 and 3 in KPZ]. The implication of Theorem 6 is that the asymptotic critical values needed to implement the proposed sequential approach remain the same as in the AR(1) case. The finite-sample effects of serially correlated errors will be examined via Monte Carlo simulations in section 9.

An important practical issue regarding the implementation of the sequential approach with serially correlated errors concerns the choice of l_T in finite samples. Based on extensive simulation experiments, we found the following approach to be both computationally efficient as well as deliver robust results with respect to selecting the true number of breaks. First, we determine the lag length using BIC based on the stable $I(0)$ and stable $I(1)$ null hypotheses, respectively. The maximum of the two lag lengths is then used to compute the $Wmax_1$ and $UDmax_1$ (or $Wmax_2$ and $UDmax_2$ in the trending case) tests in step 1 of the sequential procedure. Second, upon a rejection, estimate the *unrestricted* single break model over the full sample for each allowable lag length (zero to, say, l_{max}) whereby all regression coefficients

including those of the lagged differences are allowed to change across regimes. Choose the lag length as the minimizer of the BIC over $[0, l_{max}]$. Third, use the lag length thus determined to compute the test statistics in the two regimes specified by the estimated break date. Subsequently, at each stage "j" of the procedure, choose the lag length by minimizing the BIC across permissible lag lengths where each model is estimated by global minimization of the sum of squared residuals allowing for j breaks. The test statistics in the $(j+1)$ estimated regimes are then computed using this choice. In this way, the choice of the lag length adapts to the null hypothesis under consideration at each step (j versus $(j+1)$ breaks). This approach to lag selection was observed to dominate an approach based on a fixed number of lags as well as one where the lag choice was made once and for all under either the stable $I(0)$ or $I(1)$ null hypothesis.

8 Asymptotic Critical Values

This section details the computation of the asymptotic critical values relevant for applying the proposed approach and tabulates these values for a range of choices of the trimming parameter ϵ. In particular, we present the appropriate quantiles of the test statistics $W_1(1)$, $Wmax_1, W_2(1)$, $Wmax_2, G_2(1)$ and $UDmax_2$ noting that the quantiles for $G_1(1)$ and $UDmax_1$ are available from Bai and Perron (2003b).

Given the non-standard nature of the limit distributions, the critical values are obtained by Monte-Carlo simulations. The Brownian motions $B_1(.)$ and $B_2(.)$ are approximated by partial sums of i.i.d. standard normal deviates with step size equal to 500. For the statistics $W_1(1)$, $Wmax_1, W_2(1)$ and $Wmax_2$ 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. Finally, we evaluate the expressions appearing in the limit distributions at the vector of break fractions obtained earlier. For the statistics $G_2(1)$ and $UDmax_2$, we first generate a random sample of $T = 500$ observations from the standard normal distribution. Next, we apply the Bai and Perron (2003a) dynamic programming algorithm to obtain the minimized sum of squared residuals and the corresponding vector of break fractions and evaluate the limit expressions at this vector. The procedure is repeated 5000 times to obtain the required quantiles of the limit distributions.

Asymptotic critical values for implementing the sequential test of l versus $(l+1)$ breaks $(l = 0, 1, ..., 5)$ are provided in Tables 1a and 1b for three trimming choices: $\epsilon = 0.15, 0.20, 0.25$.

For $\epsilon = 0.15$, the upper bound A needed to be compute $Wmax_1, Wmax_2$, $UDmax_1$ and $UDmax_2$ was set to five while for $\epsilon = 0.20$ and $\epsilon = 0.25$, A was chosen as three and two, respectively. A GAUSS program for computing the critical values for other choices of ϵ and/or A is available upon request.

9 Monte Carlo Evidence

This section conducts a set of Monte Carlo experiments designed to assess the viability of the proposed procedure in finite samples for a variety of data generating processes and thereby evaluate the adequacy of the asymptotic approximations derived in the foregoing sections for sample sizes typically encountered in empirical applications. In particular, we compare the performance of the robust sequential approach advocated in section 4 with the commonly employed BP approach in terms of their relative ability to estimate the true number of breaks, when the data generating process is characterized by zero, one or two breaks, respectively. We also document the finite sample performance of the approaches recommended in section 6 for distinguishing between pure level/trend break processes and those that are accompanied by concurrent persistence shifts.

To account for serial correlation, we implement the lag selection procedure outlined in section 7 with $l_{max} = 4$. Given that the sequential tests are computed from successively smaller subsamples, it is important to use a relatively small number of lags from a parsimony standpoint when constructing the various statistics. Nevertheless, we also implemented the procedure with $l_{max} = 6$ and found the results to be inferior in terms of rue break selection probabilities across procedures compared to using $l_{max} = 4$.

In all experiments, $\{e_t\}$ denotes a sequence of $i.i.d.$ $N(0,1)$ variables. The errors $\{u_t\}$ are generated by the 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) $\rho = 0$, $\theta = -0.5$. The sequential test (for both procedures) at each step is evaluated at the 10% nominal level (i.e., $\eta = .10$ in section 4).[2] Three sample size are considered: $T = 200; 400; 600$. The trimming parameter was set at $\epsilon = .15$ and the maximum number of breaks at $A = 5$. All experiments are based on 1000 replications. Sections 9.1-9.3 present results on the finite sample adequacy of the robust sequential and BP approaches in estimating the true number

[2] We also considered using $\eta = .05$ or $.01$ but found that that the underestimation probabilities were considerable in many cases and that $\eta = .10$ appeared to provide the best compromise in terms of the size-power tradeoff.

of breaks and section 9.4 reports evidence on the viability of the approaches recommended in section 6.

9.1 The Case with No Break

In the no break case, the time series y_t is generated as

- DGP-0:

$$\begin{aligned} y_t &= \widetilde{\alpha} y_{t-1} + u_t \\ y_0 &= 0 \end{aligned} \tag{26}$$

Based on the ARMA specification for u_t, the persistence parameter defined as the sum of the autoregressive (AR) coefficients in the implied AR representation for y_t is

$$\alpha = \frac{\widetilde{\alpha} + \rho(1 - \widetilde{\alpha}) - \theta}{1 - \theta}$$

For a given (ρ, θ) configuration, we choose $\widetilde{\alpha}$ such that α equals a pre-specified value. Specifically, we consider five values of α: $.5, .7, .9, .95, 1$. For instance, when $\alpha = .7$ and $(\rho, \theta) = (.5, 0)$, $\widetilde{\alpha} = .4$. Of course, $\widetilde{\alpha} = \alpha$ for $(\rho, \theta) = (0, 0)$. Also, for all (ρ, θ) configurations considered here, $\widetilde{\alpha} = 1$ when $\alpha = 1$.[3] Holding the persistence parameter fixed across different serial correlation scenarios facilitates the comparison of results across these scenarios.

The results are presented in terms of probabilities of (correctly) selecting no break (P_c, "c" for "correct") and overestimating the number of breaks (P_0, "0" for "overestimation"). The robust sequential and BP procedures are denoted by Seq_R and Seq_{BP}, respectively. The findings are reported in Table 2. First, consider the case $\rho = \theta = 0$. When the process is relatively less persistent ($\alpha \leq .7$), there is no difference in the performance of the two procedures. However, as α increases further, the differences become quite prominent with the BP procedure selecting at least one break far more frequently. For instance, with $\alpha = .9$ and $T = 200$, the Seq_R selects no break with 94% probability compared to only 71% for Seq_{BP} (see Remark 3). As shown in section 3, the BP test has incorrect asymptotic size when $\alpha = 1$. This fact is reflected in the break selection probabilities for this case where Seq_{BP} favors a persistence change model with probability at least 75% across sample sizes. In sharp contrast, Seq_R selects the stable model with probability at least 80% across all values of α and T considered. Interestingly, as the sample size increases from $T = 200$, there is a

[3] This holds as long as $|\rho| < 1,\ |\theta| < 1$.

noticeable decline in P_c for the Seq_R procedure when $\alpha \in \{.9, .95\}$, while the corresponding decline for $\alpha = 1$ is only marginal. Such behavior can be explained by the fact that the KPZ limit distributions are derived under $\alpha = 1$ so that the tests do not have the correct size when the process has a root close to but equal to unity (under a fixed α with $|\alpha| < 1$, the KPZ tests diverge while with a local to unity parameterization $\alpha_T = 1 + c/T,\ c < 0$, the limit distributions depend on the non-centrality parameter c). With serial correlation, the picture is qualitatively similar to the i.i.d. case with Seq_{BP} exhibiting substantial over-rejections in favor of models with at least one break while Seq_R displaying robust performance in selecting the no break model across the different data generating configurations.

9.2 The Case with One Break

We consider a variety of DGPs depending on whether the process is characterized by switches between $I(1)$ and $I(0)$ regimes or between $I(0)$ regimes with different persistence parameters as well as the direction of such switches, i.e., high to low persistence or vice-versa. Specifically, the following DGPs are included in our analysis:

	For $t \leq [T\lambda_1^0]$	For $t \geq [T\lambda_1^0] + 1$
DGP-1	$y_t = \widetilde{\alpha} y_{t-1} + u_t$	$y_t = y_{t-1} + u_t$
DGP-2	$y_t = y_{t-1} + u_t$	$y_t - y_{[T\lambda_1^0]} = \widetilde{\alpha}(y_{t-1} - y_{[T\lambda_1^0]}) + u_t$
DGP-3	$y_t = \widetilde{\alpha}_1 y_{t-1} + u_t$	$y_t = \widetilde{\alpha}_2 y_{t-1} + u_t$

DGP-1 and DGP-2 are processes that involve a switch between an $I(1)$ and an $I(0)$ regime while DGP-3 involves a switch from one $I(0)$ regime to another, i.e., DGP-3 represents an $I(0)$-preserving switch. As in section 9.1, for a given (ρ, θ) combination, the parameters $\widetilde{\alpha}, \widetilde{\alpha}_1, \widetilde{\alpha}_2$ are chosen so that the implied persistence parameters $\alpha, \alpha_1, \alpha_2$ equals a fixed value. For DGP-1 and DGP-2, we present results for $\alpha \in \{.5, .7\}$. For DGP-3, we take $\alpha_1, \alpha_2 \in \{.2, .9\}$ and define $\alpha = \alpha_2 - \alpha_1$. When $\rho = \theta = 0$, we consider three values for the location of the break: $\lambda_1^0\{.3, .5, .7\}$. For the serially correlated cases, the findings are presented only for $\lambda_1^0 = .5$, as the pattern of results was found to be similar to the i.i.d. case for the other break locations. For each DGP and parameter combination, we computed the probabilities of (correctly) selecting the one break model (P_c), more than one break (P_0) and no break (P_u, "u" for underestimation).

The findings are reported in Table 3a for $\rho = \theta = 0$. As in the no break case, the BP procedure is substantially over-sized when there is an $I(1)$ segment in the data resulting

in large values of P_0 for DGP-1 and DGP-2. In fact, even for DGP-3 where the break is $I(0)$-preserving, Seq_{BP} selects more than one break quite frequently. For instance, with α = -.7, $T = 200$ and $\lambda_1^0 = .3$, BP selects an overspecified model in 38% of the samples. On the other hand, the proposed procedure shows much better performance in identifying the true model across all three break locations for a given DGP. There are, however, some differences in the performance of Seq_R across DGPs. For example, when $T = 200$ and $\alpha = .7$, the probability of underestimation can be substantial for DGP-1, with the stable model being preferred in as many as 78% of the samples when $\lambda_1^0 = .3$. This feature reflects the relatively low power of the KPZ test for DGP-1 when the break occurs early and the sample size is small. The underestimation frequencies, however, drop sharply when the sample size increases. In all parameter configurations but one, Seq_R selects the true model with probability at least 80% when $T \geq 400$.

The results with serially correlated errors are presented in Table 3b. Here the performance of Seq_R hinges on the nature of serial correlation. When the serial correlation is positive ($\rho > 0$ or $\theta < 0$), the procedure remains robust in that the true model selection probabilities (P_c) are comparable to those in the serially uncorrelated case. The situation is different when a negative MA component is present, however. In this case, Seq_R is subject to substantial underestimation with the stable model chosen in at least 26% of the samples when $T = 200$ across DGP 1-3. The problem is more severe when an $I(1)$ regime is present. In fact, with $T = 200$, Seq_{BP} selects the true model more frequently than Seq_R for DGP-1. The underestimation can be explained by the fact that the autoregressive approximation (25) does not adequately account for the serial correlation in this case. Therefore, just as not accounting for positive serial correlation (i.e., a unit root) causes a bias in favor of a model with structural change, negative serial correlation induces a bias against a model with structural change. This issue is, however, ameliorated as the sample size increases and Seq_R is seen to dominate Seq_{BP} when $T \geq 400$.

9.3 The Case with Two Breaks

With two breaks, the DGPs considered are the following:

	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-4	$y_t = y_{t-1} + u_t$	$y_t - y_{[T\lambda_1^0]} = \widetilde{\alpha}(y_{t-1} - y_{[T\lambda_1^0]}) + u_t$	$y_t = y_{t-1} + u_t$
DGP-5	$y_t = \widetilde{\alpha} y_{t-1} + u_t$	$y_t = y_{t-1} + u_t$	$y_t - y_{[T\lambda_2^0]} = \widetilde{\alpha}(y_{t-1} - y_{[T\lambda_2^0]}) + u_t$
DGP-6	$y_t = \widetilde{\alpha}_1 y_{t-1} + u_t$	$y_t = \widetilde{\alpha}_2 y_{t-1} + u_t$	$y_t = \widetilde{\alpha}_1 y_{t-1} + u_t$

We choose $\widetilde{\alpha},\ \widetilde{\alpha}_1, \widetilde{\alpha}_2$ in the same way as described in section 9.2 with the corresponding persistence parameters being denoted $\alpha, \alpha_1, \alpha_2$, respectively. Similarly, define $\alpha = \alpha_2 - \alpha_1$ for DGP-6. For $\rho = \theta = 0$, we consider three break locations: $(\lambda_1^0, \lambda_2^0) \in \{(.2, .7), (.3, .8), (.4, .8)\}$. In the serially correlated case, results are reported only for $(\lambda_1^0, \lambda_2^0) = (.3, .8)$, the results being qualitatively similar to the i.i.d. case for the other location configurations.

Table 4a reports the results for $\rho = \theta = 0$. In accordance with the findings in Tables 2, 3a and 3b, the BP procedure is subject to considerable size distortions in all cases, including DGP-6 where the process is $I(0)$ in each regime. The Seq_R procedure is relatively more reliable at identifying the true model across the three break locations although, as in the one break case, the underestimation probabilities can be substantial when $T = 200$ for DGP-5, where the initial switch is from an $I(0)$ to an $I(1)$ regime. For DGP-6, Seq_R outperforms Seq_{BP} for all sample sizes and parameter values even though the DGP satisfies the BP assumptions and is asymptotically valid in this case. The reason is that even in the absence of a unit root regime in the time series, the presence of a persistent autoregressive root in at least one regime is sufficient for the BP procedure to be subject to severe oversizing, even with $T = 600$. This feature again reflects the fact that the BP limit distribution (7) is a poor approximation to the corresponding finite sample distribution when the true autoregressive parameter is close to but not equal to unity (Remark 3). The findings with serially correlated errors are presented in Table 4b. As in the one break case, Seq_R suffers from underestimation in the negative MA case, but its performance is adequate with positive serial correlation when $T \geq 400$.

9.4 Breaks in Level and Trend

We now investigate the finite sample adequacy of the approaches outlined in section 6 to distinguish between level and/or trend break processes from those with persistence breaks. We assume that y_t is generated by the single break model

$$y_t = \mu_0 + \beta_0 t + \mu_1 DU_{1t} + \beta_1 DT_{1t} + y_t^*$$

where $DU_{1t} = I(t > T_1^0),\ DT_{1t} = (t - T_1^0) I(t > T_1^0)$ and y_t^* either follows the stable process (26) or one of the single break DGPs considered in section 9.2 with $\rho = \theta = 0$. Significance is evaluated at the 10% nominal level.

We first consider the non-trending case and set $\beta_0 = \beta_1 = 0$. The level parameters are set at $\mu_1^0 = 5,\ \mu_2^0 = 10$. The corresponding DGPs are denoted $0_c, 1_c, 2_c, 3_c$ and represent the counterparts of DGPs 0-3 with a level shift. The results are reported in Table 5a.

Each entry represents the probability of selecting the true model given the corresponding parameter configuration. When the process is $I(0)$-stable [DGP-0_c], the performance of the procedure is satisfactory except when the break occurs late. When there is a concurrent persistence shift [DGP 1_c-3_c], the procedure performs exceedingly well with the true model selection probabilities exceeding 95% in all cases when $T \geq 400$. The recommended approach thus appears to be quite successful at distinguishing between the two alternatives of interest, at least with a large enough sample size.

In the trending case, we set $\mu_0 = \mu_1 = 0$ and $\beta_0 = 1,\ \beta_1 = 2$. The corresponding DGPs are denoted $0_t, 1_t, 2_t, 3_t$ and represent the trend break counterparts of the DGPs 0-3. To approximate the limit distribution (23) in the $I(1)$ case, we replace the unknown break date with the estimate obtained by minimizing the sum of squared residuals from (21) and approximating the Weiner process $B_1(.)$ by partial sums of i.i.d. $N(0,1)$ variables with step size equal to 1000. The 10% critical value is then obtained based on 5000 Monte Carlo replications. For the $I(0)$ statistic, the 10% χ^2_1 critical value is 2.71. The results are reported in Table 5b. For DGP-0_t, where the process is subject to a break in trend only, the procedure is adequate in most cases, the notable exception being when $\alpha = 1$ and $\lambda_1^0 = .7$. In this latter case, the pure trend break model is rejected in more than 15% of the replications when $T \leq 400$ although the probability drops to 10% when $T = 600$. For DGPs 1_t-3_t, power is a concern only when $T = 200$, the break size is small and the break occurs early or late in the sample. When the sample size increase to $T = 400$, the procedure selects the persistence change model with higher than 95% probability across break locations and all DGPs 1_t-3_t.

In summary, the Monte Carlo evidence suggests that the proposed sequential procedure performs adequately in determining the number of breaks driving the time series, except when the errors contain a negative moving average component. It also performs considerably better than the BP procedure when the process contains a regime with a persistent autoregressive root. The procedure can therefore be applied without a priori assumptions on whether the underlying time series is characterized by only $I(0)$ regimes or a mix of $I(1)$ and $I(0)$ regimes. Our suggested approaches for distinguishing between processes with pure level/trend breaks from those that are subject to persistence shifts as well also appear to perform well in finite samples. An empirically important issue concerns the choice of the maximum allowable number of breaks in relation to the sample size. Given that the sequential tests are implemented on subsamples of the data, allowing for too many breaks with a relatively small sample size is likely to result in size distortions/low power resulting in overestimating/underestimating the number of breaks. Therefore, one should allow for

a sufficient number of observations per data segment and choose the maximum number of breaks permissible accordingly. Based on our experiments, it appears reasonable to allow for a maximum of five breaks when $T = 400$. This issue was also discussed by Kejriwal and Perron (2010b) in the context of determining the number of trend breaks based on their proposed sequential procedure.

10 Conclusion

This paper proposes a new sequential procedure for estimating the number of breaks in the persistence of a univariate time series, when it is not known a priori whether the breaks are $I(0)$-preserving in nature or associated with switches between $I(1)$ and $I(0)$ regimes. It may be viewed as an extension of the commonly applied Bai and Perron (1998) procedure to allow for one or more $I(1)$ segments in the data. Several extensions of the paper are in order, two of which are particularly worth noting. First, it would be useful to extend our approach to the case of error heteroskedasticity (or nonstationary volatility) given the pervasive evidence on volatility shifts in economic time series (McConnell and Perez-Quiros, 2000; Sensier and van Dijk, 2004). Cavaliere and Taylor (2008) construct tests of the stable $I(0)$ null hypothesis against the alternative of a single change in persistence based on a wild bootstrap procedure. Their bootstrap approach can potentially be adapted to construct versions of the KPZ and BP tests that are robust to volatility shifts (Kejriwal and Yu, 2017). Second, a more general framework for assessing persistence stability would allow the process to be stationary long memory in some regimes, i.e,. $I(d),\ d > 0$. While developing a break detection procedure within such a framework appears considerably more involved, it should pave the way for a wide range of interesting empirical applications.

References

Altissimo, F., Corradi, V. (2003), "Strong rules for detecting the number of breaks in a time series," *Journal of Econometrics* 117, 207-244.

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

Bai, J. (1994), "Least squares estimation of a shift in linear processes," *Journal of Time Series Analysis* 15, 453-472.

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

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

Bai, J., Perron, P. (2003b), "Critical values for multiple structural change tests," *Econometrics Journal* 6, 72-78.

Bai, J., Perron, P. (2006), "Multiple structural change models: a simulation analysis," in *Econometric Theory and Practice: Frontiers of Analysis and Applied Research,* D. Corbae, S.N. Durlauf, B.E. Hansen(eds.), Cambridge University Press, 212-240.

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

Cavaliere, G., Taylor, A.M.R. (2008), "Testing for a change in persistence in the presence of non-stationary volatility," *Journal of Econometrics* 147, 84-98.

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

Dickey, D.A., Fuller, W.A. (1979), "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association* 74, 427-431.

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

Harvey, D.I., Leybourne, S.J., Taylor, A.M.R. (2006), "Modified tests for a change in persistence," *Journal of Econometrics* 134, 441-469.

Hosoya, Y. (1989), "Hierarchical statistical models and a generalized likelihood ratio test," *Journal of the Royal Statistical Society, Series B* 51, 435-447.

Hsu, C-C., Kuan, C-M. (2001), "Distinguishing between trend-break models: method and empirical evidence," *Econometrics Journal* 4, 171-190.

Kejriwal, M., Perron, P. (2010a), "Testing for multiple structural changes in cointegrated regression models," *Journal of Business & Economic Statistics* 28, 503-522.

Kejriwal, M., Perron, P. (2010b), "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., Perron, P. (2012), "A note on estimating a structural change in persistence," *Economics Letters* 117, 932-935.

Kejriwal, M., Perron, P., Zhou, J. (2013), "Wald tests for detecting multiple structural changes in persistence," *Econometric Theory* 2013, 289-323.

Kejriwal, M., Yu., X. (2017), "Bootstrap tests for multiple structural changes in persistence in the presence of nonstationary volatility," Manuscript in Preparation, Purdue University.

Kim, J.Y. (2000), "Detection of change in persistence of a linear time series," *Journal of Econometrics* 54, 159-178.

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., Kim, T., Smith, V., Newbold, P. (2003), "Tests for a change in persistence against the null of difference-stationarity," *Econometrics Journal* 6, 291-311.

Leybourne, S.J., Kim , Taylor, A.M.R. (2007a), "CUSUM of squares-based tests for a change in persistence," *Journal of Time Series Analysis* 28, 408-433.

Leybourne, S.J., Kim, T., Taylor, A.M.R. (2007b), "Detecting multiple changes in persistence," *Studies in Nonlinear Dynamics & Econometrics* Vol. 11(3), Article 2.

Liu, J., Wu, S., Zidek, J.V., (1997), "On segmented multivariate regressions," *Statistica Sinica* 7, 497-525.

McConnell, M.M., Perez-Quiros, G., (2000), "Output fluctuations in the United States: What has changed since the early 1980s?," *American Economic Review* 90, 1464-1476.

Ng, S., Perron, P. (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.

Page, E. S. (1955). "A test for a change in a parameter occurring at an unknown point," *Biometrika* 42, 523-527.

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 *Palgrave Handbook of Econometrics*, K. Patterson and T.C. Mills (eds.), Palgrave Macmillan, 278-352.

Perron, P., Qu, Z. (2006), "Estimating restricted structural change models," *Journal of Econometrics* 134, 373-399.

Phillips, P.C.B., Perron, P. (1988), "Testing for a unit root in time series regression," *Biometrika* 75, 335-346.

Quandt, R.E. (1960), "Tests of the hypothesis that a linear regression system obeys two separate regimes," *Journal of the American Statistical Association* 55, 324-330.

Sensier, M., van Dijk, D. (2004), "Testing for volatility changes in US macroeconomic time series," *Review of Economics and Statistics* 86, 833-839.

Stock, J.H., Watson, M.W. (1996), "Evidence on structural instability in macroeconomic time series relations," *Journal of Business & Economic Statistics* 14, 11-30.

Taylor, A.M.R. (2005), "Fluctuation tests for a change in persistence," *Oxford Bulletin of Economics and Statistics* 67, 207-230.

Yao, Y-C. (1988), "Estimating the number of change-points via Schwarz' criterion", *Statistics and Probability Letters* 6, 181-189.

Table 1a: Asymptotic Critical Values of the $H_j(l+1|l)$ tests for $j = 1, 2$ [i.e., $cv_{w,j}(\eta_{l+1}),\ cv_{w,max}(\eta)$]

Non Trending Case									Trending Case								
		l									l						
ϵ	$1-\eta$	0	1	2	3	4	5	$Wmax_1$	ϵ	$1-\eta$	0	1	2	3	4	5	$Wmax_2$
.15	.90	8.09	8.94	9.53	9.96	10.29	10.51	9.86	.15	.90	7.28	7.96	8.39	8.72	8.91	9.10	7.71
	.95	8.99	10.00	10.52	10.91	11.20	11.44	10.90		.95	7.98	8.74	9.11	9.56	9.72	9.80	8.43
	.975	10.00	10.95	11.46	11.68	12.06	12.34	11.95		.975	8.75	8.75	9.84	10.26	10.60	10.66	9.18
	.99	11.21	12.06	12.66	1.90	13.24	13.53	13.02		.99	9.73	9.73	10.60	10.75	11.12	11.18	10.07
.20	.90	7.85	8.80	9.32	9.84	10.18	10.34	9.30	.20	.90	7.15	7.82	8.19	8.47	8.60	8.89	7.42
	.95	8.85	9.86	10.34	10.64	11.00	11.28	10.23		.95	7.82	8.48	8.91	9.09	9.20	9.38	8.08
	.975	9.88	10.67	11.29	11.72	11.99	12.10	11.16		.975	8.51	9.10	9.39	9.60	9.94	10.10	8.65
	.99	11.03	11.99	12.18	12.44	12.97	13.72	12.12		.99	9.21	9.94	10.43	10.70	10.89	11.23	9.42
.25	.90	7.61	8.52	8.98	9.40	9.76	10.07	8.63	.25	.90	6.97	7.66	8.07	8.37	8.55	8.67	7.14
	.95	8.55	9.43	10.11	10.41	10.75	10.97	9.49		.95	7.69	8.38	8.68	9.02	9.16	9.27	7.82
	.975	9.45	10.43	11.00	11.33	11.71	11.81	10.36		.975	8.41	9.03	9.30	9.44	9.64	9.94	8.51
	.99	10.77	11,71	11.99	12.17	14.44	12.62	11.57		.99	9.16	9.64	10.16	10.69	10.85	11.23	9.22

Table 1b: Asymptotic Critical Values of the BP tests [i.e., $cv_{g,1}(\eta_{l+1}),\ cv_{g,max}(\eta)$], Trending Case

		l						
ϵ	$1-\eta$	0	1	2	3	4	5	$UDmax_2$
.15	.90	10.77	12.60	13.56	14.28	14.71	15.00	11.04
	.95	14.33	15.02	15.79	16.51	16.82	16.82	12.85
	.975	14.36	15.82	16.89	17.27	17.63	18.09	14.40
	.99	16.57	17.63	18.52	18.85	19.27	19.42	16.57
.20	.90	10.43	12.20	13.27	13.95	14.54	14.88	10.59
	.95	12.23	14.01	14.96	15.63	16.26	16.59	12.40
	.975	14.12	15.68	16.60	17.13	17.47	17.87	14.13
	.99	16.34	17.47	18.27	18.85	19.42	20.03	16.34
.25	.90	10.09	11.87	12.88	13.56	14.26	14.58	10.21
	.95	11.95	13.61	14.64	15.26	15.96	16.39	12.01
	.975	13.69	15.38	16.47	16.92	17.37	17.47	13.70
	.99	15.99	17.37	18.09	18.50	19.42	19.60	16.00

Table 2: Break Selection Probabilities [DGP-0, $m = 0$]

T	α	Seq_R		Seq_{BP}		Seq_R		Seq_{BP}		Seq_R		Seq_{BP}		Seq_R		Seq_{BP}	
		P_c	P_0	P_c	P_0	P_c	P_0	P_c	P_0	P_c	P_0	P_c	P_0	P_c	P_0	P_c	P_0
		(A) $\rho = \theta = 0$				(B) $\rho = .5,\ \theta = 0$				(C) $\rho = 0,\ \theta = .5$				(D) $\rho = 0,\ \theta = -.5$			
200	.5	.90	.10	.90	.10	.92	.08	.89	.11	.90	.10	.71	.29	.90	.10	.85	.15
	.7	.87	.13	.87	.13	.92	.08	.88	.12	.88	.12	.67	.33	.90	.10	.82	.18
	.9	.94	.06	.71	.29	.89	.11	.76	.24	.92	.08	.44	.56	.92	.08	.72	.28
	.95	.94	.06	.52	.48	.88	.12	.65	.35	.92	.08	.33	.15	.91	.09	.58	.42
	1	.94	.06	.24	.76	.90	.10	.24	.76	.92	.08	.25	.75	.89	.11	.27	.73
400	.5	.90	.10	.90	.10	.91	.09	.90	.10	.89	.11	.78	.22	.92	.08	.88	.12
	.7	.88	.12	.88	.12	.91	.09	.89	.11	.86	.14	.74	.26	.90	.10	.87	.13
	.9	.82	.18	.77	.23	.83	.17	.81	.19	.87	.13	.53	.47	.86	.14	.80	.12
	.95	.86	.14	.59	.41	.82	.18	.73	.27	.89	.11	.42	.58	.88	.12	.71	.29
	1	.92	.08	.20	.80	.89	.11	.20	.80	.92	.08	.21	.79	.90	.10	.23	.77
600	.5	.90	.10	.90	.10	.92	.08	.91	.09	.91	.09	.81	.19	.93	.07	.90	.10
	.7	.89	.11	.89	.11	.92	.08	.90	.10	.87	.13	.79	.21	.91	.09	.91	.09
	.9	.80	.20	.80	.20	.86	.14	.83	.17	.80	.20	.63	.37	.86	.14	.85	.15
	.95	.80	.20	.70	.30	.80	.20	.78	.22	.83	.17	.46	.54	.84	.16	.79	.21
	1	.87	.13	.19	.81	.85	.15	.18	.82	.87	.13	.22	.78	.85	.15	.23	.77

Table 3a: Break Selection Probabilities [$m = 1,\ \rho = \theta = 0$]

DGP	α	SEQ_R			SEQ_{BP}			SEQ_R			SEQ_{BP}			SEQ_R			SEQ_{BP}		
		P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u
		$T = 200$						$T = 400$						$T = 600$					
		(A) $\lambda_1^0 = .5$																	
1	.5	.88	.12	.00	.50	.50	.00	.91	.09	.00	.46	.54	.00	.91	.09	.00	.44	.56	.00
	.7	.54	.08	.38	.49	.49	.02	.89	.11	.00	.44	.56	.00	.90	.10	.00	.45	.55	.00
2	.5	.90	.09	.01	.51	.48	.01	.89	.11	.00	.49	.51	.00	.89	.11	.00	.46	.54	.00
	.7	.75	.07	.18	.52	.44	.04	.89	.11	.00	.49	.51	.00	.87	.13	.00	.47	.53	.00
3	.7	.93	.07	.00	.72	.28	.00	.87	.13	.00	.77	.23	.00	.89	.11	.00	.85	.15	.00
	-.7	.91	.09	.00	.69	.31	.00	.89	.11	.00	.77	.23	.00	.87	.13	.00	.85	.15	.00
		(B) $\lambda_1^0 = .3$																	
1	.5	.60	.10	.30	.45	.53	.02	.91	.09	.00	.46	.54	.00	.89	.11	.00	.42	.58	.00
	.7	.17	.05	.78	.42	.51	.07	.70	.12	.18	.42	.58	.00	.87	.13	.00	.43	.57	.00
2	.5	.91	.08	.01	.50	.49	.01	.92	.08	.00	.52	.48	.00	.90	.10	.00	.46	.54	.00
	.7	.80	.07	.13	.47	.43	.10	.92	.08	.00	.52	.48	.00	.89	.11	.00	.46	.54	.00
3	.7	.90	.09	.01	.77	.22	.01	.87	.13	.00	.83	.17	.00	.86	.14	.00	.85	.15	.00
	-.7	.91	.09	.00	.62	.38	.00	.90	.10	.00	.71	.29	.00	.86	.14	.00	.77	.23	.00
		(C) $\lambda_1^0 = .7$																	
1	.5	.87	.12	.01	.50	.49	.01	.91	.09	.00	.51	.49	.00	.92	.08	.00	.47	.53	.00
	.7	.76	.11	.13	.47	.44	.09	.91	.08	.01	.49	.50	.01	.88	.12	.00	.46	.54	.00
2	.5	.74	.10	.16	.54	.46	.00	.90	.10	.00	.46	.54	.00	.89	.11	.00	.46	.54	.00
	.7	.36	.06	.58	.51	.46	.03	.81	.14	.05	.43	.57	.00	.86	.14	.00	.46	.54	.00
3	.7	.91	.09	.00	.65	.35	.00	.91	.09	.00	.73	.27	.00	.92	.08	.00	.83	.17	.00
	-.7	.88	.11	.01	.72	.28	.00	.87	.13	.00	.81	.19	.00	.89	.11	.00	.85	.15	.00

Table 3b: Break Selection Probabilities [$m = 1,\ \lambda_1^0 = .5,$ Serially Correlated Errors]

DGP	α	SEQ_R			SEQ_{BP}			SEQ_R			SEQ_{BP}			SEQ_R			SEQ_{BP}		
		P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u
		$T = 200$						$T = 400$						$T = 600$					
		(A) $\rho = .5,\ \theta = 0$																	
1	.5	.93	.07	.00	.34	.66	.00	.94	.06	.00	.38	.62	.00	.95	.05	.00	.42	.58	.00
	.7	.83	.06	.11	.46	.54	.00	.93	.07	.00	.46	.54	.00	.94	.06	.00	.45	.55	.00
2	.5	.92	.08	.00	.35	.65	.00	.93	.07	.00	.39	.61	.00	.93	.07	.00	.40	.60	.00
	.7	.91	.08	.01	.47	.53	.00	.93	.07	.00	.50	.50	.00	.93	.07	.00	.47	.53	.00
3	.7	.94	.06	.00	.76	.24	.00	.90	.10	.00	.79	.21	.00	.92	.08	.00	.86	.14	.00
	-.7	.88	.12	.00	.68	.32	.00	.88	.12	.00	.78	.22	.00	.90	.10	.00	.82	.18	.00
		(B) $\rho = 0,\ \theta = .5$																	
1	.5	.29	.09	.62	.50	.43	.07	.82	.09	.09	.51	.49	.00	.89	.10	.01	.52	.48	.00
	.7	.13	.04	.83	.46	.40	.14	.52	.05	.43	.47	.50	.03	.78	.09	.13	.50	.49	.01
2	.5	.56	.06	.38	.52	.40	.08	.87	.08	.05	.52	.48	.00	.90	.10	.00	.52	.48	.00
	.7	.31	.03	.66	.48	.37	.15	.69	.07	.24	.51	.46	.03	.87	.10	.03	.50	.50	.00
3	.7	.51	.11	.38	.51	.39	.10	.88	.11	.01	.63	.37	.00	.89	.11	.00	.75	.25	.00
	-.7	.65	.09	.26	.55	.38	.07	.87	.11	.02	.68	.31	.01	.90	.10	.00	.75	.25	.00
		(C) $\rho = 0,\ \theta = -.5$																	
1	.5	.90	.10	.00	.47	.53	.00	.94	.06	.00	.50	.50	.00	.95	.05	.00	.49	.51	.00
	.7	.68	.08	.24	.49	.50	.01	.95	.05	.00	.46	.54	.00	.93	.07	.00	.47	.53	.00
2	.5	.91	.09	.00	.49	.51	.00	.95	.05	.00	.49	.51	.00	.94	.06	.00	.49	.51	.00
	.7	.85	.09	.06	.50	.49	.01	.93	.07	.00	.50	.50	.00	.93	.07	.00	.50	.50	.00
3	.7	.92	.08	.00	.84	.16	.00	.92	.08	.00	.88	.12	.00	.93	.07	.00	.91	.09	.00
	-.7	.89	.11	.00	.79	.21	.00	.91	.09	.00	.88	.12	.00	.93	.07	.00	.88	.12	.00

Table 4a: Break Selection Probabilities [$m = 2,\ \rho = \theta = 0$]

DGP	α	SEQ_R			SEQ_{BP}			SEQ_R			SEQ_{BP}			SEQ_R			SEQ_{BP}		
		P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u
		$T = 200$						$T = 400$						$T = 600$					
		(A) $\lambda_1^0 = .3,\ \lambda_2^0 = .8$																	
4	.5	.74	.07	.19	.30	.60	.10	.89	.08	.03	.35	.63	.02	.91	.07	.02	.33	.66	.01
	.7	.49	.04	.47	.25	.46	.29	.79	.07	.14	.34	.58	.08	.85	.09	.06	.31	.65	.04
5	.5	.41	.12	.47	.46	.45	.09	.82	.17	.01	.50	.50	.00	.87	.13	.00	.51	.49	.00
	.7	.07	.03	.90	.36	.32	.32	.59	.18	.23	.46	.50	.04	.81	.17	.02	.49	.51	.00
6	.7	.78	.13	.09	.68	.27	.05	.90	.10	.00	.82	.18	.00	.88	.12	.00	.83	.17	.00
	-.7	.88	.05	.07	.47	.49	.04	.92	.08	.00	.61	.39	.00	.92	.08	.00	.70	.30	.00
		(B) $\lambda_1^0 = .2,\ \lambda_2^0 = .7$																	
4	.5	.74	.08	.18	.30	.58	.12	.90	.06	.04	.35	.62	.03	.92	.07	.01	.33	.66	.01
	.7	.40	.04	.56	.21	.45	.34	.76	.06	.18	.32	.58	.10	.87	.07	.06	.28	.69	.03
5	.5	.36	.09	.55	.46	.39	.15	.82	.15	.03	.49	.51	.00	.88	.12	.00	.50	.50	.00
	.7	.08	.03	.89	.33	.32	.35	.46	.12	.42	.48	.45	.07	.73	.15	.12	.48	.51	.01
6	.7	.73	.15	.12	.62	.29	.09	.89	.11	.00	.82	.18	.00	.89	.11	.00	.86	.14	.00
	-.7	.89	.06	.05	.48	.48	.04	.93	.07	.00	.60	.40	.00	.92	.08	.00	.71	.29	.00
		(C) $\lambda_1^0 = .4,\ \lambda_2^0 = .8$																	
4	.5	.63	.06	.31	.29	.55	.16	.80	.09	.11	.32	.62	.06	.86	.07	.07	.29	.33	.04
	.7	.38	.02	.60	.24	.47	.29	.67	.08	.25	.30	.56	.14	.78	.08	.14	.28	.65	.07
5	.5	.47	.11	.42	.50	.39	.11	.84	.15	.01	.52	.48	.00	.88	.12	.00	.53	.47	.00
	.7	.12	.02	.86	.33	.69	.36	.64	.17	.19	.48	.47	.05	.82	.16	.02	.51	.49	.00
6	.7	.78	.11	.11	.64	.30	.06	.91	.09	.00	.82	.18	.00	.88	.12	.00	.82	.18	.00
	-.7	.81	.06	.13	.49	.41	.10	.91	.09	.00	.64	.36	.00	.91	.09	.00	.71	.29	.00

Table 4b: Break Selection Probabilities [$m = 2,\ \lambda_1^0 = .3,\ \lambda_2^0 = .8,$ Serially Correlated Errors]

DGP	α	SEQ_R			SEQ_{BP}			SEQ_R			SEQ_{BP}			SEQ_R			SEQ_{BP}		
		P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u	P_c	P_0	P_u
		$T = 200$						$T = 400$						$T = 600$					
		(A) $\rho = .5,\ \theta = 0$																	
4	.5	.78	.10	.12	.19	.78	.03	.91	.04	.05	.28	.70	.02	.92	.06	.02	.27	.72	.01
	.7	.59	.09	.32	.21	.68	.11	.82	.06	.12	.34	.62	.04	.86	.06	.08	.32	.64	.04
5	.5	.64	.10	.26	.35	.63	.02	.90	.10	.00	.42	.58	.00	.90	.10	.00	.44	.56	.00
	.7	.28	.04	.68	.39	.51	.10	.84	.10	.06	.45	.55	.00	.91	.09	.00	.50	.50	.00
6	.7	.86	.13	.01	.59	.41	.00	.89	.11	.00	.78	.22	.00	.89	.11	.00	.79	.21	.00
	-.7	.90	.10	.00	.47	.53	.00	.91	.09	.00	.60	.40	.00	.92	.08	.00	.72	.28	.00
		(B) $\rho = 0,\ \theta = .5$																	
4	.5	.27	.03	.70	.21	.40	.39	.63	.05	.32	.38	.47	.15	.79	.05	.16	.40	.53	.07
	.7	.12	.02	.86	.16	.40	.44	.39	.03	.58	.29	.42	.29	.61	.05	.34	.34	.48	.18
5	.5	.04	.02	.94	.29	.24	.47	.24	.03	.73	.45	.37	.18	.54	.06	.40	.54	.41	.05
	.7	.02	.00	.98	.24	.20	.56	.07	.00	.93	.34	.33	.33	.19	.03	.78	.45	.33	.22
6	.7	.10	03	.87	.33	.24	.43	.52	.04	.44	.52	.24	.24	.76	.07	.17	.69	.25	.06
	-.7	.33	.07	.60	.24	.38	.38	.71	.05	.24	.44	.41	.15	.86	.05	.09	.56	.39	.05
		(C) $\rho = 0,\ \theta = -.5$																	
4	.5	.75	.08	.17	.28	.66	.06	.89	.04	.07	.36	.61	.03	.91	.06	.03	.35	.63	.02
	.7	.48	.05	.47	.25	.55	.20	.81	.05	.14	.36	.58	.06	.84	.06	.10	.35	.61	.04
5	.5	.52	.09	.39	.45	.50	.05	.92	.07	.01	.53	.47	.00	.92	.08	.00	.51	.49	.00
	.7	.13	.05	.82	.36	.45	.19	.73	.09	.18	.46	.53	.01	.90	.09	.01	.51	.49	.00
6	.7	.84	.12	.04	.74	.23	.03	.92	.08	.00	.87	.13	.00	.92	.08	.00	.89	.11	.00
	-.7	.90	.07	.03	.52	.44	.04	.92	.08	.00	.71	.29	.00	.96	.04	.00	.83	.17	.00

Table 5a: Level versus Persistence Breaks: True Model Selection Probabilities $[m = 1,\ \rho = \theta = 0]$

DGP	α	$T = 200$			$T = 400$			$T = 600$		
		$\lambda_1^0 = .5$	$\lambda_1^0 = .3$	$\lambda_1^0 = .7$	$\lambda_1^0 = .5$	$\lambda_1^0 = .3$	$\lambda_1^0 = .7$	$\lambda_1^0 = .5$	$\lambda_1^0 = .3$	$\lambda_1^0 = .7$
0_c	.5	.83	.85	.73	.89	.91	.78	.87	.90	.80
	.7	.78	.86	.61	.85	.89	.70	.85	.90	.77
1_c	.5	.99	.96	.92	1.0	1.0	1.0	1.0	1.0	1.0
	.7	.85	.84	.64	1.0	.97	.96	1.0	1.0	.99
2_c	.5	1.0	1.0	.99	1.0	1.0	1.0	1.0	1.0	1.0
	.7	.96	.89	.98	1.0	.99	.99	1.0	1.0	1.0
3_c	.7	1.0	1.0	.99	1.0	1.0	1.0	1.0	1.0	1.0
	-.7	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Table 5b: Trend versus Persistence Breaks: True Model Selection Probabilities $[m = 1,\ \rho = \theta = 0]$

DGP	α	$T = 200$			$T = 400$			$T = 600$		
		$\lambda_1^0 = .5$	$\lambda_1^0 = .3$	$\lambda_1^0 = .7$	$\lambda_1^0 = .5$	$\lambda_1^0 = .3$	$\lambda_1^0 = .7$	$\lambda_1^0 = .5$	$\lambda_1^0 = .3$	$\lambda_1^0 = .7$
0_t	.5	.91	.86	.87	.88	.87	.89	.86	.91	.87
	.7	.87	.86	.86	.87	.89	.88	.87	.90	.85
	1	.88	.88	.76	.92	.91	.82	.95	.84	.90
1_t	.5	.99	.93	.93	1.0	1.0	1.0	1.0	1.0	1.0
	.7	.83	.57	.64	1.0	.98	.98	1.0	1.0	1.0
2_t	.5	1.0	.92	.97	1.0	1.0	1.0	1.0	1.0	1.0
	.7	.83	.54	.72	1.0	.96	.99	1.0	1.0	1.0
3_t	.7	1.0	1.0	.99	1.0	1.0	1.0	1.0	1.0	1.0
	-.7	1.0	.99	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Technical Appendix

Let $B_1(.)$ and $B_2(.)$ denote standard independent Brownian motions on $[0,1]$ and $B(.) = [B_1(.), B_2(.)]'$. Further, let $\widetilde{B}_j^{(i)}(.)$ represent $B_j(.)$ demeaned over $[\lambda_{i-1}, \lambda_i]$, i.e., $\widetilde{B}_j^{(i)}(r) = B_j(r) - (\lambda_i - \lambda_{i-1})^{-1} \int_{\lambda_{i-1}}^{\lambda_i} B_j,\ r \in [\lambda_{i-1}, \lambda_i]$. The Brownian motions demeaned over the full sample are denoted as $\widetilde{B}_j(.) = B_j(.) - \int_0^1 B_j$. For brevity, all integrals of the form $\int_a^b g(r)dr$ are expressed as $\int_a^b g$. Also, for a generic variable a_t, define the quantities $\bar{a} = T^{-1}\sum_{t=1}^{T} a_t;\ \bar{a}_{-1} = T^{-1}\sum_{t=1}^{T} a_{t-1};\ \bar{a}_j = (T_j - T_{j-1})^{-1}\sum_{t=T_{j-1}+1}^{T_j} a_t;\ \bar{a}_{j,-1} = (T_j - T_{j-1})^{-1}\sum_{t=T_{j-1}+1}^{T_j} a_{t-1}$. Let $Y = (y_1, ..., y_T)',\ Y_{-1} = (y_0, ..., y_{T-1})',\ Y_j = (y_{T_{j-1}+1}, ..., y_{T_j})',\ Y_{j,-1} = (y_{T_{j-1}}, ..., y_{T_j-1})',\ e = (e_1, ..., e_T)',\ e_j = (e_{T_{j-1}+1}, ..., e_{T_j})',\ X = (x_1, ..., x_T)',\ x_t = (1, t)',\ X_i = (x_{T_{i-1}^0+1}, ..., x_{T_i^0})',\ \widetilde{M} = I_T - X(X'X)^{-1}X',\ \widetilde{M}_i = I_{T_i^0 - T_{i-1}^0} - X_i(X_i'X_i)^{-1}X_i'$ where I_T and $I_{T_i^0 - T_{i-1}^0}$ are identity matrices of order T and $(T_i^0 - T_{i-1}^0)$, respectively.

We first state two lemmas about the convergence of various sample moments whose proofs are standard and thus omitted.

Lemma A.1 *If $\{u_t\}$ is generated as $u_t = u_{t-1} + e_t,\ u_0 = 0$, where e_t satisfies Assumption A2, the following weak convergence results hold for $\lambda \in [0,1]$:*

a) $T^{-3/2}\sum_{t=1}^{[T\lambda]} u_{t-1} \Rightarrow \sigma \int_0^\lambda B_1$

b) $T^{-2}\sum_{t=1}^{[T\lambda]} u_{t-1}^2 \Rightarrow \sigma^2 \int_0^\lambda B_1^2$

c) $T^{-1}\sum_{t=1}^{[T\lambda]} u_{t-1}e_t \Rightarrow \sigma^2 \int_0^\lambda B_1(r)dB_1(r)$

Lemma A.2 *If $\{u_t\}$ is generated as $u_t = \alpha u_{t-1} + e_t,\ u_0 = 0,\ |\alpha| < 1$ where e_t satisfies Assumption A2, the following results hold uniformly over $\lambda \in [0,1]$:*

a) $T^{-1}\sum_{t=1}^{[T\lambda]} u_{t-1} \xrightarrow{p} 0$

b) $T^{-1}\sum_{t=1}^{[T\lambda]} u_{t-1}^2 \xrightarrow{p} \lambda\sigma^2/(1-\alpha^2)$

c) $T^{-1/2}\sum_{t=1}^{[T\lambda]} u_{t-1}e_t \Rightarrow \sigma^2(1-\alpha^2)^{-1/2}B_2(\lambda)$

Proof of Proposition 1: The BP test is defined as $G_1(k) = \sup_{\lambda \in \Lambda_\epsilon^k} G_1(\lambda, k)$ where

$$G_1(\lambda, k) = [T - 2(k+1)](SSR_0^{(0)} - SSR_{1,k}^{(0)})/[kSSR_{1,k}^{(0)}]$$

Defining $\hat{\alpha} = \left[\sum_{t=1}^{T}(y_{t-1} - \bar{y}_{-1})^2\right]^{-1}\sum_{t=1}^{T}(y_{t-1} - \bar{y}_{-1})y_t$, we have

$$SSR_0^{(0)} = \sum_{t=1}^{T}\left\{y_t - \bar{y} - \hat{\alpha}(y_{t-1} - \bar{y}_{-1})\right\}^2$$

Under $H_0^{(1)}$, $y_t = y_{t-1} + e_t,\ t \in \{1, ...T\}$ so that

$$\begin{aligned} SSR_0^{(0)} &= \sum_{t=1}^{T} \{(1-\hat{\alpha})(y_{t-1} - \bar{y}_{-1}) + e_t - \bar{e}\}^2 \\ &= -\frac{\left\{\sum_{t=1}^{T}(y_{t-1} - \bar{y}_{-1})e_t\right\}^2}{\sum_{t=1}^{T}(y_{t-1} - \bar{y}_{-1})^2} + \sum_{t=1}^{T} e_t^2 - T\bar{e}^2 \end{aligned} \tag{A.1}$$

Next, defining $\hat{\alpha}_i = \left[\sum_{t=T_{i-1}+1}^{T_i}(y_{t-1} - \bar{y}_{i,-1})^2\right]^{-1} \sum_{t=T_{i-1}+1}^{T_i}(y_{t-1} - \bar{y}_{i,-1})y_t$, we have

$$SSR_{1,k}^{(0)} = \sum_{i=1}^{k+1} \sum_{t=T_{i-1}+1}^{T_i} \{y_t - \bar{y}_i - \hat{\alpha}_i(y_{t-1} - \bar{y}_{i,-1})\}^2$$

Under $H_0^{(1)}$, we get

$$\begin{aligned} SSR_{1,k}^{(0)} &= \sum_{i=1}^{k+1} \sum_{t=T_{i-1}+1}^{T_i} \{(1-\hat{\alpha}_i)(y_{t-1} - \bar{y}_{i,-1}) + e_t - \bar{e}_i\}^2 \\ &= -\sum_{i=1}^{k+1}\left[\frac{\left\{\sum_{t=T_{i-1}+1}^{T_i}(y_{t-1} - \bar{y}_{i,-1})e_t\right\}^2}{\sum_{t=T_{i-1}+1}^{T_i}(y_{t-1} - \bar{y}_{i,-1})^2}\right] + \sum_{t=1}^{T} e_t^2 - \sum_{i=1}^{k+1}(T_i - T_{i-1})\bar{e}_i^2 \end{aligned} \tag{A.2}$$

Subtracting $(A.2)$ from $(A.1)$ and using Lemma A.1,

$$\begin{aligned} SSR_0^{(0)} - SSR_{1,k}^{(0)} \Rightarrow\ & \sigma^2 \left[\begin{array}{c} -\left(\int_0^1 \widetilde{B}_1^2\right)^{-1}\left(\int_0^1 \widetilde{B}_1 dB_1\right)^2 + \sum_{i=1}^{k+1}\left(\int_{\lambda_{i-1}}^{\lambda_i}\left[\widetilde{B}_1^{(i)}\right]^2\right)^{-1}\left(\int_{\lambda_{i-1}}^{\lambda_i}\widetilde{B}_1^{(i)}dB_1\right)^2 \\ -B_1(1)^2 + \sum_{i=1}^{k+1}(\lambda_i - \lambda_{i-1})^{-1}[B_1(\lambda_i) - B_1(\lambda_{i-1})]^2 \end{array} \right] \\ \equiv\ & \sigma^2 \left[\begin{array}{c} \sum_{i=1}^{k}\frac{[\lambda_i B_1(\lambda_{i+1}) - \lambda_{i+1}B_1(\lambda_i)]^2}{\lambda_{i+1}\lambda_i(\lambda_{i+1}-\lambda_i)} - \left(\int_0^1 \widetilde{B}_1^2\right)^{-1}\left(\int_0^1 \widetilde{B}_1 dB_1\right)^2 \\ + \sum_{i=1}^{k+1}\left(\int_{\lambda_{i-1}}^{\lambda_i}\left[\widetilde{B}_1^{(i)}\right]^2\right)^{-1}\left(\int_{\lambda_{i-1}}^{\lambda_i}\widetilde{B}_1^{(i)}dB_1\right)^2 \end{array} \right] \end{aligned}$$

The result follows since $[T - 2(k+1)]^{-1} SSR_{1,k}^{(0)} \xrightarrow{p} \sigma^2$. ▲

Proof of Theorem 1: We have

$$\begin{aligned} P(H_1(l+1|l) &> cv_{w,1}(\eta_{l+1})) \\ &= 1 - P(\max_{1 \le i \le l+1} H_1^{(i)}(1, \eta_{l+1}) \le cv_{w,1}(\eta_{l+1})) \end{aligned} \tag{A.3}$$

Consider two regimes j and j' where $j, j' \leq l+1$. Without loss of generality, assume $j < j'$. If j and j' are (non-adjacent) $I(1)$ regimes, the limit of $H_1^{(j)}$ is a function of $\{B_{1,j}(r) = B_1(r) - B_1(\lambda_{j-1}^0);\ r \in [\lambda_{j-1}^0, \lambda_j^0]\}$ while the limit of $H_1^{(j')}$ is a function of $\{B_{1,j'}(r') = B_1(r') - B_1(\lambda_{j'-1}^0);\ r' \in [\lambda_{j'-1}^0, \lambda_{j'}^0]$. This observation is based on the fact that under the null hypothesis of l breaks, $\left|\hat{T}_i - T_i^0\right| = O_p(1)$ for $i = 1, ..., l$. Since $B_{1,j}(.)$ and $B_{1,j'}(.)$ are independent, it follows that $H_1^{(j)}$ and $H_1^{(j')}$ are asymptotically independent.

Next, if j and j' are (possibly adjacent) $I(0)$ regimes, then the limit of $H_1^{(j)}$ is a function of $\{\bar{B}_j(r) = B(r) - B(\lambda_{j-1}^0);\ r \in [\lambda_{j-1}^0, \lambda_j^0]\}$ while the limit of $H_1^{(j')}$ is a function of $\{\bar{B}_{j'}(r') = B(r') - B(\lambda_{j'-1}^0);\ r' \in [\lambda_{j'-1}^0, \lambda_{j'}^0]$. Again, $H_1^{(j)}$ and $H_1^{(j')}$ are asymptotically independent to the independence of $\bar{B}_j(.)$ and $\bar{B}_{j'}(.)$. Further, $B_{1,j}(.)$ and $\bar{B}_{j'}(.)$ are independent as are $B_{1,j'}(.)$ and $\bar{B}_{j'}(.)$ (recall that $B_1(.)$ and $B_2(.)$ are independent).

The above arguments show that $H_1^{(i)}$ are asymptotically independent for $i = 1, ..., l+1$ regardless of whether the l breaks involve switches between $I(1)$ and $I(0)$ regimes or between different $I(0)$ regimes. Then, using that $H_1^{(i)}(1, \eta_{l+1})$ has asymptotic level at most η_{l+1}, we have from $(A.3)$,

$$\begin{aligned}\lim_{T\to\infty} P(H_1(l+1|l) &\leq 1 - \prod_{i=1}^{l+1}\left\{1 - \eta_{l+1}\right\} \\ &= 1 - (1-\eta_{l+1})^{l+1} = 1 - \left\{(1-\eta)^{1/(l+1)}\right\}^{l+1} = \eta\end{aligned}$$

This proves Theorem 1. ▲

Proof of Theorem 2: Define the events

$$\begin{aligned}A_0 &= \{Hmax_1(\eta) > cv_{w,max}(\eta)\} \\ A_l &= \left\{H_1(l+1|l) > cv_{w,1}(\eta_{l+1})\right\};\ \ l = 1, ..., m\end{aligned}$$

We have

$$\begin{aligned}P(\hat{m} = m) &= P\left[\left\{\bigcap_{l=0}^{m-1} A_l\right\} \bigcap A_m^c\right] \\ &\geq P\left[\bigcap_{l=0}^{m-1} A_l\right] + P\left[A_m^c\right] - 1 \\ &\geq \sum_{l=0}^{m-1} P(A_l) - (m-1) + P\left[A_m^c\right] - 1 \qquad (A.4)\end{aligned}$$

Now, $P(A_l) \to 1$ for $l = 0, ..., m-1$ since with m breaks each of the $H_1(l+1|l)$ tests are consistent, given that at least one of the $(l+1)$ regimes contain a break. Further, $\lim_{T\to\infty} P\left[A_m^c\right] \geq 1 - \eta$ by Theorem 1. Thus, from $(A.4)$, we get

$$\lim_{T\to\infty} P(\hat{m} = m) \geq m - (m-1) + 1 - \eta - 1 = 1 - \eta$$

which completes the proof of Theorem 2. ▲

Proof of Theorem 3: We have

$$\widetilde{SSR}_0^{(0)} = \left(\widetilde{M}Y - \widetilde{M}Y_{-1}\hat{\alpha}\right)' \left(\widetilde{M}Y - \widetilde{M}Y_{-1}\hat{\alpha}\right)$$

with $\hat{\alpha} = \left(Y_{-1}'\widetilde{M}Y_{-1}\right)^{-1} Y_{-1}'\widetilde{M}Y$. Under $\widetilde{H}_0^{(0)}$, we get

$$\begin{aligned}
\widetilde{SSR}_0^{(0)} &= \left[\widetilde{M}Y_{-1}(\alpha - \hat{\alpha}) + \widetilde{M}e\right]' \left[\widetilde{M}Y_{-1}(\alpha - \hat{\alpha}) + \widetilde{M}e\right] \\
&= -\left(Y_{-1}'\widetilde{M}Y_{-1}\right)^{-1} \left(e'\widetilde{M}Y_{-1}\right)^2 + e'\widetilde{M}e \\
&= -\left(Y_{-1}'\widetilde{M}Y_{-1}\right)^{-1} \left(e'\widetilde{M}Y_{-1}\right)^2 + e'e - e'X(X'X)^{-1}X'e
\end{aligned}$$

Next, with $\hat{\alpha}_i = \left(Y_{i,-1}'\widetilde{M}_i Y_{i,-1}\right)^{-1} Y_{i,-1}'\widetilde{M}_i Y_i$ we have

$$SSR_{2,k}^{(0)} = \sum_{i=1}^{k+1} \left(\widetilde{M}_i Y_i - \widetilde{M}_i Y_{i,-1}\,\hat{\alpha}_i\right)' \left(\widetilde{M}_i Y_i - \widetilde{M}_i Y_{i,-1}\,\hat{\alpha}_i\right) \tag{A.5}$$

Under $\widetilde{H}_0^{(0)}$, we get

$$\begin{aligned}
SSR_{2,k}^{(0)} &= -\sum_{i=1}^{k+1} \left[\left(Y_{i,-1}'\widetilde{M}_i Y_{i,-1}\right)^{-1} \left(Y_{i,-1}'\widetilde{M}_i e_i\right)^2\right] + \sum_{i=1}^{k+1} e_i'\widetilde{M}_i e_i \\
&= -\sum_{i=1}^{k+1} \left[\left(Y_{i,-1}'\widetilde{M}_i Y_{i,-1}\right)^{-1} \left(Y_{i,-1}'\widetilde{M}_i e_i\right)^2\right] + e'e - \sum_{i=1}^{k+1} \left[e_i'X_i(X_i'X_i)^{-1}X_i'e_i\right]
\end{aligned}$$

Denoting $F(r) = (1, r)'$ and using the facts that

$$\begin{aligned}
&T^{-1}Y_{-1}'\widetilde{M}Y_{-1} \xrightarrow{p} \sigma^2/(1-\alpha^2) \\
&T^{-1/2}Y_{-1}'\widetilde{M}e \xrightarrow{d} \left[\sigma^2/(1-\alpha^2)^{1/2}\right] B_2(1) \\
&T^{-1}Y_{i,-1}'\widetilde{M}_i Y_{i,-1} \xrightarrow{p} (\lambda_i - \lambda_{i-1})\,\sigma^2/(1-\alpha^2) \\
&T^{-1/2}Y_{i,-1}'\widetilde{M}_i e_i \xrightarrow{d} [\sigma^2/(1-\alpha^2)^{1/2}][B_2(\lambda_i) - B_2(\lambda_{i-1})] \\
&e'X(X'X)^{-1}X'e \xrightarrow{d} \sigma^2 \left[\int_0^1 F(r)dB_1(r)\right]' \left[\int_0^1 F(r)F(r)'\right]^{-1} \left[\int_0^1 F(r)dB_1(r)\right] \\
&\sum_{i=1}^{k+1} \left[e_i'X_i(X_i'X_i)^{-1}X_i'e_i\right] \xrightarrow{d} \sigma^2 \left[\int_{\lambda_{i-1}}^{\lambda_i} F(r)dB_1(r)\right]' \left[\int_{\lambda_{i-1}}^{\lambda_i} F(r)F(r)'\right]^{-1} \left[\int_{\lambda_{i-1}}^{\lambda_i} F(r)dB_1(r)\right] \\
&[T - 3(k+1)]^{-1} SSR_{2,k}^{(0)} \xrightarrow{p} \sigma^2
\end{aligned}$$

the result follows. ▲

Proof of Theorem 4: Define $z_t = 1$ for $t = 1, ..., T$ and the $[T \times (m+1)]$ matrix $\breve{Z} = [\breve{z}_1, ..., \breve{z}_{m+1}]$ where $\breve{z}_i = (0, ..., 0, z_{\hat{T}_{i-1}+1}, z_{\hat{T}_{i-1}+2}, ..., z_{\hat{T}_i}, 0, ..., 0)'$. Further, define the $(T \times 2)$ matrix $W = (w_1, ..., w_T)'$ where $w_t = (1, y_{t-1})'$ and the $[T \times 2(m+1)]$ matrix $\breve{W} = [\breve{w}_1, ..., \breve{w}_{m+1}]$ where $\breve{w}_i = (0, ..., 0, w_{\hat{T}_{i-1}+1}, w_{\hat{T}_{i-1}+2}, ..., w_{\hat{T}_i}, 0, ..., 0)'$. When evaluated at the true break dates $(T_1^0, ..., T_m^0)$, we denote the counterparts to $\breve{Z}$ and $\breve{W}$ as $\breve{Z}^0$ and $\breve{W}^0$, respectively. Finally, let $\gamma = (\gamma_1, ..., \gamma_{m+1})'$ with $\gamma_i = (c_i, \alpha_i)$. Then the Wald statistic can be expressed as

$$W^*(m) = (T - 2(m+1))(SSR_0^* - SSR_m^*)/SSR_m^*$$

where

$$\begin{aligned} SSR_0^* &= (M_{\breve{Z}}Y - M_{\breve{Z}}Y_{-1}\hat{\alpha})'(M_{\breve{Z}}Y - M_{\breve{Z}}Y_{-1}\hat{\alpha})' \\ SSR_m^* &= \left(Y - \breve{W}\hat{\gamma}\right)'\left(Y - \breve{W}\hat{\gamma}\right) \end{aligned}$$

and

$$\begin{aligned} M_{\breve{Z}} &= I_T - \breve{Z}(\breve{Z}'\breve{Z})^{-1}\breve{Z}' \\ \hat{\alpha} &= \left(Y_{-1}'M_{\breve{Z}}Y_{-1}\right)^{-1}Y_{-1}'M_{\breve{Z}}Y \\ \hat{\gamma} &= \left(\breve{W}'\breve{W}\right)^{-1}\breve{W}'Y \end{aligned}$$

Under the null hypothesis $\alpha_i = \alpha$ for all i with $|\alpha_i| < 1$, we have

$$\begin{aligned} SSR_0^* &= \left[M_{\breve{Z}}(\breve{Z}\delta + Y_{-1}\alpha + e) - M_{\breve{Z}}Y_{-1}\hat{\alpha}\right]'\left[M_{\breve{Z}}(\breve{Z}^0 + Y_{-1}\alpha + e) - M_{\breve{Z}}Y_{-1}\hat{\alpha}\right] \\ &= \left[M_{\breve{Z}}\left(\breve{Z}^0 - \breve{Z}\right)\delta + M_{\breve{Z}}e + M_{\breve{Z}}Y_{-1}(\alpha - \hat{\alpha})\right]'\left[M_{\breve{Z}}\left(\breve{Z}^0 - \breve{Z}\right)\delta + M_{\breve{Z}}e + M_{\breve{Z}}Y_{-1}(\alpha - \hat{\alpha})\right] \end{aligned}$$

Then, using the fact that $\left|\hat{T}_i - T_i^0\right| = O_p(1)$ for $i = 1, ..., m$, we can show that

$$\begin{aligned} SSR_0^* &= -\left(Y_{-1}'M_{\breve{Z}^0}Y_{-1}\right)^{-1}\left(Y_{-1}'M_{\breve{Z}^0}e\right)^2 + e'M_{\breve{Z}^0}e + o_p(1) \\ &= -\left(Y_{-1}'M_{\breve{Z}^0}Y_{-1}\right)^{-1}\left(Y_{-1}'M_{\breve{Z}^0}e\right)^2 + e'e \\ &\quad - \sum_{i=1}^{m+1}\left(T_i^0 - T_{i-1}^0\right)^{-1}\left(\sum_{t=T_{i-1}^0+1}^{T_i^0} e_t\right)^2 + o_p(1) \end{aligned} \tag{A.6}$$

Using similar arguments, we have

$$\begin{aligned} SSR_m^* &= -\left(e'\breve{W}\right)\left(\breve{W}^{0\prime}\breve{W}^0\right)^{-1}\breve{W}^{0\prime}e + e'e + o_p(1) \\ &= e'e - \sum_{i=1}^{m+1}\left(\sum_{t=T_{i-1}^0+1}^{T_i^0} w_t e_t\right)'\left(\sum_{t=T_{i-1}^0+1}^{T_i^0} w_t w_t'\right)^{-1}\left(\sum_{t=T_{i-1}^0+1}^{T_i^0} w_t e_t\right) + o_p(1) \end{aligned} \tag{A.7}$$

Subtracting $(A.7)$ from $(A.6)$, we get

$$\begin{aligned} SSR_0^* - SSR_m^* &= -\left[T^{-1}\sum_{i=1}^{m+1}\sum_{t=T_{i-1}^0+1}^{T_i^0}\left(y_{t-1}-\bar{y}_{i,-1}\right)^2\right]^{-1}\left[T^{-1/2}\sum_{i=1}^{m+1}\sum_{t=T_{i-1}^0+1}^{T_i^0}\left(y_{t-1}-\bar{y}_{i,-1}\right)e_t\right]^2 \\ &\quad -\sum_{i=1}^{m+1}\left(T_i^0-T_{i-1}^0\right)^{-1}\left(\sum_{t=T_{i-1}^0+1}^{T_i^0}e_t\right)^2 \\ &\quad +\sum_{i=1}^{m+1}\left(\sum_{t=T_{i-1}^0+1}^{T_i^0}w_te_t\right)'\left(\sum_{t=T_{i-1}^0+1}^{T_i^0}w_tw_t'\right)^{-1}\left(\sum_{t=T_{i-1}^0+1}^{T_i^0}w_te_t\right)+o_p(1) \\ &= S1+S2+S3+o_p(1) \end{aligned} \tag{A.8}$$

We consider each of the terms $S1$-$S3$ in turn. For $S1$,we have

$$T^{-1}\sum_{i=1}^{m+1}\sum_{t=T_{i-1}^0+1}^{T_i^0}\left(y_{t-1}-\bar{y}_{i,-1}\right)^2 \xrightarrow{p} \sigma^2/(1-\alpha^2)$$

$$T^{-1/2}\sum_{i=1}^{m+1}\sum_{t=T_{i-1}^0+1}^{T_i^0}\left(y_{t-1}-\bar{y}_{i,-1}\right)e_t \Rightarrow \sigma^2(1-\alpha^2)^{-1/2}B_2(1)$$

so that $S1 \Rightarrow -\sigma^2 B_2(1)^2$. For $S2$,

$$\begin{aligned} -\sum_{i=1}^{m+1}\left(T_i^0-T_{i-1}^0\right)^{-1}\left(\sum_{t=T_{i-1}^0+1}^{T_i^0}e_t\right)^2 &= -\sum_{i=1}^{m+1}\left(\frac{T_i^0-T_{i-1}^0}{T}\right)^{-1}\left(T^{-1/2}\sum_{t=T_{i-1}^0+1}^{T_i^0}e_t\right)^2 \\ &\Rightarrow -\sigma^2\sum_{i=1}^{m+1}\left(\lambda_i^0-\lambda_{i-1}^0\right)^{-1}\left[B_1(\lambda_i^0)-B_1(\lambda_{i-1}^0)\right]^2 \end{aligned}$$

Next, for $S3$ we have

$$\begin{aligned} &\sum_{i=1}^{m+1}\left(\sum_{t=T_{i-1}^0+1}^{T_i^0}w_te_t\right)'\left(\sum_{t=T_{i-1}^0+1}^{T_i^0}w_tw_t'\right)^{-1}\left(\sum_{t=T_{i-1}^0+1}^{T_i^0}w_te_t\right) \\ \Rightarrow\ &\sigma^2\sum_{i=1}^{m+1}\left(\lambda_i^0-\lambda_{i-1}^0\right)^{-1}\left[\left\{B_1(\lambda_i^0)-B_1(\lambda_{i-1}^0)\right\}^2+\left\{B_2(\lambda_i^0)-B_2(\lambda_{i-1}^0)\right\}^2\right] \end{aligned}$$

Finally, collecting the limit results for $S1$-$S3$ and using the fact that $[T-2(m+1)]^{-1}SSR_m^* \xrightarrow{p}$

σ^2, we get,

$$\begin{aligned} W^*(m) &\Rightarrow -B_2(1)^2 + \sum_{i=1}^{m+1} \left(\lambda_i^0 - \lambda_{i-1}^0\right)^{-1} \left[B_2(\lambda_i^0) - B_2(\lambda_{i-1}^0)\right]^2 \\ &\equiv \sum_{i=1}^{m} \frac{[\lambda_i^0 B_2(\lambda_{i+1}^0) - \lambda_{i+1}^0 B_2(\lambda_i^0)]^2}{\lambda_{i+1}^0 \lambda_i^0 (\lambda_{i+1}^0 - \lambda_i^0)} \end{aligned}$$

which is a $\chi^2(m)$ random variable. This proves the theorem. ▲

Proof of Theorem 5: (a) The proof is omitted since it follows using arguments very similar to those used in the proof of Theorem 4.

(b) The Wald statistic can be expressed as

$$\widetilde{W}_1(m) = (T - 3(m+1))(\widetilde{SSR}_0^{(1)} - SSR_{2,m}^{(1)})/SSR_{2,m}^{(1)}$$

with

$$\begin{aligned} \widetilde{SSR}_0^{(1)} &= \sum_{i=1}^{m+1} \sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} \left(y_t - y_{t-1} - (\hat{T}_i - \hat{T}_{i-1})^{-1} \sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} (y_t - y_{t-1}) \right)^2 \\ SSR_{2,m}^{(1)} &= \sum_{i=1}^{m+1} \sum_{t=\hat{T}_{i-1}+1}^{\hat{T}_i} (\Delta y_t - \hat{c}_i - \hat{b}_i t - \hat{\alpha}_i y_{t-1})^2 \end{aligned}$$

where $(\hat{c}_i, \hat{b}_i, \hat{\alpha}_i)$ are obtained from unrestricted OLS estimation using data in the (estimated) regime i. Again, using $\left|\hat{T}_i - T_i^0\right| = O_p(1)$ for $i = 1, , , ., m$, we have under the null hypothesis $\alpha_i = 1$ for all i,

$$\begin{aligned} \widetilde{SSR}_0^{(1)} &= -\sum_{i=1}^{m+1} (T_i^0 - T_{i-1}^0)^{-1} \left(\sum_{t=T_{i-1}^0+1}^{T_i^0} e_t \right)^2 + e'e + o_p(1) \\ SSR_{2,m}^{(1)} &= e'e - \sum_{i=1}^{m+1} \left\{ (e_i' X_i)(X_i' X_i)^{-1} (X_i' e_i) \right\} - \sum_{i=1}^{m+1} \left(Y_{i,-1}' \widetilde{M}_i Y_{i,-1} \right)^{-1} \left(Y_{i,-1}' \widetilde{M}_i e_i \right)^2 \end{aligned} \quad \text{(A.9)}$$

Next, define $\widetilde{Y}_{i,-1} = (\widetilde{y}_{T_{i-1}^0}, ..., \widetilde{y}_{T_i^0 - 1})'$, where $\widetilde{y}_t = \sum_{s=T_{i-1}^0}^{t} e_s$. Then from $(A.9)$, we have

$$\begin{aligned} \widetilde{SSR}_0^{(1)} - SSR_{2,m}^{(1)} &= \sum_{i=1}^{m+1} \left(\widetilde{Y}_{i,-1}' \widetilde{M}_i \widetilde{Y}_{i,-1} \right)^{-1} \left(\widetilde{Y}_{i,-1}' \widetilde{M}_i e_i \right)^2 + \sum_{i=1}^{m+1} \left\{ (e_i' X_i)(X_i' X_i)^{-1} (X_i' e_i) \right\} \\ &\quad - \sum_{i=1}^{m+1} (T_i^0 - T_{i-1}^0)^{-1} \left(\sum_{t=T_{i-1}^0+1}^{T_i^0} e_t \right)^2 + o_p(1) \end{aligned} \quad \text{(A.10)}$$

Now, we use the following facts:

$$\begin{aligned} T^{-2}\widetilde{Y}'_{i,-1}\widetilde{M}_i\widetilde{Y}_{i,-1} &\Rightarrow \sigma^2 \int_{\lambda_{i-1}^0}^{\lambda_i^0} \left[\breve{B}_1^{(i)}\right]^2 \\ T^{-1}\widetilde{Y}'_{i,-1}\widetilde{M}_i e_i &\Rightarrow \sigma^2 \int_{\lambda_{i-1}^0 1}^{\lambda_i^0} \breve{B}_1^{(i)} dB_1 \\ T^{-1/2} \sum_{t=T_{i-1}^0+1}^{T_i^0} e_t &\Rightarrow \sigma[B_1(\lambda_i^0) - B_1(\lambda_{i-1}^0)] \end{aligned} \tag{A.11}$$

Using $(A.11)$ in $(A.10)$, we get

$$\widetilde{SSR}_0^{(1)} - SSR_{2,m}^{(1)} \Rightarrow \sigma^2 \sum_{i=1}^{m+1} \left\{ \begin{array}{c} \left(\int_{\lambda_{i-1}^0}^{\lambda_i^0} \left[\breve{B}_1^{(i)}\right]^2\right)^{-1} \left(\int_{\lambda_{i-1}^0 1}^{\lambda_i^0} \breve{B}_1^{(i)} dB_1\right)^2 + \\ \left(\int_{\lambda_{i-1}^0}^{\lambda_i^0} F(r) dB_1(r)\right)' \left(\int_{\lambda_{i-1}^0}^{\lambda_i^0} F(r)F(r)'\right)^{-1} \left(\int_{\lambda_{i-1}^0}^{\lambda_i^0} F(r) dB_1(r)\right) \\ - \frac{\left[B_1(\lambda_i^0) - B_1(\lambda_{i-1}^0)\right]^2}{\lambda_i^0 - \lambda_{i-1}^0} \end{array} \right\}$$

The result then follows from recognizing that $(T - 3(m+1))^{-1} SSR_{2,m}^{(1)} \xrightarrow{p} \sigma^2$. ▲