ELSEVIER

Journal of Econometrics 91 (1999) 299–323

JOURNAL OF Econometrics

www.elsevier.nl/locate/econbase

Likelihood ratio tests for multiple structural changes

Jushan Bai*

Department of Economics, E52-274B, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received 1 April 1997; received in revised form 1 June 1998; accepted 12 August 1998

Abstract

This paper proposes a likelihood-ratio-type test for multiple structural changes in regression models. The model allows for lagged-dependent variables and trending regressors. The limiting distribution of the test is derived. We show that asymptotic critical values can be obtained analytically. In addition, the number and the locations of change points can be consistently determined via the test procedure. The method is straightforward to implement. © 1999 Elsevier Science S.A. All rights reserved.

JEL classification: C12; C22; C13; C52

Keywords: Structural change; Multiple change points; Hypothesis testing; Dynamic models; Trending regressors; Limiting distribution

1. Introduction

Testing for structural change has always been an important issue in econometrics because a myriad of political and economic factors can cause the relationships among economic variables to change over time. Since the early work of Chow (1960) and Quandt (1960), numerous studies have been undertaken on the issue of structural changes.[1] However, most of the existing work focuses on testing for a single change. In this paper, we propose a likelihood-ratio-type test for multiple changes. The proposed test is an exact

* Tel. (617) 253-6217; fax: (617) 253-1330; e-mail: jbai@mit.edu

[1] This is evidenced by the two special volumes of this Journal edited, respectively, by Broemeling (1982), and Dufour and Ghysels (1996), as well as by a number of monographs on this subject, e.g., Poirier (1979), Kramer (1989), and Hackl and Westlund (1991).

0304-4076/99/$ - see front matter
PII: S0304-4076(98)00079-7

likelihood-ratio test under the assumption of normality. Since normality is not assumed, the test may be considered a pseudo-Gaussian likelihood-ratio test. The limiting distribution of the test is derived in this paper. Critical values can be calculated analytically.

When testing for a single break, the limiting distribution of the test statistic is derived in the absence of breaks (typically by direct application of the functional central limit theorem under the null of no change). In this paper, both null and alternative hypotheses allow for the existence of break points and thus involve estimating these points. As a result, this analysis differs considerably from the existing literature. In particular, functional central limit theorems cannot be directly applied to derive the limiting distribution. The behavior of break-point estimators under both the null and the alternative hypotheses must be taken into account.

In addition to having the capability of detecting the presence of a break, the proposed test can be used to determine the number and locations of breaks in the data. This is of both theoretical and practical interest. Under the current practice, when the null hypothesis is rejected, a single break point is estimated. In many cases, there is no reason to assume that only one break point is present, especially for data covering an extended period of time. The possibility of multiple changes is consistent with the notion of multiple equilibria proposed by economic theory, with each change (regime) representing a new equilibrium condition. Rudebusch and Diebold (1994) explain the links between the structural-change model and multiple equilibria of economic theory. Hegwood and Papell (1996) suggest the existence of multiple regimes in the real exchange-rate data. In an interesting application of the structural change method, Willard et al. (1996) study how people during the Civil War period responded to various events that were happening around them and compare the relative importance of those events (as implied by the data) to the accounts of traditional historians.

The null hypothesis of our test assumes ℓ break points, whereas the alternative hypothesis assumes $\ell + 1$ break points. (When $\ell = 0$, the test reduces to the usual test of no change against a single change.) Due to the format of these hypotheses, when the test is performed repeatedly while augmenting the value of ℓ, the number of break-points can be consistently estimated.

A related test is set out in Bai and Perron (1994). They also propose to test ℓ versus $\ell + 1$ breaks. Their test does not have a likelihood-ratio interpretation. Whenever an additional break point is to be estimated, it is obtained conditional on the previously obtained break points. For example, suppose that ℓ break points have been computed, the $(\ell + 1)$th break point is estimated from one of the subsamples separated by the given ℓ break points. In contrast, in the likelihood-ratio setting, the model is estimated optimally under both null and alternative hypotheses. This entails estimating ℓ breaks simultaneously under the null and estimating $\ell + 1$ breaks simultaneously under the alternative. The limiting distribution of the LR test is shown to be of a similar – but not identical

– form to that of the Bai–Perron test. Although its theoretical justification requires considerable effort, the LR procedure itself is straightforward to implement.

This paper makes two additional contributions to the existing literature. First, the procedure can deal with dynamic regressors (lagged dependent variables). The existing literature contains many results admitting dynamic regressors, for example, Kramer et al. (1988). In those studies, the model, and particularly the regressors, are stationary under the null.[2] In a dynamic setting, once a shift occurs, the regressors cease to be stationary. Since our null hypothesis allows for shifts, both the dependent and the explanatory variables are non-stationary.[3] Second, we consider testing multiple breaks in polynomial trends. We not only characterize the limiting distribution, we also derive analytical expressions for asymptotic critical values.

The remainder of this paper is organized as follows. In Section 2, we state the model and assumptions. Section 3 defines the test statistic and gives the main result. The assumption of regime-wise asymptotic stationarity is assumed in this section. Section 4 extends the test to include polynomial trends. Limited Monte Carlo results are reported in Section 5, and some remarks are given in Section 6. Technical proofs are provided in the Appendix.

2. Model and assumption

Consider the following m-break model:

$$\begin{aligned} y_t &= z_t'\delta_1 + u_t, && t = 1, 2, \ldots, k_1^0, \\ y_t &= z_t'\delta_2 + u_t, && t = k_1^0 + 1, \ldots, k_2^0, \\ &\vdots && \vdots \\ y_t &= z_t'\delta_{m+1} + u_t, && t = k_m^0 + 1, \ldots, T, \end{aligned} \tag{1}$$

where y_t is the observed dependent variable at time t; z_t $(q \times 1)$ is a vector of covariates; δ_j $(j = 1, \ldots, m+1)$ is a vector of coefficients with $\delta_i \neq \delta_{i+1}$ $(i = 1, \ldots, m)$; and u_t is the disturbance at time t. The break points $k_1^0, \ldots, k_m^0$ as well as the number of breaks m are also unknown. When m is known, all parameters including the break points can be easily estimated. Their statistical properties are investigated by Bai and Perron (1994). In practice, the number of

[2] Some exact tests may not require stationarity, but require more specific distributional assumptions such as normality, e.g., Dufour and Kiviet (1996).

[3] This paper, however, does not consider integrated regressors. For this, readers are referred to Vogelsang (1994), Banerjee et al. (1992), Hansen (1992), and Zivot and Andrews (1992).

breaks, m, is unknown. It is of importance to make an inference about the value of m. We propose a test statistic for testing the null hypothesis of ℓ versus the alternative of $\ell + 1$ breaks. We show how this test can lead to a consistent estimate for m. Before discussing the test statistics, we make additional assumptions as follows.

Assumption 1. The regressors are regime-wise asymptotically stationary. More specifically, let $\Delta k_i^0 = k_{i+1}^0 - k_i^0$, for $i = 0, \ldots, m$,

$$\operatorname{plim} \frac{1}{\Delta k_i^0} \sum_{t=k^0{}_i+1}^{k^0{}_i+[\Delta k^0{}_i v]} z_t z_t' = vQ_i$$

uniformly in $v \in [0, 1]$, where Q_i is a positive definite matrix. Throughout, we adopt the convention that $k_0^0 = 0$ and $k_{m+1}^0 = T$.

Assumption 2. $\{u_t, \mathscr{F}_t\}$ forms a sequence of martingale differences where $\mathscr{F}_t = \sigma$-field $\{z_s, u_s; s \leqslant t\}$ with $Eu_t^2 = \sigma^2$ for all t, and $\sup_t \mathrm{E}|u_t|^{4+\delta} < \infty$.

Assumption 3. The functional central limit theorem holds for $z_t u_t$ in each regime. That is, for $i = 0, \ldots, m$,

$$\frac{1}{(\Delta k_i^0)^{1/2}} \sum_{t=k_i^0+1}^{k_i^0+[\Delta k^0{}_i v]} z_t u_t' \Rightarrow W_{i,q}(v),$$

where $W_{i,q}(v)$ is a q-dimensional Wiener process on $[0, 1]$ with variance–covariance matrix $\sigma^2 v Q_i$, and $W_{1,q}, \ldots, W_{m+1,q}$ are independent of each other.

Assumption 4. $\delta_i \neq \delta_{i+1}$, $k_i^0 = [T\tau_i^0]$, $\tau_i^0 \in (0,1)$ with $\tau_i{}^0 < \tau_{i+1}^0$ $(i = 1, \ldots, m)$.

Assumption 1 is satisfied by within-regime i.i.d. regressors having a positive-definite variance–covariance matrix. Assumption 1 is also satisfied by any within-regime second-order stationary process such that the strong law of large numbers holds for $z_t z_t'$. In these cases, $Q_i = \mathrm{E} z_t z_t'$. Assumption 1 is clearly weaker than the global asymptotic stationarity of the following:

$$\operatorname{Plim} \frac{1}{T} \sum_{t=1}^{[Tv]} z_t z_t' = vQ, \tag{2}$$

uniformly in $v \in [0, 1]$. Autoregressive models with structural changes will not satisfy the global asymptotic stationarity (global-stationary) (2) but do satisfy Assumption 1. In this case, Q_i is the second-moment matrix of a stationary autoregressive process with autoregressive parameter δ_i. Assumptions 2 and 3

are standard. But Assumption 2 rules out serially correlated disturbances. The result of this paper can be extended to dependent disturbances such that u_t forms a sequence of mixingales with respect to an increasing sequence of sigma-fields but uncorrelated with the regressors. For simplicity, we shall focus on uncorrelated errors, making it possible to include lagged dependent variables. Assumption 4 is a standard technical device for asymptotic purposes.

We next introduce some terminology and notation. A subsample $[i, j]$ means a subsample consists of observations from i to j (inclusive). An m-partition is a vector of m integers, $(k_1, \ldots, k_m)$, such that $1 < k_1 < k_2 < \ldots k_m < T$. Each partition divides the whole sample into $m+1$ subsamples (segments). For a small positive number $\pi > 0$, we define a set of partition, Λ_π, by

$$\Lambda_\pi = \{(k_1, \ldots, k_m): k_i - k_{i-1} \geqslant \pi T; i = 1, \ldots, m+1\}, \tag{3}$$

where $k_0 = 0$ and $k_{m+1} = T$. Each of the $m+1$ subsamples resulting from a partition in Λ_π consists of at least a positive fraction of the observations. We assume π is small so that $(k_1^0, \ldots, k_m^0) \in \Lambda_\pi$.

Let $S_T(k_1, \ldots, k_m)$ denote the sum of squared residuals corresponding to the partition $(k_1, \ldots, k_m)$. That is,

$$S_T(k_1, \ldots, k_m) = \sum_{i=1}^{m+1} \inf \phi_i \sum_{k_{i-1}+1}^{k_i} (y_t - z_t' \phi_i)^2,$$

where $k_0 = 0$ and $k_{m+1} = T$. The optimal SSR is defined as

$$S_T(\hat{k}_1, \ldots, \hat{k}_m) = \min_{(k_1, \ldots, k_m)} S_T(k_1, \ldots, k_m),$$

where the minimization is taken over Λ_π. In practice, π is set to 0.05, or 0.10. Assuming there are m breaks, then $(\hat{k}_1, \ldots, \hat{k}_m)$ forms an estimator for $(k_1^0, \ldots, k_m^0)$. If we define $\hat{\tau}_i = \hat{k}_i / T$, the break fraction, then $\hat{\tau}_i$ is T consistent for τ_i^0. That is, $T(\hat{\tau}_i - \tau_i^0) = \hat{k}_i - k_i^0 = O_p(1)$, see Bai and Perron (1994). For simplicity, we shall say $\hat{k}_i$ is T consistent for k_i^0, with the understanding that we are referring to the break fractions.

In the next section, we discuss the test statistics based on the optimal SSRs.

3. The test statistic

This section considers testing the hypothesis H_0: $m = \ell$ against H_1: $m = \ell + 1$. The test statistic is based on the difference between the optimal SSR associated with ℓ breaks and the optimal SSR associated with $\ell + 1$ breaks. Let

$(\hat{k}_1, \ldots, \hat{k}_\ell)$ be the estimator of the break point $(k_1^0, \ldots, k_\ell^0)$ under the null. Let $(\hat{k}_1^*, \ldots, \hat{k}_{\ell+1}^*)$ be the point at which the sum of squared residuals is minimized, when $\ell + 1$ break points are allowed.

The test statistic is defined as

$$\sup \mathrm{LR}_T(\ell + 1|\ell) = \frac{S_T(\hat{k}_1, \ldots, \hat{k}_\ell) - S_T(\hat{k}_1^*, \ldots, \hat{k}_{\ell+1}^*)}{\hat{\sigma}^2(\ell + 1)}, \tag{4}$$

where $\sigma^2(\ell + 1) = S_T(\hat{k}_1^*, \ldots, \hat{k}_{\ell+1}^*/T)$. When the errors are i.i.d. normal random variables, the above test is a likelihood ratio (LR) test. Because normality is not assumed, Eq. (4) may be considered a pseudo-likelihood ratio test. Using dynamic programming algorithm, the computation of Eq. (4) is straightforward and fast even for large T and large ℓ. In the Monte Carlo section below, we use Bai and Perron's (1994) program to compute (4)[4].

Theorem 1. Let Assumptions 1–4 hold. Then under H_0: $m = \ell$,

$$\sup \mathrm{LR}_T(\ell + 1|\ell) \xrightarrow{\mathrm{d}} \max\{\xi_1, \ldots, \xi_{\ell+1}\},$$

where $\xi_1, \ldots, \xi_{\ell+1}$ *are independent and have the representation*

$$\xi_i = \sup_{\eta_i \leq s \leq 1-\eta_i} \sum_{j=1}^{q} \frac{B_{i,j}(s)^2}{s(1-s)} \tag{5}$$

here $\eta_i = \pi/(\tau_i^0 - \tau_{i-1}^0)$, *and the* $B_{i,j}(\mu)$ *are independent and standard Brownian bridges on* $[0, 1]$.

Although the ξ_i's are independent, they do not necessarily have the same distribution because the η_i's may not be the same. Regardless of identical distribution, critical values are easy to obtain, as the limiting distribution has a known analytical density function.

Corollary 1. Under the assumptions of Theorem 1,

$$\lim_{T \to \infty} P(\sup \mathrm{LR}_T(\ell + 1|\ell) > c) = 1 - \prod_{i=1}^{\ell+1} [1 - G_i(c)],$$

[4] The computer program in GAUSS is available upon request.

where

$$G_i(c) = \frac{c^{q/2}\exp(-c/2)}{2^{q/2-1}\Gamma(q/2)}\left[\left(1-\frac{q}{c}\right)\log\frac{(1-\eta_i)}{\eta_i}+\frac{2}{c}+o(c^{-2})\right].$$

Given a size α, the corresponding critical value c can be easily computed from the above formula. For large c and small η_i the first term inside the brackets of $G_i(c)$ dominates, and thus the term $2/c$ may be ignored. In practice, the $\eta_i = \pi/(\tau_i^0 - \tau_{i-1}^0)$ is unknown, even though π is under the control of a researcher. It can be replaced, however, by its estimated value such that $\hat{\eta}_i = \pi/(\hat{\tau}_i - \hat{\tau}_{i-1})$ for $i = 1, \ldots, \ell$, where $\hat{\tau}_0 = 0$ and $\hat{\tau}_{\ell+1} = 1$. Under the null hypothesis, $\hat{\tau}_i$ converges to $\tau_i{}^0$ at rate T, yielding good estimates for η_i. The analytical expression is convenient for automation (computer programming). The expression $G_i(c)$ is derived in DeLong (1981).

Bai and Perron (1994) suggest an alternative (conditioning) procedure to test ℓ versus $\ell+1$ breaks. Their test is also based on the difference between the SSR associated with ℓ breaks and that associated with $\ell+1$ breaks. However, the $\ell+1$ break points are obtained not simultaneously but sequentially. More specifically, the $(\ell+1)$th break point is obtained conditional on the ℓ break points obtained previously, and the $(\ell+1)$th break point is picked from the $\ell+1$ subsamples separated by the break points $\hat{k}_1, \ldots, \hat{k}_\ell$. This additional break point is chosen in the subsample where the sum of squared residuals achieves the greatest reduction.[5] The limiting distribution of their test is the maximum of $\ell+1$ i.i.d. random variables of the form (5). The test proposed in this paper has a different limiting distribution from that of Bai and Perron's sequential test. The present test is conceptually simpler, and has a likelihoodratio interpretaion. Although the technical proof is more demanding than the sequential test of Bai and Perron (1994), the test is easy to compute and implement.

3.1. Consistency of the test

The test procedure proposed in this paper is consistent. That is, under the alternative hypothesis that $m \geqslant \ell+1$, then $\sup \mathrm{LR}_T(\ell+1|\ell) \to \infty$ with probability tending to 1. Therefore, when more than ℓ breaks exist, the null hypothesis of ℓ breaks will be rejected with probability tending to 1. We can

[5] The study of this conditioning procedure rather than LR was partly due to our more primitive understanding of the LR procedure back then. Particularly, it was less well understood in terms of the form of LR's limiting distribution, as well as the technical apparatus needed in justifying its limiting distribution. The LR's deeper understanding came only recently as reflected in this paper. However, Bai and Perron's procedure does have an advantage that no simultaneous estimation of the break points is necessary when we use it repeatedly as ℓ increases.

actually show that $\sup \mathrm{LR}_T(\ell + 1|\ell) = \mathrm{O}_p(T)$ under the alternative hypothesis. To see this, suppose there are $\ell + 1$ breaks. When only ℓ breaks are allowed in estimation, at least one break point cannot be consistently estimated. Lemma 2 of Bai and Perron (1994) then implies

$$S_T(\hat{k}_1, \ldots, \hat{k}_1) - \sum_{t=1}^{T} u_t^2 = \mathrm{O}_p(T) \to \infty .$$

On the other hand, when $\ell + 1$ breaks are allowed in estimation,

$$\sum_{t=1}^{T} u_t^2 - S_T(\hat{k}_1^*, \ldots, \hat{k}_{\ell+1}^*) = \mathrm{O}_p(1).$$

Thus the numerator of the test statistic is $\mathrm{O}_p(T)$ and the denominator is $\mathrm{O}_p(1)$. So the test statistic is $\mathrm{O}_p(T)$ under the alternative hypothesis. This conclusion remains true when more than $\ell + 1$ breaks exist under the alternative. The details are omitted.

3.2. Estimating the number of breaks

One advantage of our test is its capability of identifying the number of breaks in the data. The procedure is described in Bai and Perron (1994), although they use a different test. Here we give a brief summary. One begins with a test of no-break versus a single break.[6] If the hypothesis is rejected, one proceeds to test the null of a single break versus two breaks, and so forth. This process is repeated until the test $\sup \mathrm{LR}_T(\ell + 1|\ell)$ fails to reject the null hypothesis of no additional breaks. The estimated number of breaks is equal to the number of rejections.[7]

Let $\hat{m}$ denote this number, and let m_0 denote the true number of breaks. We have the following result.

Theorem 2. Let Assumptions 1–4 hold. If the size α_T converges to zero slowly enough (for the test based on $\sup \mathrm{LR}_T(\ell + 1|\ell)$ *to remain consistent), then*

$$P(\hat{m} = m_0) \to 1.$$

[6] One may start with H_0: $\ell = \ell_0 > 0$ if at least ℓ_0 breaks are known to exist.

[7] Alternatively, this procedure may be used in a reverse order. That is, instead of increasing the value of ℓ, we decrease its value (starting with a reasonable initial value) until the null of no-change is rejected. The number of break points is equal to the number specified by the alternative hypothesis of the first rejection. The theoretical property of this reversed procedure, however, remains to be studied.

The proof of this theorem is identical to that of Proposition 8 in Bai and Perron (1998) and is thus omitted.

Let α be the size of the test and c be the corresponding critical value calculated from the limiting distribution in Theorem 1. Because $1-\prod_{i=1}^{\ell+1}[1-G_i(c)] \leqslant \sum_{i=1}^{\ell+1} G_i(c) \leqslant Kc^{q/2}\exp(-c/2)\exp(-c/3)$ for all large c, we have $\alpha \leqslant \exp(-c/3)$. This implies $c \leqslant -3\log(\alpha)$ (for all small α or equivalently large c). Suppose α_T is chosen such that $\alpha_T = 1/T$, then $c_T \leqslant 3\log T$. Under the alternative hypothesis, we know $\sup \mathrm{LR}_T(\ell+1|\ell)$ is of order $\mathrm{O_p}(T)$. Consequently, we will reject the null hypothesis with probability tending to 1. This implies that one will not underestimate the true number of break points for large samples. The probability of overestimating the number of breaks is no more than α_T (see Bai and Perron, 1998). So if $\alpha_T \to 0$ slowly (say no more quickly than $1/T$), we can consistently estimate the number of breaks. In practice, T is fixed, and so is α_T. Simulation shows that $\alpha = 0.05$ works satisfactorily.

Recently, Andrews et al. (1996) consider an optimal test for regression models under the assumption of normality. The null hypothesis is no-change, and the alternative hypothesis allows for multiple changes. In contrast, the present test allows for structural changes under the null.[8] It is this feature that enables us to determine the number of breaks via hypothesis testing. In addition, critical values of the Andrews–Lee–Ploberger test are not available except for a single change, and do not appear to be straightforward to obtain even for large samples. In comparison, the critical values of the LR test of this paper can be obtained analytically. Finally, we point out that a likelihood-ratio-type test for the null of no-change ($m = 0$) against alternative multiple changes ($m = \ell$) is derived in Bai and Perron (1998) in the absence of trending regressors. However, rejection of the null of no-change does not provide information as to the number of breaks in the data.

4. Trending regressors

In the previous sections, regime-wise asymptotic stationarity is assumed, ruling out trending regressors. We now derive the corresponding results in the presence of such regressors. Consider the following model with polynomial trends:

$$y_t = w_t'\gamma_i + z_t'\delta_i + u_t, \quad k_{i-1}^0 + 1 \leqslant t \leqslant k_i^0, (i = 1, 2, \ldots, m+1) \tag{6}$$

where $w_t' = (1, t, \ldots, t^p)$, $k_0^0 = 0$, and $k_{m+1}^0 = T$. The regressor z_t $(q \times 1)$ satisfies the assumptions stated in the previous section. It is assumed that $(\gamma_i, \delta_i) \neq (\gamma_{i+1}, \delta_{i+1})$.

[8] This paper does not address the issue of optimality. For a general treatment of this issue, see Sowell (1996).

Test statistic for testing H_0: $m = \ell$ versus H_1: $m = \ell + 1$ has the same format as before: it is constructed using the difference between the optimal SSR associated with ℓ breaks and that associated with $\ell + 1$ breaks. For $s \in [0, 1]$, define the matrix

$$A(s) = \begin{pmatrix} s & \frac{1}{2}s^2 & \cdots & \frac{1}{p+1}s^{p+1} \\ \frac{1}{2}s^2 & \frac{1}{3}s^3 & \cdots & \frac{1}{p+2}s^{p+2} \\ \vdots & \vdots & & \vdots \\ \frac{1}{p+1}s^{p+1} & \frac{1}{p+2}s^{p+2} & \cdots & \frac{1}{2p+1}s^{p+1} \end{pmatrix},$$

which is the (appropriately) normalized limit of $\sum_{t=1}^{[Ts]} w_t w_t'$.

Let $\lambda_1(s), \ldots, \lambda_{p+1}(s)$ denote the eigenvalues of $A(1)^{-1/2}A(s)A(1)^{-1/2}$. (They are also the eigenvalues of $A(s)\ A(1)^{-1}$.) These eigenvalues are continuous and increasing functions of s with $\lambda_i(0) = 0$, and $\lambda_i(1) = 1$. The limiting distribution of the test statistic is characterized in the next theorem.

Theorem 3. Let Assumptions 1–4 hold. Under H_0 of ℓ breaks, the test statistic satisfies

$$\sup \mathrm{LR}_T(\ell + 1|\ell) \xrightarrow{\mathrm{d}} \max\{\psi_1, \ldots, \psi_{\ell+1}\}$$

where

$$\psi_i = \sup_{\eta_i \leqslant s \leqslant (1-\eta_i)} \left[\sum_{j=1}^{p+1} \frac{B_{ij}^2(\lambda_j(s))}{\lambda_j(s)(1-\lambda_j(s))} + \sum_{j=p+2}^{p+q+1} \frac{B_{ij}^2(s)}{s(1-s)} \right],$$

the $B_{ij}(\cdot)$ are standard Brownian bridges on $[0, 1]$, and the η_i are defined in Theorem 1.

Corollary 2. Under the assumptions of Theorem 1,

$$\mathrm{P}(\sup \mathrm{LR}_T(\ell + 1|\ell) > c) = 1 - \prod_{i=1}^{\ell+1} [1 - H_i(c)], \tag{7}$$

where

$$H_i(c) \approx \frac{c^{r/2} \exp(-c/2)}{2^{r/2-1}\Gamma(r/2)} \left(\frac{1}{r} - \frac{1}{c}\right) \times \left[\frac{1}{2} \log \frac{\det A(1-\eta_i)\det(A(1) - A(\eta_i))}{\det A(\eta_i)\det(A(1) - A(1-\eta_i))} q \log \frac{(1-\eta_i)}{\eta_i} \right],$$

with $r = p + q + 1$ and det B as the determinant of B.

Theorem 3 and its corollary hold for general trending regressors beyond polynomial ones. All that is needed is that the limiting matrix $A(s)$ be positive definite and increasing. That is, $A(u) > A(v)$ for $u > v$. However, under the specific structure of polynomial trends, $H_i(c)$ can be simplified. It can be shown that

$$\det(A(1) - A(1-s)) = \det A(s) = s^{(p+1)^2} \det A(1).$$

This implies that

$$\frac{1}{2}\log\frac{\det A(1-\eta_i)\det(A(1)-A(\eta_i))}{\det A(\eta_i)\det(A(1)-A(1-\eta_i))} = (p+1)^2\log\frac{1-\eta_i}{\eta_i}.$$

Therefore,

$$H_i(c) \approx \frac{c^r/2\exp(-c/2)}{2^{r/2-1}\Gamma(r/2)}\left(\frac{1}{r}-\frac{1}{c}\right)\left[[(p+1)^2+q]\log\frac{(1-\eta_i)}{\eta_i}\right]. \tag{8}$$

Because $(1/r - 1/c)[(p+1)^2+q] = (1-r/c)[(p+1)^2+q]/r$, the value of $H_i(c)$ is proportional to $G_i(c)$ of Corollary 1 corresponding to $r = p+1+q$ stationary regressors. The proportionality factor is $[(p+1)^2+q]/r$. This implies the following interesting fact. When testing for no-change versus a single change ($\ell = 0$), the asymptotic critical value of a size α test for model (6) is equal to that of a size $\alpha(p+1+q)/((p+1)^2+q)$ test with $p+1+q$ stationary regressors. The latter is tabulated in many papers, see Andrews (1993).

5. Some numerical results

To illustate the procedure and to indicate the finite sample performance of the LR test, we report results of a limited set of sampling experiments.

The sampling experiment considers three different sets of regressors, simple linear regression, autoregression, and linear trend. For each set of regressors, a sample of $T = 150$ observations is generated from a model with 2 breaks (3 regimes). The first break is set at $k_1^0 = 50$ and the second at $k_2^0 = 100$. We may write the data generating process as

$$y_t = a_i + b_i z_t + u_t, \quad k_{i-1}^0 + 1 \leqslant t \leqslant k_i^0, \quad i = 1, 2, 3,$$

where (a_i, b_i) is regime i's regression parameter, and $k_0^0 = 0$, and $k_3^0 = T$. The three sets of regressors are (I): z_t i.i.d. normal N(1,1); (II) $z_t = y_{t-1}$, and (III): $z_t = t$. The disturbances are i.i.d. standard normal for all cases. Let $\delta = [(a_1,b_1), (a_2,b_2), (a_3,b_3)]$.

The design coefficients are

(I) (regression): $\delta = [(1.0, 1.0), (1.5, 1.5), (2.0, 2.0)]$
(II) (autoregression): $\delta = [(10.0, 0.5), (10.0, 0.4), (10.0, 0.5)]$,
(III) (linear trend): $\delta = [(1.0, 1.0), (1.1, 1.1), (1.2, 1.2)]$.

For design (II), the intercept is not subject to shifts. In addition, the autoregressive slope coefficient has a small shift in magnitude. But because of the large value of the intercept, the expected value of $z_t^2 = y_{t-1}^2$ is large. This implies that the break points can be accurately estimated because the variance of the break-point estimator is inversely related to $(b_i - b_{i+1})^2 E z_t^2$ (assume no intercept shift). For all designs, the model is estimated by imposing the restriction that each regime has at least 5 observations. This defines π or Λ_π in Eq. (3) implicitly.

Table 1 reports the percentage rejections from 5000 repetitions for testing H_0: $m = \ell$ against H_1: $m = \ell + 1$ for $\ell = 0, 1, 2$. Two nominal sizes are considered: $\alpha = 0.05$ and $\alpha = 0.10$. The percentages in the last two columns can be viewed as actual sizes (after dividing by 100) because the null hypothesis is true. The percentages in the first four columns can be viewed as powers (again after dividing by 100). For the linear regression and autoregressive models, the actual sizes correspond well with the nominal sizes. But for the linear trend model, the actual sizes are somewhat below the nominal ones.

Table 2 gives the selected number of breaks via hypothesis testing with size $\alpha = 0.05$. The test sup $\mathrm{LR}(\ell + 1|\ell)$ is repeatedly performed by increasing the value of ℓ, until the null hypothesis is accepted. For each sample, the estimated number of breaks, $\hat{m}$, is equal to the number of rejections. Table 2 shows the distribution of $\hat{m}$ out of 5000 simulated samples. For all three models, over 90% of the time, the method leads to correct identification of the number of breaks.

We now consider the relative performance of the LR test with Bai and Perron's Conditional Test sup $F_T(\ell + 1|\ell)$. In the latter test, the $(\ell + 1)$th break point is estimated assuming the first ℓ break points are given (estimated in previous steps). The limiting distribution of the conditional test is given in

Table 1
Percentage rejections of sup $\mathrm{LR}(\ell + 1|\ell)$ (5000 repetitions)

H_0	$m = 0$		$m = 1$		$m = 2$	
H_1	$m = 1$		$m = 2$		$m = 3$	
Size	0.05	0.10	0.05	0.10	0.05	0.10
Model						
(I)	100.0	100.0	95.7	97.5	4.8	10.1
(II)	100.0	100.0	99.9	99.9	5.4	10.7
(III)	100.0	100.0	100.0	100.0	3.3	6.7

Table 2
Distribution of the estimated number of breaks based on sup $LR(\ell + 1|\ell)$ (5000 repetitions)

$\hat{m}$	0	1	2	3	$\geqslant 4$
Model					
(I)	0	233	4524	240	3
(II)	0	3	4726	267	4
(III)	0	0	4835	163	2

Proposition 7 of Bai and Perron (1998). To assess the relative performance of the conditional and LR procedures, Monte Carlo experiments are conducted using the same data generating process described earlier. Tables 3 and 4 report the results associated with the conditional test. These tables show that the conditional procedure also does a satisfactory job. In comparion with Tables 1 and 2, the sup LR test shows an improved preformance.

Taken together, the sup LR test has reasonable size and power properties. Its performance in determining the number of breaks is quite satisfactory. An added attraction lies in its feasiblity and ease of use.

6. Discussion

6.1. Partial structural change

Thus far, we allow all regression coefficients to vary. That is, all coefficients are re-estimated whenever a break occurs. Such a setup is called full structural change. If it is known that some of the coefficients do not change, the constraint of no change should be imposed. This will increase the efficiency of the estimated regression parameters as well as the power of hypothesis testing. This setup is called partial structural change.

A general partial-change model is described by

$$y_t = x_t'\beta + z_t'\delta_i + u_t, \quad k_{i-1}^0 + 1 \leqslant t \leqslant k_i^0,$$

for $i = 1, \ldots, m + 1$. The coefficient β is constant for the whole sample. Bai and Perron (1998) develop a fast-computing algorithm for partial change models. Under some general conditions for $\{x_t, z_t\}$, as described in Bai and Perron (1998), all preceding results hold. In particular, even though x_t contains trending regressors, Theorem 1 holds as long as z_t is regime-wise asymptotically stationary. That is, the presence of regressor x_t does not alter the theoretical results in Theorems 1–3. These claims can be proved rigorously at a greater technical expense.

Table 3
Percentage rejections of the conditional test (5000 repetitions)

H_0	$m=0$		$m=1$		$m=2$	
H_1	$m=1$		$m=2$		$m=3$	
Size	0.05	0.10	0.05	0.10	0.05	0.10
Model						
(I)	100.0	100.0	89.9	94.7	4.3	8.1
(II)	100.0	100.0	97.5	98.7	5.4	10.1
(III)	100.0	100.0	100.0	100.0	6.4	10.9

Table 4
Distribution of the estimated number of breaks based on the conditional test (5000 repetitions)

$\hat{m}$	0	1	2	3	$\geqslant 4$
Model					
(I)	0	504	4211	270	15
(II)	0	124	4133	678	65
(III)	0	0	4679	310	11

It is noted that the tests of this paper are constructed using the full sample, which can take advantage of partial changes. This is important particularly for multiple changes because the savings in the number of degrees of freedom can be substantial. Sample-splitting methods frequently used in applied work are unable to make use of partial changes.

6.2. *Multiple-threshold time series models*

The procedure proposed here may be extended to threshold time series models (Tong, 1990). One interesting topic is to identify the number of thresholds. To date, the determination of the number of thresholds in a model is based on a less formal approach. This is usually accomplished by plotting and identifying the turning points in the cusum of recursive residuals. Hansen (1996) recently suggests a sample-splitting method to study multiple-threshold models. By arranging the data in an appropriate way, the threshold time series model can be turned into a structural-change problem, at least computationally. Thus an LR-type test statistic for testing ℓ thresholds versus $\ell+1$ thresholds can be easily constructed. The underlying theoretical distribution, however, is different from the one given here. Chan (1991) considers the null of linearity against a single threshold. In view of the result of this paper and that of Chan,

conjectures can be made about the limiting distribution of the LR type test for multiple thresholds. However, the underlying theoretical pinning needs a more closer examination.

Acknowledgements

This paper has benefited from seminar presentations at the MIT/Harvard Econometrics workshop as well as the 1996 ASSA meetings in New Orleans. Partial financial support is acknowledged from the National Science Foundation under Grants SBR-9414083 and SBR-9709508.

Appendix: Proofs

Under the null hypothesis of ℓ changes, $(\hat{k}_1, \ldots, \hat{k}_\ell)$ is T consistent for $(k_1^0, \ldots, k_\ell^0)$ (see Bai and Perron, 1998). When there are actually ℓ breaks but $\ell+1$ are allowed in the estimation, we shall show that ℓ of them will be T consistent. More specifically, let $(\hat{k}_1^*, \ldots, \hat{k}_{\ell+1}^*)$ denote the estimated $\ell+1$ break points. There exists a subset $\{i_1, i_2, \ldots, i_\ell\}$ of $\{1, 2, \ldots, \ell+1\}$ such that $|\hat{k}_{i_v}^* - k_v^0| = \mathrm{Op}(1)$ $(v = 1, \ldots, \ell)$.

Lemma A.1. Let Assumptions A1–A4 hold. Under the null hypothesis of ℓ breaks, there are ℓ and exactly ℓ of $\{\hat{k}_1^*, \ldots, \hat{k}^*_{\ell+1}\}$ that will be T consistent for $(k_1^0, \ldots, k_\ell^0)$. That is, $|\hat{k}_{i_v}^* - k_v^0| = \mathrm{Op}(1)$ for some $\{i_1, \ldots, i_\ell\} \subset \{1,2, \ldots, \ell+1\}$.

Proof. We first prove consistency. That is, there exists a subset $\{i_1, \ldots, i_\ell\}$ of $\{1, 2, \ldots, \ell+1\}$ such that for every $\varepsilon > 0$, we have $\mathrm{P}(|\hat{k}_{i_v}^* - k_v^0| > \varepsilon T) < \varepsilon (v = 1, \ldots, \ell)$ for all large T. Suppose such a subset does not exist, it then implies that there exists a break point which cannot be consistently estimated. Then, from the result of Bai and Perron (1998, Lemma 2)

$$\sum_{t=1}^{T} u_t^2 - S_T(\hat{k}_1^*, \ldots, \hat{k}_{\ell+1}^*) = -\mathrm{O_p}(T) \to -\infty .$$

This contradicts with the least squares principle that

$$\sum_{t=1}^{T} u_t^2 - S_T(\hat{k}_1^*, \ldots, \hat{k}_{\ell+1}^*) \geqslant \sum_{t=1}^{T} u_t^2 - S_T(\hat{k}_1, \ldots, \hat{k}_\ell) \geqslant 0.$$

Next suppose, without loss of generality, that $\hat{k}_1^*, \ldots, \hat{k}_{i-1}^*, \hat{k}_{i+1}^*, \ldots, \hat{k}_{\ell+1}^*$ are consistent. This implies that $\hat{k}_{i-1}^*$ is close to k_{i-1}^0, and $\hat{k}_{i+1}^*$ is close to k_i^0. By the

definition of Λ_π, $\hat{k}_i^*$ cannot be close to either k_{i-1}^0 or k_i^0. That is, there exists $\varepsilon_0 > 0$ such that with large probability $\hat{k}_i^* - k_{i-1}^0 \geqslant \varepsilon_0 T$, and $k_i^0 - \hat{k}_i^* \geqslant \varepsilon_0 T$. This implies that the subsample $[1, \hat{k}_i^*]$ contains $i-1$ true break points. If we use this subsample to estimate the true break points, the solution coincides with[9] $(\hat{k}_1^*, \ldots, \hat{k}_{i-1}^*)$. Since there are $i-1$ true break points, and we allow $i-1$ breaks in the estimation, by the result of Bai and Perron (1998), the resulting estimator $(\hat{k}_1^*, \ldots, \hat{k}_{i-1}^*)$ is T-consistent for $(k_1^0, \ldots, k_{i-1}^0)$. A similar argument applies to the subsample $[\hat{k}_i^*, T]$. And so $\hat{k}_{i+1}^*, \ldots, \hat{k}_{\ell+1}^*$ are T-consistent for $k_i^0, \ldots, k_\ell^0$.

Lemma A.2. Let Assumptions 1–4 and H_0 hold. Let $j_1, \ldots, j_\ell$ be integers (may be negative). For every $M < \infty$, uniformly over $|j_i| \leqslant M$ $(i = 1, \ldots, \ell)$ we have

$$S_T(k_1^0 + j_1, \ldots, k_\ell^0 + j_\ell) - S_T(k_1^0, \ldots, k_\ell^0) - \sum_{i=1}^{\ell} W_i(j_i) = o_p(1),$$

where for $i = 1, \ldots, \ell$, $W_i(0) = 0$ and

$$W_i(j) = \begin{cases} (\delta_{i+1} - \delta_i)' \sum_{k_i^0 - j}^{k_i^0} z_t z_t' (\delta_{i+1} - \delta_i) - 2(\delta_{i+1} - \delta_i)' \sum_{k_i^0 - j}^{k_i^0} z_t u_t, & j = -1, -2, \ldots \\ (\delta_{i+1} - \delta_i)' \sum_{k_i^0 + 1}^{k_i^0 + j} z_t z_t' (\delta_{i+1} - \delta_i) + 2(\delta_{i+1} - \delta_i)' \sum_{k_i^0 + 1}^{k_i^0 + j} z_t u_t, & j = 1, 2, \ldots \end{cases}$$

Proof. To be concrete, we shall analyze the case of $j_i > 0$ for $i = 1, \ldots, \ell$. Other cases can be analyzed similarly. Let $(\hat{\delta}_1, \ldots, \hat{\delta}_{\ell+1})$ be the estimated regression parameters corresponding to the partition $(k_1^0 + j_1, \ldots, k_\ell^0 + j_\ell)$. The resulting SSR for the total sample can be written as the summation of the SSRs of $2\ell + 1$ segments, with each segment involving observations from a single true regime only. More specifically,

$$S_T(k_1^0 + j_1, \ldots, k_\ell^0 + j_\ell) = \sum_{i=1}^{\ell+1} \sum_{k_{i-1}^0 + j_{i-1} + 1}^{k_i^0} [(y_t - z_t' \hat{\delta}_i)^2 + \sum_{i=1}^{\ell} \sum_{k_i^0 + 1}^{k_i^0 + j_i} (y_t - z_t' \hat{\delta}_i)^2, \tag{A.1}$$

with $k_0^0 = 0$ and $j_0 = 0$. Similar to the above, we can express $S_T(k_1^0, \ldots, k_\ell^0)$ as the summation of SSR's of the same $(2\ell + 1)$ segments. Let $(\hat{\delta}_1^0, \ldots, \hat{\delta}_{\ell+1}^0)$ be the

[9] Notice the fact that,

$$\min_{k_1, \ldots, k_{i-1}, k_{i+1}, \ldots, k_{\ell+1}} S_T(k_1, \ldots, k_{i-1}, \hat{k}_i^*, k_{i+1}, \ldots, k_{\ell+1}) \equiv S_T(\hat{k}_1^*, \ldots, \hat{k}^*_{\ell+1}).$$

estimated regression parameters based on the partition $(k_1^0, \ldots, k_\ell^0)$. Then $S_T(\hat{k}_1^0, \ldots, \hat{k}_\ell^0)$ is given by Eq. (A.1) with the first $\hat{\delta}_i$ replaced by $\hat{\delta}_i^0$ and the second replaced by $\hat{\delta}_{i+1}^0$ (not $\hat{\delta}_i^0$ because of no misspecification in regime spells). Thus,

$$S_T(k_1^0 + j_1, \ldots, k_\ell^0 + j_\ell) - S_T(k_1^0, \ldots, k_\ell^0) = \sum_{i=1}^{\ell+1} \sum_{k_{i-1}^0 + j_{i-1} + 1}^{k_i^0}$$

$$[(y_t - z_t'\hat{\delta}_i)^2 - (y_t - z_t'\hat{\delta}_i^0)^2] + \sum_{i=1}^{\ell} \sum_{k_i^0 + 1}^{k_i^0 + j_i} [(y_t - z_t'\hat{\delta}_i)^2 - (y_t - z_t'\hat{\delta}_{i+1}^0)^2]. \quad \text{(A.2)}$$

Because j_i is bounded for each i, all estimated regression parameters are root T consistent for the true regression parameters. This, together with the boundedness of j_i, allows us to write the second term of the right hand side of Eq. (A.2) as

$$\sum_{i=1}^{\ell} \sum_{k_i^0 + 1}^{k_i^0 + j_i} \{[u_t + z_t'(\delta_{i+1} - \delta_i)]^2 - u_t^2\} + o_p(1) = \sum_{i=1}^{\ell} W_i(j_i) + o_p(1).$$

This follows from $y_t - z_t'\hat{\delta}_i = u_t - z_t'(\hat{\delta}_i - \delta_{i+1}) = u_t - z_t'(\delta_i - \delta_{i+1}) - z_t'(\hat{\delta}_i - \delta_i) = u_t - z_t'(\delta_i - \delta_{i+1}) + z_t' O_p(T^{-1/2})$. Similarly, $y_t - z_t'\hat{\delta}_{i+1}^0 = u_t - z_t' O_p(T^{-1/2})$.

Next, we shall argue that the first term on the right-hand side of Eq. (A.2) converges to zero in probability. The said term is equal to

$$-2 \sum_{i=1}^{\ell} \left[(\hat{\delta}_i - \hat{\delta}_i^0)' \sum_{k_{i-1}^0 + j_{i-1} + 1}^{k_i^0} z_t u_t \right] + \sum_{i=1}^{\ell} \left[[(\hat{\delta}_i - \delta_i) + (\hat{\delta}_i^0 - \delta_i)]' \right.$$

$$\left. \sum_{k_{i-1} + j_{i-1} + 1}^{k_i^0} (z_t z_t')(\hat{\delta}_i - \hat{\delta}_i^0) \right].$$

It is not difficult to prove that $\hat{\delta}_i - \hat{\delta}_i^0 = O_p(1/T)$ for all i (follows from the boundedness of j_i). Using this fact we see that each of the two expressions above is $O_p(T^{-1/2})$. □

The next lemma states that when an additional break is allowed in estimation, Lemma 2 should be modified by adding an extra term, which will determine the limiting distribution of the test statistic. Since exactly ℓ estimated breaks will be consistent by Lemma 1, one of them will not be consistent. The location of this inconsistent break point can be in any of the $(\ell + 1)$-segments separated by the ℓ-consistent ones. The following lemma gives the asymptotic behavior of the SSR when an extra break point (denoted by h) is in the ith segment

$(i = 1, 2, \ldots, \ell + 1)$. In the remainder of this paper, we use $D(k_1, k_2)$ to denote the sum of squared residuals for the subsample $[k_1 + 1, k_2]$. That is, $D(k_1, k_2) = \min_\delta \sum_{t=k_1+1}^{k_2} (y_t - z_t'\delta)^2$.

Lemma A.3. Let Assumptions 1–4 hold. Under the assumption of ℓ breaks, we have

$$S_T(h, k_1^0 + j_1, \ldots, k_\ell^0 + j_\ell) - S_T(k_1^0, \ldots, k_\ell^0) - \sum_{s=1}^{\ell} W_s(j_s) + \Gamma_{1,T}(h) = o_p(1),$$

$$S_T(k_1^0 + j_1, h, k_2^0 + j_2, \ldots, k_\ell^0 + j_\ell) - S_T(k_1^0, \ldots, k_\ell^0) - \sum_{s=1}^{\ell} W_s(j_s) + \Gamma_{2,T}(h) = o_p(1),$$

$$S_T(k_1^0 + j_1, \ldots, k_\ell^0 + j_\ell, h) - S_T(k_1^0, \ldots, k_\ell^0) - \sum_{s=1}^{\ell} W_s(j_s) + \Gamma_{\ell+1,T}(h) = o_p(1),$$

where

$$\Gamma_{i,T}(h) = D(k_{i-1}^0, k_i^0) - D(k_{i-1}^0, h) - D(h, k_i^0) \tag{A.3}$$

and the $o_p(1)$ is uniform in $|j_i| \leqslant M$ and in $h \in V_i$ for equation i, with

$$V_i = \{h: h \in [k_{i-1}^0 + j_{i-1} + \pi T, k_i^0 + j_i - \pi T]\}. \tag{A.4}$$

So for $h \in V_i$, we have $(k_1^0 + j_1, \ldots, k_{i-1}^0 + j_{i-1}, h, k_i^0 + j_i, \ldots, k_\ell^0 + j_\ell) \in \Lambda_\pi$.

We note that $\Gamma_{i,T}(h)$ is the difference between the restricted and unrestricted sums of squared residuals (allowing a break at h) for the subsample $[k_{i-1}^0 + 1, k_i^0]$. From the standard result, if $j_{i-1} = 0$ and $j_i = 0$ in V_i, then

$$\sup_{h \in V_i} \Gamma_{i,T}(h)/\sigma^2 \xrightarrow{d} \xi_i, \tag{A.5}$$

where ξ_i is defined in Eq. (5). When j_{i-1} and j_i are non-zero but bounded, the limiting distribution is the same.

Lemma A.3 can be put in a more compact form, for $i = 1, \ldots, \ell + 1$

$$S_T(k_1^0 + j_1, \ldots, k_{i-1}^0 + j_{i-1}, h, k_i^0 + j_i, \ldots, k_l^0 + j_\ell) - S_T(k_1^0, \ldots, k_\ell^0) = \sum_{s=1}^{\ell} W_s(j_s) - \Gamma_{i,T}(h) + o_p(1), \tag{A.6}$$

where for $i = 1$, the first term on the left-hand side should be understood as $S_T(h, k_1^0 + j_1, \ldots, k_\ell^0 + j_\ell)$, and for $i = \ell + 1$, as $S_T(k_1^0 + j_1, \ldots, k_{\ell+1}^0 + j_{\ell+1}, h)$.

Proof of Lemma. A.3. Again, consider the case of $j_s \geqslant 0$ for all s. We shall prove the second equation of Lemma A.3. The rest are similar (the first and last are actually simpler). Note that $S_T(k_1^0 + j_1, h, k_2^0 + j_2, \ldots, k_\ell^0 + j_\ell) = D(0, k_1^0 + j_1) + D(k_1^0 + j_1, h) + D(h, k_2^0 + j_2) + \sum_{i=2}^{\ell} D(k_i^0 + j_i, k_{i+1}^0)$. By adding and subtracting $D(k_1^0 + j_1, k_2^0 + j_2)$, we have

$$\begin{aligned} S_T(k_1^0 + j_1, h, k_2^0 + j_2, \ldots, k_\ell^0 + j_\ell) &= D(k_1^0 + j_1, h) + D(h, k_2^0 + j_2) \\ &\quad - D(k_1^0 + j_1, k_2^0 + j_2) \\ &\quad + S_T(k_1^0 + j_1, \ldots, k_\ell^0 + j_\ell). \end{aligned}$$

The last term does not depend on h. By Lemma A.2, all we need is

$$\begin{aligned} &D(k_1^0 + j_1, h) + D(h, k_2^0 + j_2) - D(k_1^0 + j_1, k_2^0 + j_2) \\ &\quad = D(k_1^0, h) + D(h, k_2^0) - D(k_1^0, k_2^0) + o_p(1), \end{aligned} \tag{A.7}$$

where $o_p(1)$ is uniform in $h \in V_2$. This result will imply that the test statistic based on the sample $[k_1^0 + j_1, k_2^0 + j_2]$ will be asymptotically equivalent to the test based on the sample $[k_1^0, k_2^0]$. This should be obvious given the boundedness of j_1 and j_2. Below is the formal proof. Note that Eq. (A.7) is implied by the following two equations:

$$D(k_1^0 + j_1, h) - D(k_1^0 + j_1, k_2^0 + j_2) = D(k_1^0, h) - D(k_1^0, k_2^0 + j_2) + o_p(1), \tag{A.8}$$

$$D(h, k_2^0 + j_2) - D(k_1^0, k_2^0 + j_2) = D(h, k_2^0) - D(k_1^0, k_2^0) + o_p(1) \tag{A.9}$$

because adding Eqs. (A.8) and (A.9) yields Eq. (A.7).

Consider Eq. (A.8). Let $\hat{\delta}_2^\dagger$ and $\hat{\delta}_2^\diamond$ be estimators of δ_2 using samples $[k_1^0 + 1, h]$ and $[k_1^0 + 1, k_2^0 + j_2]$, respectively. Then

$$D(k_1^0, h) = \sum_{t=k_1^0+1}^{k_1^0+j_1} (y_t - z_t'\hat{\delta}_2^\dagger)^2 + \sum_{k_1^0+j_1+1}^{h} (y_t - z_t'\hat{\delta}_2^\dagger)^2 \tag{A.10}$$

and

$$D(k_1^0, k_2^0 + j_2) = \sum_{t=k_1^0+1}^{k_1^0+j_1} (y_t - z_t'\hat{\delta}_2^{\diamond})^2 + \sum_{k_1^0+j_1+1}^{k_2^0+j_2} (y_t - z_t'\hat{\delta}_2^{\diamond})^2. \tag{A.11}$$

Since both $\hat{\delta}_2^{\dagger}$ and $\hat{\delta}_2^{\diamond}$ are consistent for δ_2, and because j_1 is bounded, the difference between the two first terms of Eqs. (A.10) and (A.11) is $o_p(1)$, and therefore,

$$D(k_1^0, h) - D(k_1^0, k_2^0 + j_2) = \sum_{k_1^0+j_1+1}^{h} (y_t - z_t'\hat{\delta}_2^{\dagger})^2 - \sum_{k_1^0+j_1+1}^{k_2^0+j_2} (y_t - z_t'\hat{\delta}_2^{\diamond})^2 + o_p(1). \tag{A.12}$$

Let $\hat{\delta}_2^{\ddagger}$ be the estimator of δ_2 based on the sample $[k_1^0 + j_1 + 1, h]$. Then $D(k_1^0 + j_1, h) = \sum_{k^0{}_1+j_1+1}^{h}(y_t - z_t'\hat{\delta}_2^{\ddagger})^2$. Because j_1 is bounded, and because $h - k_1^0 \geqslant cT$ for some $c > 0$, $\hat{\delta}_2^{\dagger} - \hat{\delta}_2^{\ddagger} = O_p(T^{-1})$. This implies that the second term of Eq. (A.10) is $D(k_1^0 + j_1, h) + o_p$. [See the proof of the first expression of Eq. (4).] Similarly, the second term of Eq. (A.11) is $D(k_1^0 + j_1, k_2^0 + j_2) + o_p$. In view of Eq. (A.12), we obtain Eq. (A.8). The proof of Eq. (A.9) is similar. □

Proof of Theorem 1. If $|\hat{k}_i - k_i^0| \leqslant M$ for all i, then $S_T(\hat{k}_1, \ldots, \hat{k}_\ell) = \min_{j_1, \ldots, j_\ell} S_T(k_1^0 + j_1, \ldots, k_\ell^0 + j_\ell)$, here and in what follows, the minimization with respect to $j_1, \ldots, j_\ell$ is taken over $|j_i| \leqslant M$, $\forall i$. By Lemma A.2,

$$\min_{j_1, \ldots, j_\ell} S_T(k_1^0 + j_1, \ldots, k_\ell^0 + j_\ell) - S_T(k_1^0, \ldots, k_\ell^0) = \sum_{i=1}^{\ell} \min_{|j| \leqslant M} W_i(j) + o_p(1). \tag{A.13}$$

If $|\hat{k}_{i_v}^* - k_v^0| \leqslant M$ $(v = 1, \ldots, \ell)$, then

$$S_T(\hat{k}_1^*, \ldots, \hat{k}_{\ell+1}^*) = \min_{1 \leqslant i \leqslant \ell+1} \min_{j_1, \ldots, j_\ell; h \in V i} S_T(k_1^0 + j_1, \ldots, k_{i-1}^0 + j_{i-1}, h, k_i^0 + j_i, \ldots, k_\ell^0 + j_\ell),$$

where V_i is defined in Eq. (A.4). Note that for $h \in V_i$, we have $(k_1^0 + j_1, \ldots, k_{i-1}^0 + j_{i-1}, h, k_i^0 + j_i, \ldots, k_\ell^0 + j_\ell) \in \Lambda_\pi$. The above says that the

location of the inconsistent break point will be in the segment where the sum of squared residuals is reduced the most.

Lemma A.3 is equivalent to

$$\min_{1\leqslant i\leqslant \ell+1}\ \min_{j_1,\ldots,\,j_\ell;h\in V_i} S_T(k_1^0+j_1,\ldots,k_i^0+j_i,h,k_{i+1}^0+j_{i+1},\ldots,k_\ell^0+j_\ell)$$
$$-S_T(k_1^0,\ldots,k_\ell^0)=\min_{1\leqslant i\leqslant \ell+1}\ \min_{j_1,\ldots,\,j_\ell;h\in V_i}\left\{\sum_{s=1}^{\ell} W_s(j_s)-\Gamma_{i,T}(h)\right\}+o_p(1). \tag{A.14}$$

Let $\bar{V}_i=\{k\colon k\in[k_{i-1}^0+\pi T-M,\,k_i^0-\pi T+M]\}$ and $\underline{V}_i=\{k\colon k\in[k_{i-1}^0+\pi T+M,\,k_i^0-\pi T-M]\}$. Then $\underline{V}_i\subset V_i\subset\bar{V}_i$. Neither $\underline{V}_i$ nor $\bar{V}_i$ depends on the j_s's. We have

$$\min_{1\leqslant i\leqslant \ell+1}\ \min_{j_1,\ldots,\,j_\ell;h\in V_i}\left\{\sum_{s=1}^{\ell} W_s(j_s)-\Gamma_{i,T}(h)\right\}$$
$$\geqslant\sum_{s=1}^{\ell}\min_{|j|\leqslant M} W_s(j)+\min_i\ \min_{h\in\bar{V}_i}\ -\Gamma_{i,T}(h) \tag{A.15}$$

and, replacing $\bar{V}_i$ by $\underline{V}_i$,

$$\min_{1\leqslant i\leqslant \ell+1}\ \min_{j_1,\ldots,\,j_\ell;h\in V_i}\left\{\sum_{s=1}^{\ell} W_s(j_s)-\Gamma_{i,T}(h)\right\}$$
$$\leqslant\sum_{s=1}^{\ell}\min_{|j|\leqslant M} W_s(j)+\min_i\ \min_{h\in\underline{V}_i}\ -\Gamma_{i,T}(h) \tag{A.16}$$

Adding and subtracting $S_T(k_1^0,\ldots,k_\ell^0)$ and using Eqs. (A.13), (A.14) and (A.15), we obtain

$$S_T(\hat{k}_1,\ldots,\hat{k}_\ell)-S_T(\hat{k}_1^*,\ldots,\hat{k}_{\ell+1}^*)\leqslant -\min_{1\leqslant i\leqslant \ell+1}\ \min_{h\in\bar{V}_i}\ -\Gamma_{i,T}(h)+o_p(1), \tag{A.17}$$

provided that $|\hat{k}_i-k_i^0|\leqslant M$ and $|\hat{k}_{i_v}^*-k_v^0|\leqslant M$. Similarly, replacing $\bar{V}_i$ by $\underline{V}_i$, we obtain

$$S_T(\hat{k}_1,\ldots,\hat{k}_\ell)-S_T(\hat{k}_1^*,\ldots,\hat{k}_{\ell+1}^*)\geqslant\max_{1\leqslant i\leqslant \ell+1}\ \max_{h\in\underline{V}_i}\Gamma_{i,T}(h)+o_p(1) \tag{A.18}$$

provided that $|\hat{k}_i - k_i^0| \leqslant M$ and $|\hat{k}_{i_v}^* - k_v^0| \leqslant M$. By the result of Bai and Perron (1998), $P(\exists i \text{ s.t. } |\hat{k}_i - k_i^0| > M) < \varepsilon$ and by Lemma A.1, $P(\exists v \text{ s.t. } |\hat{k}_{i_v}^* - k_v^0| > M) < \varepsilon$ for large M. These results and Eq. (A.17) imply that

$$P(\{S_T(\hat{k}_1, \ldots, \hat{k}_\ell) - S_T(\hat{k}_1^*, \ldots, \hat{k}_{\ell+1}^*)\}/\sigma^2 \geqslant x)$$
$$\leqslant 2\varepsilon + P\left(\max_i \max_{h \in \bar{V}_i} \Gamma_{i,T}(h)/\sigma^2 + o_p(1) \geqslant x\right) \tag{A.19}$$

From Eq. (A.5), $\max_i \max_{h \in \bar{V}_i} \Gamma_{i,T}(h)/\sigma^2 + o_p \xrightarrow{d} \max_i \xi_i$. It follows from Eq. (A.19) that (because ε is arbitrary and $\hat{\sigma}^2(\ell+1) \xrightarrow{p} \sigma^2$)

$$\lim_{T \to \infty} P(\sup LR_T(\ell+1|\ell) \geqslant x) \leqslant P\left(\max_i \xi_i \geqslant x\right). \tag{A.20}$$

Similarly, using Eq. (A.18) and noting that Eq. (A.5) holds when V_i is replaced by $\underline{V}_i$, we obtain

$$\lim_{T \to \infty} P(\sup LR_T(\ell+1|\ell) \leqslant x) \leqslant P\left(\max_i \xi_i \leqslant x\right). \tag{A.21}$$

Inequalities Eqs. (A.20) and (A.21) imply

$$\lim_{T \to \infty} P(\sup LR_T(\ell+1|\ell) \geqslant x) = P\left(\max_i \xi_i \geqslant x\right).$$

This proves Theorem 1. □

Proof of Theorem 3. The more demanding part of the proof is the same as that of Theorem 1. In particular, the proof of Eqs. (A.17) and (A.18) needs no change. It remains to show that $\sup_h \Gamma_{i,T}(h)/\sigma^2 \xrightarrow{d} \psi_i$, where ψ_i is specified in the theorem, and $\Gamma_{i,T}(h) = D(k_{i-1}^0, k_i^0) - D(k_{i-1}^0, h) - D(h, k_i^0)$, and the supremum with respect to h is taken over the set $\bar{V}_i$ or $\underline{V}_i$. We note that $\Gamma_{i,T}(h)$ is simply the difference between the restricted sum of squared residuals and the unrestricted sum of squared residuals (allowing for one break at h), computed for the segment $[k_{i-1}^0 + 1, k_i{}^0]$. We consider the case of $i = 1$. Other cases are the same. Write N for k_1^0. Let $x_t = (w_t', z_t')'$, and $X_h = (x_1, x_2, \ldots, x_h)'$ $(h = 1, \ldots, N)$. Let $U_h = (u_1, \ldots, u_h)'$ $(h = 1, \ldots, N)$. Then

$$D(0, N) - D(0, h) - D(h, N) = V_h'(X_h'X_h - (X_h'X_h)(X_N'X_N)^{-1}(X_h'X_h))^{-1}V_h,$$

where $V_h = X_h'U_h - (X_h'X_h)(X_N'X_N)^{-1}X_N'U_N$. Let $\Upsilon_N = \mathrm{diag}(N^{1/2}, N^{3/2}, \ldots, N^{(2p+1)/2}, N^{1/2}I_q)$, where I_q is the $q \times q$ identity matrix. Because a constant is included in the regressor, we may assume $Ez_t = 0$. Then $\Upsilon_N^{-1}(X_{[Ns]}'X_{[Ns]})\Upsilon_N^{-1} \to \mathrm{diag}(A_s, sR_{zz}) \equiv Q_s$, where $R_{zz} = \mathrm{plim}\, N^{-1}\sum_1^N z_t z_t'$ and $A_s = A(s)$ is defined in the text. Furthermore, $\Upsilon_N^{-1}X_{[N\cdot]}U_{[N\cdot]} \Rightarrow \sigma W(\cdot)$, where $W(s)$ is a Gaussian process with $EW(s)W(r)' = Q_{s\wedge r}$. Let $V(s) = W(s) - Q_sQ_1^{-1}W(1)$. We have

$$\begin{aligned} D(0,N) - D(0,[Ns]) - D([Ns],N) &\Rightarrow \sigma^2 V(s)'(Q_s - Q_sQ_1^{-1}Q_s)^{-1}V(s) \\ &= \sigma^2 V^*(s)'(Q_1^{-1/2}Q_sQ_1^{-1/2} \\ &\quad - (Q_1^{-1/2}Q_sQ_1^{-1/2})^2)^{-1}V^*(s), \quad \text{(A.22)} \end{aligned}$$

where $V^*(s) = Q_1^{-1/2}V(s)$. The variance–covariance matrix of $V^*(s)$ is $Q_1^{-1/2}Q_sQ_1^{-1/2} - (Q_1^{-1/2}Q_sQ_1^{-1/2})^2$. Let $\lambda_1(s), \ldots, \lambda_{p+1}(s)$ be the eigenvalues of $A_1^{-1/2}A_sA_1^{-1/2}$ (they are chosen to be continuous and increasing in s), and let $\lambda_{p+1+i}(s) = s$ for $i = 1, \ldots, q$. Then there exists an orthogonal matrix C_s, such that $C_sQ_1^{-1/2}Q_sQ_1^{-1/2}C_s' = \mathrm{diag}(\lambda_i(s), i = 1, \ldots, p+1+q)$. Thus $C_sV^*(s)$ is a vector of Gaussian processes with $E[C_sV^*(s)V^*(r)'C_r'] = \mathrm{diag}[\lambda_i(s\wedge r) - \lambda_i(s)\lambda_i(r); i = 1, \ldots, p+1+q]$. The last q components of $C_sV^*(s)$ are standard Brownian bridges, and the first $p+1$ components are time-scaled Brownian bridges. Thus the limiting distribution in Eq. (A.22) has the following representation:

$$\sum_{j=1}^{p+1} \frac{B_j(\lambda_j(s))^2}{\lambda_j(s)(1-\lambda_j(s))} + \sum_{j=p+2}^{p+1+q} \frac{B_j(s)^2}{s(1-s)}, \quad \text{(A.23)}$$

where the $B_j(\cdot)$ are standard Brownian bridges on $[0, 1]$. Finally note that, the smallest h in $\underline{V}_1$ (or in $\overline{V}_1$) satisfies $h/N \to \eta_1$, and the largest h satisfies $h/N \to 1 - \eta_1$. This, together with Eq. (A.23), yields the representation of ψ_1 given in the theorem. The independence of $\psi_1, \ldots, \psi_{\ell+1}$ follows from the fact that they are derived from non-overlapping segments. □

Proof of Corollary 2. We only need to find the expression for $H_1(c) = P(\psi_1 > c)$. Let $t_i = \frac{1}{2}\log(\lambda_i(s)/(1-\lambda_i(s))$. Using the argument of Chan (1991), we have

$$P(\psi_1 > c) \approx 1 - \exp\left(-2\chi_{p+1+q}(c)\left(\frac{c}{p+1+q} - 1\right)\sum_{i=1}^{p+1+q}\int_{\eta_1}^{1-\eta_1} \frac{dt_i}{ds}ds\right),$$

where $\chi_{p+1+q}(c)$ is the density function of a Chi-square random variable with $p+1+q$ degrees of freedom. From $\exp(-x) \approx 1 - x$ for small x,

and

$$\int_{\eta_1}^{1-\eta_1} \mathrm{d}t_i = \frac{1}{2}\log\frac{\lambda_i(1-\eta_1)(1-\lambda_i(\eta_1))}{\lambda_i(\eta_1)(1-\lambda_i(1-\eta_1))},$$

we have

$$\mathrm{P}(\psi_1 > c) \approx 2\chi_{p+1+q}(c)\left(\frac{c}{p+1+q} - 1\right)$$
$$\left[\frac{1}{2}\sum_{i=1}^{p+1}\log\frac{\lambda_i(1-\eta_1)(1-\lambda_i(\eta_1))}{\lambda_i(\eta_1)(1-\lambda_i(1-\eta_1))} + q\log\frac{1-\eta_1}{\eta_1}\right],$$

since the last q eigenvalues are identity functions. Using the fact that $\lambda_1(s)\cdots\lambda_{p+1}(s) = \det(A_1^{-1/2}A_sA_1^{-1/2}) = \det(A_1^{-1})\det A(s)$ and $(1-\lambda_1(s))\cdots(1-\lambda_{p+1}(s)) = \det(I - A_1^{-1}A_sA_1^{-1}) = \det(A_1^{-1})\det[A_1 - A(s)]$, and substituting the density function for $\chi_{p+1+q}(c)$, we obtain the expression of $H_1(c)$ given in the corollary. □

References

Andrews, D.W.K., 1993. Testing for structural instability and structural change with unknown change point. Econometrica 61, 821–856.

Andrews, D.W.K., Lee, I., Ploberger, W., 1996. Optimal changepoint tests for normal linear regression. Journal of Econometrics 70, 9–38.

Bai, J., Perron, P., 1998. Estimating and testing for multiple structural changes in linear models. Econometrica 66, 47–78.

Banerjee, A., Lumsdaine, R.L., Stock, J.H., 1992. Recursive and sequential tests of the unit root and trend break hypothesis: theory and international evidence. Journal of Business and Economic Statistics 10, 271–287.

Broemeling, L. (Ed.), 1982. Structural change in econometrics. Journal of Econometrics 19, 1–182.

Chan, K.S., 1991. Percentage points of likelihood ratio tests for threshold autoregression. Journal of Royal Statistical Society, Series B 53, 691–696.

Chow, G.C., 1960. Tests of equality between sets of coefficients in two linear regressions. Econometrica 28, 591–605.

DeLong, D.M., 1981. Crossing probabilities for a square root boundary by a Bessel process. Communications in Statistics-Theory and Methods A 10 (21), 2197–2213.

Diebold, F.X., Rudebusch, G.D., 1994. Measuring business cycles: a modern perspective. NBER Working paper, No. 4643.

Dufour, J.M., Ghysels, E. (Eds.), 1996. Recent developments in the econometrics of structural change. Journal of Econometrics, 70, 1–316.

Dufour, J.M., Kiviet, J.F., 1996. Exact tests for structural change in first-order dynamic models. Journal of Econometrics 70, 39–68.

Hackl, P.D., Westlund, A.H., (Eds.), 1991. Economics Structural Change: Analysis and Forecasting. Springer-Verlag, New York.

Hansen, B., 1996. Sample-splitting and threshold estimation. Manuscript, Department of Economics, Boston College.

Hansen, B.E., 1992. Tests for parameter instability in regressions with I(1) processes. Journal of Business and Economic Statistics 10, 321–335.

Hegwood, N.D., Papell, D.H., 1996. Quasi purchasing power parity. Manuscript, Department of Economics. University of Houston.

Kramer, W., 1989. Econometrics of Structural Change. Springer, New York.

Kramer, W., Ploberger, W., Alt, R., 1988. Testing for structural changes in dynamic models. Econometrica 56, 1355–1370.

Quandt, R.E., 1960. Tests of the hypothesis that a linear regression system obeys two separate regimes. Journal of American Statistical Association 55, 320–330.

Sowell, F., 1996. Optimal tests for parameter instability in the generalized method of moments framework. Econometrica 64, 1085–1107.

Tong, H., 1990. Non-Linear Time Series: A Dynamical System Approach. Oxford University Press, Oxford.

Vogelsang, T., 1994. Testing for a shift in mean without having to estimate serial correlation parameters. Manuscript, Department of Economics. Cornell University.

Willard, K.L., Guinnane, T.W., Rosen, S.H., 1996. Turning points in the civil war: views from the Greenback market. American Economic Review 86, 1001–1018.

Zivot, E., Andrews, D.W.K., 1992. Further evidence on the great crash, the oil price shock and the unit root hypothesis. Journal of Business and Economic Statistics 10, 251–271.