Journal of Econometrics 146 (2008) 59–73

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

The limit distribution of the estimates in cointegrated regression models with multiple structural changes[☆]

Mohitosh Kejriwal[a], Pierre Perron[b,*]

[a] Krannert School of Management, Purdue University, 403 West State Street, West Lafayette, IN 47907, United States

[b] Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, United States

ARTICLE INFO

Article history:
Received 11 August 2007
Received in revised form
21 March 2008
Accepted 4 July 2008
Available online 11 July 2008

JEL classification:
C22

Keywords:
Change-point
Break dates
Unit roots
Cointegration
Confidence intervals

ABSTRACT

We study estimation and inference in cointegrated regression models with multiple structural changes allowing both stationary and integrated regressors. Both pure and partial structural change models are analyzed. We derive the consistency, rate of convergence and the limit distribution of the estimated break fractions. Our technical conditions are considerably less restrictive than those in Bai et al. [Bai, J., Lumsdaine, R.L., Stock, J.H., 1998. Testing for and dating breaks in multivariate time series. Review of Economic Studies 65, 395–432] who considered the single break case in a multi-equations system, and permit a wide class of practically relevant models. Our analysis is, however, restricted to a single equation framework. We show that if the coefficients of the integrated regressors are allowed to change, the estimated break fractions are asymptotically dependent so that confidence intervals need to be constructed jointly. If, however, only the intercept and/or the coefficients of the stationary regressors are allowed to change, the estimates of the break dates are asymptotically independent as in the stationary case analyzed by Bai and Perron [Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78]. We also show that our results remain valid, under very weak conditions, when the potential endogeneity of the non-stationary regressors is accounted for via an increasing sequence of leads and lags of their first-differences as additional regressors. Simulation evidence is presented to assess the adequacy of the asymptotic approximations in finite samples.

1. Introduction

Issues related to structural change have received considerable attention in the statistics and econometrics literature (see Perron (2006), for a survey). In the last fifteen years or so, substantial advances have been made in the econometrics literature to cover models at a level of generality that allows a host of interesting practical applications in the context of unknown change points. These include models with general stationary regressors and errors that can exhibit temporal dependence and heteroskedasticity. Andrews (1993) and Andrews and Ploberger (1994) provide a comprehensive treatment of the problem of testing for structural change assuming that the change point is unknown. Bai (1997) studies the least squares estimation of a single change point in regressions involving stationary and/or trending regressors. He derives the consistency, rate of convergence and the limiting distribution of the change point estimator under general conditions on the regressors and the errors. Bai and Perron (1998) extend the testing and estimation analysis to the case of multiple structural changes, while Bai and Perron (2003) present an efficient algorithm to obtain the break dates corresponding to the global minimizers of the sum of squared residuals. Perron and Qu (2006) consider the case in which restrictions within or across regimes are imposed. Qu and Perron (2007) cover the more general case of multiple structural changes in a system of equations allowing arbitrary restrictions on the parameters.

When dealing with non-stationary variables, the literature is less extensive. With respect to testing, Hansen (1992b) develops tests of the null hypothesis of no change in models where all coefficients are allowed to change. An extension to partial changes has been analyzed by Kuo (1998). The tests considered are the Sup and Mean LM tests directed against an alternative of a one time change in parameters. Hao (1996) also suggests the use of the exponential LM test. Seo (1998) considers the Sup, Mean and Exp versions of the LM test within a cointegrated VAR setup. The Sup and Mean LM tests in this setup are shown to have a similar asymptotic distribution as the Sup and Mean LM tests of Hansen (1992b). Kejriwal and Perron (2008) show that such tests can suffer

☆ Perron acknowledges financial support for this work from the National Science Foundation under Grant SES-0649350. We are grateful to two referees for useful comments.

* Corresponding author. Tel.: +1 617 353 4389.
E-mail addresses: mkejriwa@purdue.edu (M. Kejriwal), perron@bu.edu (P. Perron).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.07.001

from important lack of power in finite samples and be subject to a non-monotonic power function such that the power decreases as the magnitude of the break increases. They suggest modified Sup–Wald type tests that perform considerably better.

With respect to estimation, Perron and Zhu (2005) analyze the properties of parameter estimates in models where the trend function exhibits a slope change at an unknown date and the errors can be either stationary or have a unit root. With integrated variables, the case of most interest is that of a framework in which the variables are cointegrated. Accounting for parameter shifts is crucial in cointegration analysis since it normally involves long spans of data which are more likely to be affected by structural breaks. The goal is then to test whether the cointegrating relationship has changed and to estimate the break dates and form confidence intervals for them. In this respect, an important paper is that of Bai et al. (1998) who consider a single break in a multi-equations system and show the estimates obtained by maximizing the likelihood function to be consistent. They also obtain a limit distribution of the estimate of the break date under a shrinking shift scenario assuming that the coefficients associated with the trend and the non-stationary regressors shrink faster than those pertaining to the stationary regressors.

The aim of this paper is to provide a comprehensive treatment of issues related to estimation and inference with multiple structural changes, occurring at unknown dates, in cointegrated regression models. Our work builds on that of Bai and Perron (1998) who undertake a similar treatment in a stationary framework. Our framework is general enough to allow both stationary and non-stationary variables in the regression. The assumptions regarding the distribution of the error processes are mild enough to allow for general forms of serial correlation and conditional heteroskedasticity, as well as mild forms of unconditional heteroskedasticity. Moreover, we analyze both pure and partial structural change models. A partial change model is useful in allowing potential savings in the number of degrees of freedom, an issue particularly relevant for multiple changes. It is also important in empirical work since it helps to isolate the variables which are responsible for the failure of the null hypothesis. The parameter estimates of the regression coefficients and the break dates are obtained by minimizing the sum of squared residuals. We derive the consistency, rate of convergence and limiting distribution of the estimated break fractions under much weaker conditions than those in Bai et al. (1998). We show that if the coefficients of the integrated regressors are allowed to change, the estimated break fractions are asymptotically dependent so that confidence intervals need to be constructed jointly. Methods to construct such confidence intervals are discussed. If, however, only the intercept and/or the coefficients of the stationary regressors are allowed to change, the estimates of the break dates are asymptotically independent as in the stationary framework analyzed by Bai and Perron (1998). Though our theoretical results hold under much weaker conditions than those of Bai et al. (1998) and allow for multiple breaks, our analysis is restricted to a single cointegrating vector unlike theirs which is valid in a multi-equations system which, thereby allows multiple cointegrating vectors. In the multiple break case, the fact that the estimated break fractions are asymptotically dependent complicates the analysis considerably and the extension to a multi-equations system is outside the scope of this paper.

This article is organized as follows. Section 2 presents the model and assumptions. In Section 3, we derive the consistency, rate of convergence and limiting distribution of the estimates of the break dates. Section 4 presents the results of simulation experiments to assess the adequacy of the asymptotic approximations in finite samples. Section 5 offers concluding remarks and all technical derivations are included in a mathematical Appendix.

2. The model and assumptions

Consider the following linear regression model with m breaks ($m+1$ regimes):

$$y_t = c_j + z_{ft}'\delta_f + z_{bt}'\delta_{bj} + x_{ft}'\beta_f + x_{bt}'\beta_{bj} + u_t \quad (t = T_{j-1}+1, \ldots, T_j) \tag{1}$$

for $j = 1, \ldots, m+1$, where $T_0 = 0$, $T_{m+1} = T$ and T is the sample size. In this model, $x_{ft}\,(p_f \times 1)$ and $x_{bt}\,(p_b \times 1)$ are vectors of $I(0)$ variables while $z_{ft}\,(q_f \times 1)$ and $z_{bt}\,(q_b \times 1)$ are vectors of $I(1)$ variables defined by

$$z_{ft} = z_{f,t-1} + u_{zt}^f$$
$$z_{bt} = z_{b,t-1} + u_{zt}^b$$
$$x_{ft} = \mu_f + u_{xt}^f$$
$$x_{bt} = \mu_b + u_{xt}^b$$

where z_{f0} and z_{b0} are assumed, for simplicity, to be either $O_p(1)$ random variables or fixed finite constants. For ease of reference, the subscript b on the error term stands for "break" and the subscript f stands for "fixed" (across regimes). By labeling the regressors x_{ft} and x_{bt} as $I(0)$, we mean that the partial sums of the associated noise components satisfy a functional central limit theorem. The conditions imposed are discussed below. We then label a variable as $I(1)$ if it is the accumulation of an $I(0)$ process.

The break points $(T_1, \ldots, T_m)$ are treated as unknown. This is a partial structural change model in which the coefficients of only a subset of the regressors are subject to change while the remaining coefficients are effectively estimated using the entire sample. When $p_f = q_f = 0$, a pure structural change model is obtained where all coefficients are allowed to change across regimes.[1] We can express (1) in matrix form as:

$$Y = G\alpha + \bar{W}\gamma + U$$

where $Y = (y_1, \ldots, y_T)'$, $G = (Z_f, X_f)$, $Z_f = (z_{f1}, \ldots, z_{fT})'$, $X_f = (x_{f1}, \ldots, x_{fT})'$, $U = (u_1, \ldots, u_T)'$, $W = (w_1, \ldots, w_T)'$, $w_t = (z_{bt}', x_{bt}')'$, $\gamma = (\delta_{b1}', \beta_{b1}', \ldots, \delta_{b,m+1}', \beta_{b,m+1}')'$, $\alpha = (\delta_f', \beta_f')'$ and $\bar{W}$ is the matrix which diagonally partitions W at the m-partition $(T_1, \ldots, T_m)$, that is, $\bar{W} = \text{diag}(W_1, \ldots, W_{m+1})$ with $W_i = (w_{T_{i-1}+1}, \ldots, w_{T_i})'$ for $i = 1, \ldots, m+1$. The data generating process is assumed to be

$$Y = G\alpha^0 + \bar{W}^0\gamma^0 + U \tag{2}$$

where α^0, γ^0 and $(T_1^0, \ldots, T_m^0)$ are the true values of the parameters and the matrix $\bar{W}^0$ is the one that partitions W at $(T_1^0, \ldots, T_m^0)$.

As a matter of notation, "$\xrightarrow{p}$" denotes convergence in probability, "$\xrightarrow{d}$" convergence in distribution and "$\Rightarrow$" weak convergence in the space $D[0, 1]$ under the Skorohod metric. Also, $x_t = (x_{ft}', x_{bt}')'$, $u_{xt} = (u_{xt}^{f\prime}, u_{xt}^{b\prime})'$, $z_t = (z_{ft}', z_{bt}')'$, $u_{zt} = (u_{zt}^{f\prime}, u_{zt}^{b\prime})'$, $\xi_t = (u_t, u_{zt}^{f\prime}, u_{zt}^{b\prime}, u_{xt}^{f\prime}, u_{xt}^{b\prime})'$, $\mu = (\mu_f', \mu_b')'$ and $\lambda = \{\lambda_1, \ldots, \lambda_m\}$ is the vector of break fractions defined by $\lambda_i = T_i/T$ for $i = 1, \ldots, m$. We make the following assumptions on the regressors and the elements of the noise component ξ_t.

[1] Note that (1) assumes a particular normalization of the cointegrating vector. Ng and Perron (1997) study the normalization problem in a two variable models. They show that the least squares estimator can have very poor finite sample properties when normalized in one direction but can be well behaved when normalized in the other. This occurs when one of the variables is a weak random walk or is nearly stationary. They suggest to use as regressand the variable for which the spectral density at frequency zero of the first differences is smallest.

- Assumption A1: Let $f_t' = (z_{ft}', x_{ft}', z_{bt}', x_{bt}')$, $F = (f_1, \ldots, f_T)'$ and $\bar{F}^0$ be the diagonal partition of F at $(T_1^0, \ldots, T_m^0)$ such that $\bar{F}^0 = \text{diag}(F_1^0, \ldots, F_{m+1}^0)$. Define the matrix $D_{0i} = \text{diag}((T_i^0 - T_{i-1}^0)^{-1} I_{q_f}, (T_i^0 - T_{i-1}^0)^{-1/2} I_{p_f}, (T_i^0 - T_{i-1}^0)^{-1} I_{q_b}, (T_i^0 - T_{i-1}^0)^{-1/2} I_{p_b})$. We assume that for each $i = 1, \ldots, m+1$, $D_{0i} F_i^{0\prime} F_i^0 D_{0i}$ converges to a random matrix not necessarily the same for all i.
- Assumption A2: Define the matrix $\tilde{D} = \text{diag}(T^{-1/2} I_{q_f}, I_{p_f}, T^{-1/2} I_{q_b}, I_{p_b})$. There exists an $l_0 > 0$ such that for all $l > l_0$, the minimum eigenvalues of $A_{il} = (1/l)\tilde{D} \sum_{T_i^0+1}^{T_i^0+l} f_t f_t' \tilde{D}$ and of $A_{il}^* = (1/l)\tilde{D} \sum_{t=T_i^0-l}^{T_i^0} f_t f_t' \tilde{D}$ are bounded away from zero ($i = 1, \ldots, m+1$).
- Assumption A3: The matrix $B_{kl} = \sum_{t=k}^{l} w_t w_t'$ is invertible for $l - k \geq q_b + p_b$.
- Assumption A4: Let $\xi_t^* = (u_{zt}^{f\prime}, u_{zt}^{b\prime}, u_{xt}^{f\prime}, u_{xt}^{b\prime})'$ and $p = p_f + p_b + q_f + q_b$. The vector $\{\xi_t^* u_t\}$ satisfies Assumption A4 in Qu and Perron (2007). Define the L_r-norm of a random matrix X as $\|X\|_r = (\sum_i \sum_j E|X_{ij}|^r)^{1/r}$ for $r \geq 1$ and $\mathcal{F}_t = \sigma$-*field* $\{\ldots, \xi_{t-1}^*, \xi_t^*, \ldots, u_{t-2}, u_{t-1}\}$. If $\xi_t^* u_t$ is weakly stationary within each segment, then (a) $\{\xi_t^* u_t, \mathcal{F}_t\}$ forms a strongly mixing (α-mixing) sequence with size $-4r/(r-2)$ for some $r > 2$, (b) $E(u_t) = 0$ and $\|\xi_t^* u_t\|_{2r+\delta} < M < \infty$ for some $\delta > 0$, (c) Let $S_{k,j}(\ell) = \sum_{T_{j-1}^0+\ell+1}^{T_{j-1}^0+\ell+k} \xi_t^* u_t$, $j = 1, \ldots, m+1$, for each $e \in R^n$ of length 1, var $(\langle e, S_{k,j}(0)\rangle) \geq v(k)$ for some function $v(k) \to \infty$ as $k \to \infty$ (with $\langle\cdot\rangle$, the usual inner product). If $\xi_t^* u_t$ is not weakly stationary within each segment, we assume that (a)–(c) holds, and in addition, that there exists a positive definite matrix $\Omega = [w_{i,s}]$ such that for any $i, s = 1, \ldots, p$, we have, uniformly in ℓ, $\left|k^{-1} E\left((S_{k,j}(\ell))_i (S_{k,j}(\ell))_s\right) - w_{i,s}\right| \leq C_2 k^{-\psi}$, for some C_2, $\psi > 0$. It is also assumed that $\{\xi_t\}$ satisfies the conditions stated in this assumption.
- Assumption A5: $E(u_{xt} u_t) = 0$.
- Assumption A6: $0 < \lambda_1^0 < \cdots < \lambda_m^0 < 1$ with $T_i^0 = [T\lambda_i^0]$.
- Assumption A7: Let $\bar{\gamma}_{T,i} = \gamma_{T,i+1}^0 - \gamma_{T,i}^0$, then $\bar{\gamma}_{T,i} = \text{diag}(T^{-1/2} I_{q_b}, I_{p_b}) \bar{\gamma}_i v_T$, for some $\bar{\gamma}_i$ independent of T, where $v_T > 0$ is a scalar satisfying $v_T \to 0$ and $T^{1/2} v_T \to \infty$ as $T \to \infty$.

Assumption A1 is needed for multiple linear regressions involving both stationary and integrated regressors and simply indicates that sample moments of the regressors exists when scaled as stated. Assumption A2 ensures that there is no local collinearity problem so that the break points can be identified. The use of the weighting matrices D_{0i} and $\tilde{D}$ is due to the presence of both $I(1)$ and $I(0)$ regressors. Assumption A3 is a standard invertibility requirement to have well defined estimates. Assumption A4 determines the dependence structure of the processes $\xi_t^* u_t$ and ξ_t. In particular, they imply that $\xi_t^* u_t$ and ξ_t are short memory processes having bounded fourth moments. The assumptions are imposed to obtain a functional central limit theorem, a generalized Hájek and Rényi (1955) type inequality and a strong law of large numbers that allow us to show the estimates to be consistent and to derive their rate of convergence. The conditions are mild in the sense that they allow for substantial conditional heteroskedasticity and autocorrelation. Also, if no autocorrelation is present, i.e., $\{\xi_t^* u_t\}$ and $\{\xi_t\}$ are martingale difference sequences with respect to the filtration $\mathcal{F}_t$, then even the weak stationarity assumption can be dropped and ξ_t allowed to be unconditionally heteroskedastic with bounded fourth moments. The conditions for this to hold are very general (see, e.g., Davidson (1994)). It can be shown to apply to a large class of linear processes including those generated by all stationary and invertible ARMA models. Note that Assumption A4 could be replaced by other sufficient conditions that can yield the main ingredients stated above.[2]

Assumption A5 specifies that the stationary regressors are contemporaneously uncorrelated with the errors. This is a standard requirement to obtain consistent estimates. It is important to note that no such assumption is imposed on the correlation between the innovations to the $I(1)$ regressors and the errors. Hence, we allow endogenous $I(1)$ regressors. Assumption A6 implies asymptotically distinct breaks, i.e. each regime contains a positive fraction of the sample even in the limit. Assumption A7 implies a shrinking shifts asymptotic framework where the magnitudes of the shifts converge to zero as the sample size increases. Specifically, we assume that the coefficients associated with the $I(1)$ variables shrink at a faster rate than those associated with the $I(0)$ variables. Note that Bai and Perron (1998) assume all coefficients to shrink at the same rate since all regressors in their framework are assumed to be stationary. Moreover, since breaks of larger magnitude are easier to identify, consistency of the break fractions assuming a small magnitude of shift imply consistency for breaks of larger magnitude.

Our set of assumptions is considerably weaker than those of Bai et al. (1998) who impose the following conditions: (a) the errors u_t are independent of the regressors at all leads and lags, which precludes, among other things, endogenous $I(1)$ regressors, (b) the noise components are linear processes with *i.i.d.* errors, (c) some bound on the expectation of some functions of the squared regressors (see their Assumptions 3.3 and 3.4), (d) zero mean stationary regressors. Hence, our framework allows a much wider variety of models that are of interest in applied work. For the rate at which the magnitude of the breaks shrink to zero, Bai et al. (1998) also impose the requirement that $T^{1/2} v_T / \log(T) \to \infty$ as $T \to \infty$. Our condition is slightly weaker.

Estimates of the parameters are obtained by minimizing the global sum of squared residuals. For each m-partition $(T_1, \ldots, T_m)$, denoted $\{T_j\}$, the associated least squares estimates of α and γ are obtained by minimizing

$$\text{SSR}_T(T_1, \ldots, T_m) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} \left[y_t - c_i - z_{ft}'\delta_f - x_{ft}'\beta_f - z_{bt}'\delta_{bi} - x_{bt}'\beta_{bi}\right]^2. \quad (3)$$

Let $\hat{\alpha}(\{T_j\})$ and $\hat{\gamma}(\{T_j\})$ be the resulting estimates. Substituting these into the objective function and denoting the resulting sum of squared residuals as $S_T(T_1, \ldots, T_m)$, the estimates of the break dates $(\hat{T}_1, \ldots, \hat{T}_m)$ are such that

$$(\hat{T}_1, \ldots, \hat{T}_m) = \arg\min_{T_1, \ldots, T_m} S_T(T_1, \ldots, T_m). \quad (4)$$

Throughout, we also impose the following assumption on the set of permissible partitions where ε acts as a trimming parameter and imposes a minimal length for each regime.

- Assumption A8: The minimization (4) is taken over all partitions $(T_1, \ldots, T_m) = (T\lambda_1, \ldots, T\lambda_m)$ in the set

$$\Lambda_\varepsilon = \left\{(\lambda_1, \ldots, \lambda_m); \left|\lambda_{j+1} - \lambda_j\right| \geq \varepsilon, \lambda_1 \geq \varepsilon, \lambda_m \leq 1 - \varepsilon\right\}. \quad (5)$$

This assumption is not very restrictive given that ε can be small. The estimates of the regression coefficients are then $\hat{\alpha} = \hat{\alpha}(\{\hat{T}_j\})$

[2] Examples of such conditions are discussed by Dehling and Philipp (1982), Altissimo and Corradi (2003) and Lavielle and Moulines (2000), among others.

and $\hat{\gamma} = \hat{\gamma}(\{\hat{T}_j\})$. The estimates of the parameters can be obtained using the algorithm of Bai and Perron (2003) with no modification since the algorithm itself is valid irrespective of the nature of the regressors and errors given that it is designed to obtain estimates of the break dates that minimize the global sum of squared residuals in a regression with some or all coefficients allowed to change across a pre-specified number of regimes.

Finally, note that trends of the form $(t/T)^i$ for $i = 1, \ldots, d$, say, are allowed for the $I(0)$ regressors. Extending the analysis to $I(0)$ and $I(1)$ variables with unscaled trends of the form t^i is straightforward for $I(0)$ variables with minor modifications of the scaling matrices in Assumptions A1, A2 and A7. The results stated below about the consistency and rate of convergence go through, though not the result about the limit distribution, which would require some modifications. We can allow $I(1)$ regressors with trend of the form

$$z^*_{ft} = \rho_f t + z_{ft}$$
$$z^*_{bt} = \rho_b t + z_{bt}$$

by including a linear time trend as a regressor in (1). Moreover, since z^*_{bt} behaves asymptotically like a time trend, the rate of decrease of the shifts needs to be specified as $\delta^0_{b,i+1} - \delta^0_{b,i} = T^{-1}\bar{\delta}_{b,i}v_T$ and $\bar{\delta}_{b,i}\rho_b \neq 0$. Following the arguments in Bai et al. (1998), all results about consistency, rate of convergence and even limit distribution carry through.

3. Consistency and rates of convergence

3.1. Consistency

Let $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_m)$ with corresponding true values $\lambda^0 = (\lambda^0_1, \ldots, \lambda^0_m)$. The following Theorem states the consistency of $\hat{\lambda}$ for λ^0.

Theorem 1. *Under* A1–A8*: $\hat{\lambda}_k \xrightarrow{p} \lambda^0_k$, $k = 1, \ldots, m$.*

To prove the theorem, we need to establish two lemmas. Let $\hat{u}_t = y_t - g'_t\hat{\alpha} - w'_t\hat{\gamma}_k$, for $t \in [\hat{T}_{k-1} + 1, \hat{T}_k]$ and $d_t = g'_t(\hat{\alpha} - \alpha^0) + w'_t(\hat{\gamma}_k - \hat{\gamma}^0_j)$, for $t \in [\hat{T}_{k-1} + 1, \hat{T}_k] \cap [T^0_{j-1} + 1, T^0_j]$ $(k, j = 1, \ldots, m + 1)$. By the properties of projections,

$$\sum_{t=1}^{T} \hat{u}_t^2 \leq \sum_{t=1}^{T} u_t^2. \tag{6}$$

Since $\hat{u}_t = u_t - d_t$,

$$\sum_{t=1}^{T} \hat{u}_t^2 = \sum_{t=1}^{T} u_t^2 + \sum_{t=1}^{T} d_t^2 - 2\sum_{t=1}^{T} u_t d_t. \tag{7}$$

Now we have the following first lemma:

Lemma 1. *Assume* A1–A8 *hold, then $\sum_{t=1}^{T} u_t d_t = O_p(T^{1/2}v_T)$.*

To prove the Theorem, we will prove that $\sum_{t=1}^{T} d_t^2 > 2\sum_{t=1}^{T} u_t d_t$ in the limit as $T \to \infty$. To do this, we show that $\sum_{t=1}^{T} d_t^2$ diverges at a faster rate than $\sum_{t=1}^{T} u_t d_t$ if any estimate of the break fractions is not consistent. This gives us the desired contradiction from (6). The following lemma states the rate of divergence of $\sum_{t=1}^{T} d_t^2$ in such a case.

Lemma 2. *Assume* A1–A8 *hold and that $\hat{\lambda}_j \overset{p}{\nrightarrow} \lambda^0_j$ for some j; then*

$$\lim_{T\to\infty} \sup P\left(T^{-2}\sum_{t=1}^{T} d_t^2 > C\left\|\delta^0_{bj} - \delta^0_{bj+1}\right\|^2\right) > \epsilon_0$$

for some $C > 0$ and $\epsilon_0 > 0$.

Since $\sum_{t=1}^{T} d_t^2 > T^2C\|\delta^0_{bj} - \delta^0_{bj+1}\|^2 > TC'v_T^2$, $\sum_{t=1}^{T} d_t^2 > 2\sum_{t=1}^{T} u_t d_t$ if $Tv_T^2/T^{1/2}v_T \to \infty$, which follows from Assumption A7. This proves Theorem 1.

Remark 1. If the breaks occur only in the coefficients associated with the $I(0)$ regressors, Lemma 1 still holds. Moreover, we can show that $\sum_{t=1}^{T} d_t^2 > TC\|\beta^0_{bj} - \beta^0_{bj+1}\|^2 > TC'v_T^2$ which also proves Theorem 1 in this case.

3.2. Rates of convergence

We now show that $\hat{\lambda}_k$ converges to λ^0_k at rate Tv_T^2, which is stated in the following theorem.

Theorem 2. *Under* A1–A8*, we have $Tv_T^2(\hat{\lambda}_k - \lambda^0_k) = O_p(1)$ for $k = 1, \ldots, m$.*

Remark 2. Note that Bai et al. (1998) assume $T^{1/2}v_T/\log T \to \infty$. Thus our results on consistency and rates of convergence are valid under weaker conditions.

Given the above rate of convergence, the limiting distribution of the estimates of the regression coefficients is the same as that obtained when the break dates are known.

Proposition 1. *Let $\theta = (\alpha, \gamma)$, $\bar{F}^0 = (G, \bar{W}^0)$. Define the $((m + 1)(q_b + p_b) \times (m + 1)(q_b + p_b))$ matrices*

$$\tilde{D}_{m+1} = \mathrm{diag}(T^{-1}I_{q_b}, T^{-1/2}I_{p_b}, \ldots, T^{-1}I_{q_b}, T^{-1/2}I_{p_b})$$
$$\tilde{D}_T = \mathrm{diag}(T^{-1}I_{q_f}, T^{-1/2}I_{p_f}, \tilde{D}_{m+1}).$$

Assume A1–A8 *hold, then we have $\tilde{D}_T^{-1}(\hat{\theta} - \theta^0) \xrightarrow{d} H^{-1}\kappa$, where $H = p\lim_{T\to\infty}(\tilde{D}_T\bar{F}^{0\prime}\bar{F}^0\tilde{D}_T)$ and κ is the limiting distribution of $\tilde{D}_T\bar{F}^{0\prime}U$.*

3.3. The limiting distribution of the estimates of the break dates

We now consider the limit distribution of the estimates of the break dates. We first impose the following condition:

- Assumption A9: Let $\Delta T^0_j = T^0_j - T^0_{j-1}$; for $j = 1, \ldots, m$, as $\Delta T^0_j \to \infty$, uniformly in $s \in [0, 1]$,
  $(\Delta T^0_j)^{-1}\sum_{t=T^0_{j-1}+1}^{T^0_{j-1}+[s\Delta T^0_j]} x_t x'_t \to_p sQ_{xx}$ where
  $$Q_{xx} = \begin{bmatrix} Q^{ff}_{xx} & Q^{fb}_{xx} \\ Q^{bf}_{xx} & Q^{bb}_{xx} \end{bmatrix}$$
  is a nonrandom positive definite matrix.

Assumption A9 rules out trending regressors with a stationary noise component. Note also that we have imposed the same distribution for the $I(0)$ regressors across segments, contrary to what is customary in the recent literature. Relaxing this assumption would be relatively straightforward following the same arguments as in Bai and Perron (1998), but allowing for heterogeneity in the distribution of the errors underlying the $I(1)$ regressors would be considerably more difficult. Instead of having a limit distribution in terms of standard Wiener processes, we would have time-deformed Wiener processes according to the variance profile of the errors through time; see, e.g., Cavaliere and Taylor (2007). This would lead to important complications given that, as shown below, the limit distribution of the estimates of the break dates depends on the whole time profile of the limit Wiener processes. For these reasons, we restrict the analysis to the case of homogeneous distributions across segments, and do so for both the $I(1)$ and $I(0)$ regressors as well as the error of the regression. Given this, Assumption A4 implies the following distributional results which also sets up the notation to be used.

$$A(v_1,\ldots,v_m)=\sum_{i=1}^{m}|v_i|$$

$$B(v_1,\ldots,v_m)=\sum_{i=1}^{m}\left[\frac{\bar{\gamma}_i'\begin{pmatrix}(\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_i^0)\sigma\eta^{(i)}(v_i)\\ \sigma\eta^{(i)}(v_i)\\ (Q_{xu}^{bb})^{1/2}W_{xb}^{(i)}(v_i)\end{pmatrix}}{\left\{\bar{\gamma}_i'\begin{pmatrix}(\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_i^0)W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2} & (\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_i^0) & (\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_i^0)\mu_b'\\ W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2} & 1 & \mu_b'\\ \mu_b W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2} & \mu_b & Q_{xx}^{bb}\end{pmatrix}\bar{\gamma}_i\right\}^{1/2}}\right] \quad \text{and}$$

$$\eta^{(i)}(v_i)=\begin{bmatrix}0 & \text{if } v_i=0\\ \eta_1^{(i)}(v_i) & \text{if } v_i<0\\ \eta_2^{(i)}(v_i) & \text{if } v_i>0\end{bmatrix}$$

$$W_{xb}^{(i)}(v_i)=\begin{bmatrix}0 & \text{if } v_i=0\\ W_{xb,1}^{(i)}(v_i) & \text{if } v_i<0\\ W_{xb,2}^{(i)}(v_i) & \text{if } v_i>0\end{bmatrix}$$

$$\Pi_i=\bar{\gamma}_i'\begin{pmatrix}(\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_i^0)W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2} & (\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_i^0) & (\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_i^0)\mu_b'\\ W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2} & 1 & \mu_b'\\ \mu_b W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2} & \mu_b & Q_{xx}^{bb}\end{pmatrix}\bar{\gamma}_i$$

Box I.

- The vector $\xi_t=(u_t,u_{zt}^{f\prime},u_{zt}^{b\prime},u_{xt}^{f\prime},u_{xt}^{b\prime})'$, of dimension $n=q_f+p_f+q_b+p_b+1$, satisfies the following multivariate functional central limit theorem:

$$T^{-1/2}\sum_{t=1}^{[Tr]}\xi_t\Rightarrow B(r)$$

where $B(r)=(B_1(r),B_z^f(r)',B_z^b(r)',B_x^f(r)',B_x^b(r)')'$ is a n vector Brownian motion with symmetric covariance matrix

$$\Omega=\begin{pmatrix}\sigma^2 & \Omega_{1z}^f & \Omega_{1z}^b & \Omega_{1x}^f & \Omega_{1x}^b\\ \Omega_{z1}^f & \Omega_{zz}^{ff} & \Omega_{zz}^{fb} & \Omega_{zx}^{fb} & \Omega_{zx}^{ff}\\ \Omega_{z1}^b & \Omega_{zz}^{bf} & \Omega_{zz}^{bb} & \Omega_{zx}^{bf} & \Omega_{zx}^{bb}\\ \Omega_{x1}^f & \Omega_{xz}^{ff} & \Omega_{xz}^{fb} & \Omega_{xx}^{ff} & \Omega_{xx}^{fb}\\ \Omega_{x1}^b & \Omega_{xz}^{bf} & \Omega_{xz}^{bb} & \Omega_{xx}^{bf} & \Omega_{xx}^{bb}\end{pmatrix}\begin{matrix}1\\ q_f\\ q_b\\ p_f\\ p_b\end{matrix}$$

$$=\lim_{T\to\infty}T^{-1}E(S_TS_T')=\Sigma+\Lambda+\Lambda'$$

where $S_T=\sum_{t=1}^{T}\xi_t$, $\Sigma=\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}E(\xi_t\xi_t')$ and

$$\Lambda=\lim_{T\to\infty}T^{-1}\sum_{j=1}^{T-1}\sum_{t=1}^{T-j}E(\xi_t\xi_{t+j}').$$

with $\sigma^2>0$ and $p\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}u_t^2=\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}E[u_t^2]\equiv\sigma_u^2$. When the errors are stationary processes, we also have $\Omega=2\pi f_\xi(0)$ where $f_\xi(0)$ is the spectral density function of $\{\xi_t\}$ at zero frequency.
- $T^{-1/2}\sum_{t=1}^{[Tr]}(u_{xt}^f,u_{xt}^b)u_t\Rightarrow Q_{xu}^{1/2}W_x^*(r)$, where $W_x^*(r)=(W_{xf}^*(r)',W_{xb}^*(r)')'$ is a (p_f+p_b) vector of independent Wiener processes and

$$Q_{xu}=\left(\sum_{h=-\infty}^{\infty}E(u_{xt}u_t^2u_{xt-h}')\right)\equiv\begin{bmatrix}Q_{xu}^{ff} & Q_{xu}^{bf}\\ Q_{xu}^{fb} & Q_{xu}^{bb}\end{bmatrix}.$$

We shall also impose the following condition:

- Assumption A10: For all t and s: (a) $E(u_{xt}u_tz_s')=0$; (b) $E(u_{xt}u_tu_s)=0$; (c) $E(u_{xt}u_tu_{xs}')=0$.

Assumption A10 restricts somewhat the class of models applicable but is quite mild. Sufficient, though not necessary, conditions for it to hold are: for (a) that the $I(0)$ regressors are uncorrelated with the errors contemporaneously even conditional on the $I(1)$ variables; for (b) that the autocovariance structure of the $I(0)$ regressors be independent of the errors and, similarly, for (c) that the autocovariance structure of the errors be independent of the $I(0)$ regressors. This assumption is needed to guarantee that $W_x^*(\cdot)$ and $B(\cdot)$ are uncorrelated and, being Gaussian, are therefore independent. Without this condition, the analysis would be much more complex.

Finally, we need the following assumption which rules out cointegration among the $I(1)$ regressors:

- Assumption A11: $\begin{pmatrix}\Omega_{zz}^{bb} & \Omega_{zz}^{fb}\\ \Omega_{zz}^{bf} & \Omega_{zz}^{ff}\end{pmatrix}>0$.

Consistent estimates of the matrices Σ and Λ (and hence Ω) are $\hat{\Sigma}=T^{-1}\sum_{t=1}^{T}\hat{\xi}_t\hat{\xi}_t'$ and $\hat{\Lambda}=T^{-1}\sum_{j=1}^{T-1}w(j/l_T)\sum_{t=1}^{T-j}\hat{\xi}_t\hat{\xi}_{t+j}$, where $\hat{\xi}_t=(\hat{u}_t,\Delta z_{ft}',\Delta z_{bt}',(x_{ft}-\bar{x}_f)',(x_{bt}-\bar{x}_b)')'$ with $\hat{u}_t$ the OLS residuals from regression (1), $\bar{x}_i=T^{-1}\sum_{t=1}^{T}x_{it}$ $(i=f,b)$ and $w(j/l_T)$ is a kernel function that is continuous and even with $w(0)=1$ and $\int_{-\infty}^{\infty}k^2(x)\mathrm{d}x<\infty$. Also, $l_T\to\infty$ as $T\to\infty$ and $l_T=o(T^{1/2})$. Consistency of these covariance matrix estimates has been shown in Hansen (1992a). The following proposition states the limiting distribution of the estimates of the break dates under strict exogeneity of the $I(1)$ regressors.

Theorem 3. *For $i=1,\ldots,m$, let $\eta^{(i)}(v_i)$ and $W_{xb}^{(i)}(v_i)$ are two sided Wiener processes independent of each other. Also, $\eta_1^{(i)}(v_i)$ and $\eta_2^{(i)}(v_i)$ are independent, and $\eta_j^{(i)}(v_i)$ $(j=1,2)$ are independent across i. Similarly, $W_{xb,1}^{(i)}(v_i)$ and $W_{xb,2}^{(i)}(v_i)$ are independent, and $W_{xb,j}^{(i)}(v_i)$ $(j=1,2)$ are independent across i. Then, under Assumptions* A1–A11 *and* $\Omega_{1z}^f=\Omega_{1z}^b=0$,

$$\left(v_T^2\Pi_1(\hat{T}_1-T_1^0),\ldots,v_T^2\Pi_m(\hat{T}_m-T_m^0)\right)\Rightarrow\arg\max_{(v_1,\ldots,v_m)}H(v_1,\ldots,v_m)$$

with

$$H(v_1,\ldots,v_m)=B(v_1,\ldots,v_m)-\frac{1}{2}A(v_1,\ldots,v_m)$$

where $A(v_1,\ldots,v_m)$ and $B(v_1,\ldots,v_m)$ are given in Box I.

This theorem shows how different inference about the break dates is when the coefficient of an $I(1)$ regressor is allowed to change. In this case, the limit distribution of the estimates of the break dates are not asymptotically independent even if the break dates are separated by a positive fraction of the sample size. This can be seen by noting that the function $H(v_1, \ldots, v_m)$ involves the same Wiener processes (those corresponding to the $I(1)$ regressors) evaluated at $v_1, \ldots, v_m$. This gives rise to the correlation between the estimates of the break dates even in the limit. This result contrasts with the case of regression models with $I(0)$ regressors, in which case the limit distributions of the estimates of the break dates are asymptotically independent (see Bai and Perron (1998)). As shown in Corollary 3 below, this asymptotic independence continues to hold when $I(1)$ regressors are included but their coefficients are not allowed to change. Note that from a computational aspect, evaluating the limit distribution involves considering 2^m cases corresponding to the possible combinations of the signs of $v_1, \ldots, v_m$. In the single break case, the limit distribution is different from that in Bai et al. (1998) because we do not assume zero mean stationary regressors. Hence, the limit distribution is a function of μ_b, which can nevertheless be consistently estimated.

The cumulative distribution function of the random variable $\text{argmax}_{(v_1,\ldots,v_m)} H(v_1, \ldots, v_m)$ does not have a tractable analytical formula and hence needs to be obtained by simulations. Accordingly, we first generate a realization of $H(v_1, \ldots, v_m)$ by replacing the true value of the parameters with consistent estimates and simulating the Brownian motion processes over an appropriate range, say, $[-M, M]$. Then, we obtain the global maximum of the function $H(v_1, \ldots, v_m)$ over $(v_1, \ldots, v_m) \in [-M, M] \times \cdots \times [-M, M]$. This is repeated until we have an estimate of the joint distribution over an appropriate range. A standard method to construct joint confidence intervals is to use the so called Bonferroni procedure. In this case, we simulate the marginal distributions of $\hat{T}_1, \ldots, \hat{T}_m$ and form $(1 - \alpha/m)$% confidence intervals for each. The joint confidence interval at significance level α is then the intersection of the intervals for each of the m break dates (see, e.g., Gourieroux and Monfort (1995, p. 218)). Other methods of constructing joint confidence intervals are discussed in Lehmann and Romano (2005).

Often, special cases of the general regression model (1) are used. We classify them in two categories: (a) models involving only $I(1)$ regressors; (b) models involving both $I(1)$ and $I(0)$ regressors.

Category (a), *Models with* $I(1)$ *variables only* ($p_f = p_b = 0$, for all cases):

1. $c_j = 0$ for all $j = 1, \ldots, m+1$, $q_f = 0$: $y_t = z_{bt}'\delta_{bj} + u_t$;
2. $q_f = 0$: $y_t = c_j + z_{bt}'\delta_{bj} + u_t$;
3. $q_b = 0$: $y_t = c_j + z_{ft}'\delta_f + u_t$;
4. $c_j = c$ for all $j = 1, \ldots, m+1$, $q_f = 0$: $y_t = c + z_{bt}'\delta_{bj} + u_t$;
5. no restriction: $y_t = c_j + z_{ft}'\delta_f + z_{bt}'\delta_{bj} + u_t$;
6. $c_j = c$ for all $j = 1, \ldots, m+1$: $y_t = c + z_{ft}'\delta_f + z_{bt}'\delta_{bj} + u_t$.

Category (b), *Models with both* $I(1)$ *and* $I(0)$ *variables:*

1. $c_j = c$ for all $j = 1, \ldots, m+1$, $p_f = q_b = 0$: $y_t = c + z_{ft}'\delta_f + x_{bt}'\beta_{bj} + u_t$;
2. $c_j = c$ for all $j = 1, \ldots, m+1$, $p_b = q_f = 0$: $y_t = c + z_{bt}'\delta_{bj} + x_{ft}'\beta_f + u_t$;
3. $c_j = c$ for all $j = 1, \ldots, m+1$, $p_f = q_f = 0$: $y_t = c + z_{bt}'\delta_{bj} + x_{bt}'\beta_{bj} + u_t$;
4. $p_f = q_f = 0$: $y_t = c_j + z_{bt}'\delta_{bj} + x_{bt}'\beta_{bj} + u_t$;
5. $p_b = q_b = 0$: $y_t = c_j + z_{ft}'\delta_f + x_{ft}'\beta_f + u_t$;
6. $p_b = q_f = 0$: $y_t = c_j + z_{bt}'\delta_{bj} + x_{ft}'\beta_f + u_t$;
7. $p_f = q_b = 0$: $y_t = c_j + z_{ft}'\delta_f + x_{bt}'\beta_{bj} + u_t$;
8. $q_f = 0$: $y_t = c_j + z_{bt}'\delta_{bj} + x_{ft}'\beta_f + x_{bt}'\beta_{bj} + u_t$;
9. $q_b = 0$: $y_t = c_j + z_{ft}'\delta_f + x_{ft}'\beta_f + x_{bt}'\beta_{bj} + u_t$;

We now give a brief description of each of the models in the two categories. Consider first Category (a). Case 1 is a pure structural change model without an intercept in which all $I(1)$ coefficients are allowed to change across regimes. Case 2 is a pure structural change model which allows for a change in the intercept as well. Case 3 is a partial change model in which only the intercept is allowed to change. Case 4 is again a partial change model where the intercept is not allowed to change. Cases 5 and 6 are block partial models in which a subset of the $I(1)$ coefficients is allowed to change. In Category (b), Cases 1 to 3 are partial change models where the intercept is not allowed to change across regimes. Case 4 is a pure change model where all $I(1)$ and $I(0)$ coefficients as well as the intercept is allowed to change. Case 5 is a partial change model, which involves only an intercept shift. Case 6 is a partial change model where the $I(0)$ coefficients are not allowed to change. Similarly, Case 7 is a partial change model where the $I(1)$ coefficients are not allowed to change. Cases 8–9 are block partial models in which a subset of coefficients of at least one type of regressor is not allowed to change. The limit distribution stated in Theorem 3 simplifies according to particular subgroups of these special cases, as stated in the following Corollaries.

Corollary 1. *For Cases* (1), (4) *and* (6) *in Category* (a) *and Case* (2) *in Category* (b) *we have*

$$\Pi_i = \bar{\gamma}_i'(\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_i^0) W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2}\bar{\gamma}_i$$

$$B(v_1, \ldots, v_m) = \sum_{i=1}^{m}\left[\frac{\bar{\gamma}_i'(\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_i^0)\sigma\eta^{(i)}(v_i)}{\left\{\bar{\gamma}_i'(\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_i^0) W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2}\bar{\gamma}_i\right\}^{1/2}}\right].$$

The cases covered by this Corollary are those for which the constant is held fixed (or constrained to be zero) and only coefficients on $I(1)$ regressors are allowed to change.

Corollary 2. *For Cases* (2) *and* (5) *in Category* (a) *and Case* (6) *in Category* (b), *we have*

$$\Pi_i = \bar{\gamma}_i'\begin{pmatrix}(\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_i^0) W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2} & (\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_i^0) \\ W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2} & 1\end{pmatrix}\bar{\gamma}_i$$

$$B(v_1, \ldots, v_m) = \sum_{i=1}^{m}\left[\frac{\bar{\gamma}_i'\begin{pmatrix}(\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_i^0)\sigma\eta^{(i)}(v_i) \\ \sigma\eta^{(i)}(v_i)\end{pmatrix}}{\left\{\bar{\gamma}_i'\begin{pmatrix}(\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_i^0) W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2} & (\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_i^0) \\ W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2} & 1\end{pmatrix}\bar{\gamma}_i\right\}^{1/2}}\right].$$

The cases covered by this corollary are those for which the constant is allowed to change and none of the coefficients on the $I(0)$ regressors is allowed to change. The combination of Corollaries 1 and 2 show that the limit distribution of the break dates is different whether the constant is allowed to change or not, when coefficients on $I(1)$ regressors change. This is in contrast to the stationary case where having a pure or partial structural change model implies the same limit distribution.

Corollary 3. *For Case* (3) *in Category* (a) *and Cases* 1, 5 *and* 9 *in Category* (b), *we have*

$$\Pi_i = \bar{\gamma}_i'\bar{\gamma}_i$$

$$B(v_1, \ldots, v_m) = \sum_{i=1}^{m}\left[\frac{\bar{\gamma}_i'\sigma\eta^{(i)}(v_i)}{\left\{\bar{\gamma}_i'\bar{\gamma}_i\right\}^{1/2}}\right].$$

$$\Pi_i = \bar{\gamma}_i' \begin{pmatrix} (\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_i^0) W_z^b(\lambda_i^0)' (\Omega_{zz}^{bb})^{1/2} & (\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_i^0)\mu_b' \\ \mu_b W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2} & Q_{xx}^{bb} \end{pmatrix} \bar{\gamma}_i$$

$$B(v_1,\ldots,v_m) = \sum_{i=1}^{m} \left[\frac{\bar{\gamma}_i' \begin{pmatrix} (\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_i^0)\sigma\eta^{(i)}(v_i) \\ (Q_{xu}^{bb})^{1/2} W_{xb}^{(i)}(v_i) \end{pmatrix}}{\left\{ \bar{\gamma}_i' \begin{pmatrix} (\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_i^0) W_z^b(\lambda_i^0)' (\Omega_{zz}^{bb})^{1/2} & (\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_i^0)\mu_b' \\ \mu_b W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2} & Q_{xx}^{bb} \end{pmatrix} \bar{\gamma}_i \right\}^{1/2}} \right]$$

Box II.

The cases covered by this Corollary are those for which no coefficient associated with $I(1)$ regressors are allowed to change. Note that in this case, the limits do not involve the Wiener processes corresponding to the $I(1)$ regressors. This is because the coefficients of the $I(1)$ regressors are not allowed to change and, hence, they do not matter asymptotically as far as the limit distribution of the break dates are concerned. Here, the estimates of the break dates are therefore asymptotically independent and we have, for $i = 1,\ldots,m$,

$$\frac{\bar{\gamma}_i'^2}{\sigma^2} v_T^2(\hat{T}_i - T_i^0) \Rightarrow \arg\max_v \left\{ \eta^{(i)}(v) - \frac{|v|}{2} \right\}$$

where

$$\eta^{(i)}(v) = \begin{cases} \eta_1^{(i)}(v) & \text{if } v < 0 \\ \eta_2^{(i)}(v) & \text{if } v > 0 \end{cases}$$

which reduces to the limit distribution stated in Bai and Perron (1998) for the stationary case.

Corollary 4. *For Case* 3 *in Category* (b), *we have in* Box II.

Here coefficients on the $I(1)$ and $I(0)$ regressors are allowed to change but the constant is not. The following Corollary shows how allowing the constant to vary or not affects the limit distribution when both $I(1)$ and $I(0)$ regressors are allowed to change.

Corollary 5. *The limit distribution for Cases* 4 *and* 8 *in Category* (b) *is the same as that for the general case given by regression (1).*

Here all coefficients associated with $I(1)$ variables are allowed to change as well as the constant and the coefficients on some $I(0)$ regressors. When combined with the other results, this points to the fact that, when at least one coefficient associated with an $I(1)$ regressor is allowed to change, what influences the limit distribution are: (1) the number of $I(1)$ regressors whose coefficients are allowed to change; (2) the number of $I(0)$ regressors whose coefficients are allowed to change, and (3) whether the constant is allowed to change. Of particular interest is the fact that the limit distribution obtained holding a subset of the coefficients fixed is the same that prevails when not including these regressors. This is different from what occurs when doing hypothesis testing about whether breaks are present or not (see, Kejriwal and Perron (2008)). In this case, including regressors whose coefficients are not allowed to change affects the limit distribution of the tests.

3.4. Extension to the dynamic OLS regression model

To deal with the possibility of endogenous $I(1)$ regressors, a popular method is to use the so-called dynamic OLS regression where leads and lags of the first-differences of the $I(1)$ variables are added as regressors, as suggested by Saikkonen (1991) and Stock and Watson (1993). The regression model is then

$$y_t = \hat{c}_i + z_{ft}'\hat{\delta}_f + x_{ft}'\hat{\beta}_f + z_{bt}'\hat{\delta}_{bi} + x_{bt}'\hat{\beta}_{bi} + \sum_{j=-k_T}^{k_T} \Delta z_{t-j}'\hat{\Pi}_j + \hat{v}_t^* \quad (t = T_{j-1}+1,\ldots,T_j) \tag{8}$$

for $j = 1,\ldots,m+1$, where $v_t^* = v_t + \sum_{|j|>k_T} \zeta_{z,t-j}'\Pi_j \equiv v_t + e_t$. Note that since these additional regressors are introduced to modify the limit distribution of the estimates of the main parameters of the models, the coefficients associated with the leads and lags of the first-differenced $I(1)$ variables are held fixed across regimes.[3] It is well known that to obtain estimates that are asymptotically optimal and test statistics that have the usual chi-square limit distribution, k_T must increase at some rate as T increases. This makes the problem different from that discussed so far, since in the previous sections the number of regressors is held fixed as the sample size increases. The aim of this section is to show that, under some conditions, the results discussed so far remain valid in this context. To establish this, we need the following assumption on the rate of increase of k_T.

- Assumption A11: As $T \to \infty$, $k_T \to \infty$, $k_T^2/T \to 0$, $k_T \sum_{|j|>k_T} \|\Pi_j\| \to 0$, and $k_T^{1/2} v_T \to \infty$.

Except for the last condition, this assumption is the same as used in Kejriwal and Perron (in press) who showed that the results stated in Saikkonen (1991) under more restrictive conditions remain valid. For instance, if the Π_j are eventually exponentially decaying (as in the case of an ARMA process), the lower bound condition $k_T \sum_{|j|>k_T} \|\Pi_j\| \to 0$ permits a logarithmic rate of increase for k_T so that a data dependent rule based on an information criterion can be used to select the number of leads and lags. Given the fact that the magnitude of the breaks is decreasing as T increases, here things are more complex and there is a need for an additional lower bound condition given by the requirement that $k_T^{1/2} v_T \to \infty$. The relevant result is stated in the following Theorem.

Theorem 4. *Suppose that the break dates are obtained by minimizing the sum of squared residuals from regression (8), then under Assumptions* A1–A11, *Theorems* 1–3 *remain valid.*

Remark 3. Bai et al. (1998) assume that $T^{1/2}v_T/\log T \to \infty$. Suppose that k_T is selected using data dependent rules based on information criteria such as the AIC or BIC so that $k_T = O(\log T)$. Then we need $(\log T)^{1/2} v_T \to \infty$. We can show that the latter condition implies $T^{1/2}v_T/\log T \to \infty$, but not vice-versa. Hence, their assumption is not sufficient to guarantee that the same limit distribution applies when adding an increasing sequence of leads and leads of the first-differenced $I(1)$ regressors.

4. Simulation experiments

In this section, we conduct simulation experiments to assess the adequacy of the asymptotic distribution as an approximation to the finite sample distributions. We investigate the finite sample coverage rate of asymptotic confidence intervals based on the

[3] Note that the number of leads and lags of Δz_t need not be the same. We specify the same value for simplicity. Alternatively, one can interpret k_T as the maximum of the number of leads and lags.

Table 1
Coverage rates for the single break case

Model	Values	S1				S2				S3				S4			
		$T = 120$		$T = 240$		$T = 120$		$T = 240$		$T = 120$		$T = 240$		$T = 120$		$T = 240$	
		80%	90%	80%	90%	80%	90%	80%	90%	80%	90%	80%	90%	80%	90%	80%	90%
	$c_2 = 2, \delta_2 = 2, \lambda^0 = .50$	.92	.94	.95	.97	.87	.90	.90	.92	.91	.93	.95	.97	.93	.94	.95	.96
	$c_2 = 2, \delta_2 = 1.5, \lambda^0 = .50$	.81	.87	.90	.93	.78	.82	.85	.88	.79	.86	.88	.92	.86	.89	.91	.92
A	$c_2 = 2, \delta_2 = 2, \lambda^0 = .25$	.87	.91	.91	.93	.83	.86	.88	.90	.83	.86	.91	.93	.91	.92	.91	.92
	$c_2 = 2, \delta_2 = 1.5, \lambda^0 = .25$	.78	.84	.85	.90	.77	.83	.80	.85	.75	.83	.85	.91	.85	.88	.88	.90
	$c_2 = 2, \delta_2 = 2, \lambda^0 = .75$	.92	.94	.96	.97	.89	.92	.93	.94	.92	.95	.96	.97	.95	.96	.96	.97
	$c_2 = 2, \delta_2 = 1.5, \lambda^0 = .75$	.85	.89	.92	.95	.82	.87	.87	.89	.83	.87	.91	.94	.87	.90	.94	.95
	$c_2 = 2.5, \lambda^0 = .50$	.86	.91	.87	.93	.91	.97	.90	.96	.81	.86	.86	.92	.80	.86	.85	.90
	$c_2 = 2, \lambda^0 = .50$	.76	.84	.77	.86	.88	.93	.89	.96	.69	.76	.75	.83	.67	.77	.76	.83
B	$c_2 = 2.5, \lambda^0 = .25$	.83	.88	.86	.92	.92	.96	.89	.96	.79	.85	.84	.90	.79	.85	.86	.92
	$c_2 = 2, \lambda^0 = .25$	.74	.83	.80	.87	.84	.92	.89	.95	.66	.75	.77	.84	.68	.76	.77	.84
	$c_2 = 2.5, \lambda^0 = .75$	.81	.87	.85	.91	.88	.95	.89	.95	.72	.79	.83	.89	.73	.79	.83	.88
	$c_2 = 2, \lambda^0 = .75$	.70	.78	.75	.83	.84	.91	.87	.95	.58	.67	.72	.80	.59	.67	.72	.79
	$\delta_2 = 2, \lambda^0 = .50$	.92	.94	.95	.96	.88	.91	.91	.93	.92	.94	.94	.96	.93	.94	.95	.96
	$\delta_2 = 1.5, \lambda^0 = .50$	.84	.88	.87	.91	.78	.84	.83	.86	.81	.86	.89	.92	.87	.89	.91	.93
C	$\delta_2 = 2, \lambda^0 = .25$	.88	.92	.93	.95	.82	.85	.88	.90	.86	.90	.92	.95	.90	.91	.92	.93
	$\delta_2 = 1.5, \lambda^0 = .25$	.79	.85	.86	.91	.75	.83	.79	.84	.77	.83	.84	.90	.81	.84	.88	.90
	$\delta_2 = 2, \lambda^0 = .75$	.93	.95	.96	.97	.90	.91	.92	.94	.92	.94	.96	.97	.94	.95	.96	.97
	$\delta_2 = 1.5, \lambda^0 = .75$	.86	.90	.92	.95	.82	.86	.87	.90	.85	.89	.91	.94	.89	.91	.92	.94
	$c_2 = 2, \delta_2 = 2, \lambda^0 = .50$	.91	.94	.94	.95	.88	.90	.92	.94	.91	.93	.94	.96	.93	.94	.96	.97
	$c_2 = 2, \delta_2 = 1.5, \lambda^0 = .50$	.81	.86	.90	.93	.79	.83	.84	.88	.82	.87	.87	.90	.87	.90	.92	.93
D	$c_2 = 2, \delta_2 = 2, \lambda^0 = .25$	.88	.92	.92	.94	.85	.88	.87	.90	.85	.89	.91	.93	.92	.93	.94	.96
	$c_2 = 2, \delta_2 = 1.5, \lambda^0 = .25$	.76	.83	.85	.90	.72	.77	.78	.83	.73	.81	.83	.88	.85	.88	.87	.90
	$c_2 = 2, \delta_2 = 2, \lambda^0 = .75$	.93	.95	.96	.97	.88	.90	.93	.94	.93	.95	.94	.96	.94	.95	.96	.97
	$c_2 = 2, \delta_2 = 1.5, \lambda^0 = .75$	.86	.89	.90	.93	.78	.84	.87	.91	.83	.88	.90	.93	.88	.90	.92	.93

limiting distribution for cases where the Data Generating Process (DGP) involves one and two breaks. Our choice of data generating processes enables comparisons with the results obtained by Bai et al. (1998) for the single break case. The number of replications used is 1000. The level of trimming is set at $\epsilon = 0.15$. The sample sizes used are $T = 120$ and $T = 240$. Throughout, we consider a single $I(1)$ regressor z_t generated by:

$$z_t = z_{t-1} + \eta_t$$

and an $I(0)$ regressor $x_t \sim i.i.d.\ N(1, 1)$. Here $\eta_t \sim i.i.d.\ N(0, 1)$ and independent of x_t. Also, let $e_t \sim i.i.d.\ N(0, 1)$ where $\mathrm{Cov}(\eta_t, e_t) = \mathrm{Cov}(x_t, e_t) = 0$.

4.1. The case with one break

We consider four models: the first involving shifts in both the intercept and the cointegrating coefficient, the second involving a shift in the intercept only, the third involving only a shift in the cointegrating coefficient and the fourth which is the same as the first except that x_t is also included in the regression but its coefficient is not allowed to change. The models considered are the following:

- Model-A (Change in intercept and slope):
  $$y_t = c_1 + \delta_1 z_t + u_t \quad \text{if } t \le \left[T\lambda^0\right]$$
  $$y_t = c_2 + \delta_2 z_t + u_t \quad \text{if } t > \left[T\lambda^0\right].$$
- Model-B (Change in intercept only):
  $$y_t = c_1 + z_t + u_t \quad \text{if } t \le \left[T\lambda^0\right]$$
  $$y_t = c_2 + z_t + u_t \quad \text{if } t > \left[T\lambda^0\right].$$
- Model-C (Change in slope only):
  $$y_t = 1 + \delta_1 z_t + u_t \quad \text{if } t \le \left[T\lambda^0\right]$$
  $$y_t = 1 + \delta_2 z_t + u_t \quad \text{if } t > \left[T\lambda^0\right].$$
- Model-D: Same as Model-A except that the (irrelevant) $I(0)$ regressor x_t, whose coefficient is not allowed to change, is also included.

For each model, we consider the following four specifications for u_t: S1 (*i.i.d.* errors, exogenous regressor): $u_t = e_t$; S2 (MA(1) errors, exogenous regressor): $u_t = e_t - 0.5e_{t-1}$; S3 (*i.i.d.* errors, endogenous regressor): $u_t = 0.5\eta_t + e_t$; S4 (MA(1) errors, endogenous regressor): $u_t = 0.5v_t + \eta_t,\ v_t = e_t - 0.5e_{t-1}$. For S2, we correct for serial correlation by constructing the long-run variance estimator based on the Quadratic Spectral kernel and an AR(1) approximation for the bandwidth. For S3, we use the dynamic OLS estimator discussed in Section 3.4 with $k_T = 2$. Finally, for S4, we use the correction for serial correlation as in S2 and the endogeneity correction as in S3.

We set $c_1 = 1$ and $\delta_1 = 1$. Table 1 presents finite sample coverage rates of 80% and 90% asymptotic confidence intervals. For model A, when the magnitude of the break in the slope is large, the confidence intervals are conservative, irrespective of the sample size. For a smaller change in slope, however, the coverage rates are usually lower for $T = 120$ but become conservative when the sample size is doubled. A similar picture applies for model B except that here the coverage rates are usually lower than the corresponding asymptotic confidence levels when the break is small and occurs at the beginning or the end of the sample. For model C, the coverage rates are adequate for small breaks and $T = 120$. With larger breaks or with $T = 240$, they are somewhat conservative. The results for model D are qualitatively similar to those for model A. The simulations presented in Bai et al. (1998) show that the confidence intervals are too tight for most of the cases considered. This is not the case in our simulations. Indeed, our results show the asymptotic approximation to be more accurate than reported in Bai et al. (1998). Overall, the coverage rates are reasonably accurate or somewhat conservative.

4.2. The case with two breaks

With two breaks the models considered are the following:

- Model-E (Change in Intercept and Slope)
  $$y_t = c_1 + \delta_1 z_t + u_t \quad \text{if } t \le [T\lambda_1^0]$$

Table 2
Coverage rates for the two breaks case ($\lambda_1^0 = 1/3, \lambda_2^0 = 2/3$)

Model	Values	S1				S2				S3				S4			
		$T = 120$		$T = 240$		$T = 120$		$T = 240$		$T = 120$		$T = 240$		$T = 120$		$T = 240$	
		80%	90%	80%	90%	80%	90%	80%	90%	80%	90%	80%	90%	80%	90%	80%	90%
	$c_2 = 1.5, c_3 = 2, \delta_2 = 1.5, \delta_3 = 2$	.72	.75	.81	.83	.78	.79	.85	.87	.67	.69	.79	.81	.70	.72	.80	.82
E	$c_2 = 2, c_3 = 3, \delta_2 = 2, \delta_3 = 3$	.89	.91	.92	.94	.90	.91	.92	.93	.84	.86	.92	.93	.87	.88	.90	.91
	$c_2 = 2, c_3 = 3, \delta_2 = 1.5, \delta_3 = 2$	.74	.77	.82	.84	.77	.79	.83	.84	.66	.70	.79	.81	.71	.74	.79	.81
F	$\delta_2 = 1.5, \delta_3 = 2$	.79	.82	.83	.85	.82	.84	.86	.88	.79	.82	.86	.88	.89	.90	.93	.94
	$\delta_2 = 2, \delta_3 = 3$	.90	.92	.93	.94	.90	.91	.93	.94	.91	.92	.92	.94	.95	.96	.97	.98
	$c_2 = 1.5, c_3 = 2, \delta_2 = 1.5, \delta_3 = 2$	.73	.76	.83	.85	.80	.82	.85	.86	.69	.73	.80	.83	.63	.67	.81	.82
G	$c_2 = 2, c_3 = 3, \delta_2 = 2, \delta_3 = 3$	.89	.91	.92	.93	.92	.93	.93	.94	.86	.88	.91	.92	.83	.86	.93	.94
	$c_2 = 2, c_3 = 3, \delta_2 = 1.5, \delta_3 = 2$	.71	.74	.81	.84	.79	.81	.86	.88	.70	.72	.79	.81	.65	.68	.82	.84

Table 3
Coverage rates for the two breaks case ($\lambda_1^0 = 1/4, \lambda_2^0 = 3/4$)

Model	Values	S1				S2				S3				S4			
		$T = 120$		$T = 240$		$T = 120$		$T = 240$		$T = 120$		$T = 240$		$T = 120$		$T = 240$	
		80%	90%	80%	90%	80%	90%	80%	90%	80%	90%	80%	90%	80%	90%	80%	90%
	$c_2 = 1.5, c_3 = 2, \delta_2 = 1.5, \delta_3 = 2$	.72	.76	.82	.84	.76	.78	.84	.86	.65	.69	.77	.81	.64	.67	.80	.83
E	$c_2 = 2, c_3 = 3, \delta_2 = 2, \delta_3 = 3$	.88	.90	.92	.93	.87	.88	.93	.94	.86	.87	.90	.92	.85	.87	.91	.92
	$c_2 = 2, c_3 = 3, \delta_2 = 1.5, \delta_3 = 2$	.76	.78	.82	.85	.78	.80	.86	.87	.69	.71	.78	.80	.67	.70	.81	.83
F	$\delta_2 = 1.5, \delta_3 = 2$	.76	.79	.84	.86	.77	.81	.87	.88	.77	.82	.86	.87	.88	.89	.93	.94
	$\delta_2 = 2, \delta_3 = 3$	.90	.92	.93	.94	.89	.90	.93	.94	.89	.90	.93	.95	.95	.96	.97	.98
	$c_2 = 1.5, c_3 = 2, \delta_2 = 1.5, \delta_3 = 2$	.72	.75	.83	.86	.76	.79	.82	.85	.69	.71	.81	.83	.67	.70	.80	.82
G	$c_2 = 2, c_3 = 3, \delta_2 = 2, \delta_3 = 3$	.90	.91	.93	.94	.89	.90	.93	.94	.86	.88	.91	.93	.85	.87	.91	.92
	$c_2 = 2, c_3 = 3, \delta_2 = 1.5, \delta_3 = 2$	.74	.77	.83	.85	.76	.78	.84	.86	.68	.71	.82	.84	.68	.72	.79	.82

$$y_t = c_2 + \delta_2 z_t + u_t \quad \text{if } [T\lambda_1^0] < t \le [T\lambda_2^0]$$
$$y_t = c_3 + \delta_3 z_t + u_t \quad \text{if } [T\lambda_2^0] < t \le T.$$

- Model-F (Change in Slope only)

$$y_t = 1 + \delta_1 z_t + u_t \quad \text{if } t \le [T\lambda_1^0]$$
$$y_t = 1 + \delta_2 z_t + u_t \quad \text{if } [T\lambda_1^0] < t \le [T\lambda_2^0]$$
$$y_t = 1 + \delta_3 z_t + u_t \quad \text{if } [T\lambda_2^0] < t \le T.$$

- Model-G: Same as Model-E except that the (irrelevant) $I(0)$ regressor x_t, whose coefficient is not allowed to change, is also included.

Again, we set $c_1 = 1$ and $\delta_1 = 1$. As in the one break case, we consider the error specifications S1–S4 with the corresponding corrections for serial correlation and/or endogeneity. The confidence intervals are constructed jointly using the Bonferroni procedure discussed in Section 3. The coverage rates are presented in Tables 2 and 3. Consider first Model E. When the change in slope is small, the coverage rates are inadequate; however, the confidence intervals become conservative as the magnitude of the change increases. An interesting feature is that the coverage rates remain almost unaffected when the magnitude of the intercept change increases but the magnitude of the slope change remains the same. For Model F, the confidence intervals are again conservative provided the magnitude of the breaks is large. Again, the results for model G are similar to those for model E. For all models, the coverage rates increase when the sample size increases. Note that, given the use of an asymptotic framework with shrinking shifts, the accuracy of the approximations need not improve as the sample size increases when the magnitudes of the breaks are fixed. Overall, the coverage rates are reasonably accurate and, as expected, somewhat conservative with larger sample sizes. Hence, from these limited experiments, we can conclude that the limiting distributions derived provide useful approximations in finite samples.

5. Conclusion

This paper has presented a comprehensive treatment of issues related to estimation and inference in cointegrated regression models with multiple structural breaks. We analyzed models with $I(1)$ variables only as well as models which incorporate both $I(0)$ and $I(1)$ regressors. The breaks are allowed to occur in the intercept, the cointegrating coefficients, the parameters of the $I(0)$ regressors or any combination of these. The results show that confidence intervals for the break dates need to be constructed jointly whenever the coefficients associated with the $I(1)$ regressors are allowed to change even if the break dates are separated by a positive fraction of the sample size. A comparison of various methods to construct such confidence intervals in terms of their performance in finite samples is an important avenue for future research.

Appendix

We use $\|.\|$ to denote the Euclidean norm, i.e. $\|x\| = (\sum_{i=1}^{p} x_i^2)^{1/2}$ for $x \epsilon R^p$. For a matrix A, we use the vector-induced norm, i.e. $\|A\| = \sup_{x \neq 0} \|Ax\| / \|x\|$. We have $\|A\| \le \left[tr(A'A)\right]^{1/2}$. Also, for a projection matrix P, $\|PA\| \le \|A\|$. $W_1, W_z^f, W_z^b, W_x^f, W_x^b$ are independent Wiener processes with dimensions corresponding to those of $B_1, B_z^f, B_z^b, B_x^f, B_x^b$. We also define the matrices $D_{1T} = \text{diag}(T^{-1}I_{q_f}, T^{-1/2}I_{p_f})$, $D_{2T} = \text{diag}(T^{-1}I_{q_b}, T^{-1/2}I_{p_b})$ and $D_{3T} = \text{diag}(T^{-1/2}I_{q_b}, I_{p_b})$. Henceforth, we will refer to Bai and Perron (1998) as [BP]. We first state a series of lemmas which will be used subsequently.

Lemma A.1 (Qu and Perron, 2007). *Let $(\eta_i)_{i \ge 1}$ be a sequence of mean zero R^d-valued random vectors. Define $\mathcal{F}_k$ as an increasing σ-field generated by $(\eta_i)_{i \le k}$. Suppose $(\eta_i)_{i \ge 1}$ satisfies Assumption* A4. *We have* (a) *(Generalized Hajek–Renyi inequality) there exists an $L < \infty$*

such that, for every $c > 0$ *and* $k_0 > 0$, $P(\sup_{k \geq k_0} k^{-1} \| \sum_{i=1}^{k} \eta_i \| > c) \leq (L/c^2 k_0)$; (b) *(FCLT)* $T^{-1/2} \sum_{t=1}^{[Tr]} \eta_t \Rightarrow \Omega W^*(r)$ *where* $W^*(r)$ *is a d-vector of independent Wiener processes and* $\Rightarrow$ *denotes weak convergence under the Skorohod topology;* (c) *(SLLN)* $k^{-1} \sum_{i=1}^{k} \eta_i \rightarrow^{a.s.} 0$ *as* $k \to \infty$.

The following Lemma is a direct consequence of Lemma A.1(b) applied to ξ_t.

Lemma A.2. *Under* A4, *we have uniformly over all* $0 < r < s < 1$: (a) $\sum_{t=[Tr]}^{[Ts]} \xi_t = O_p(T^{1/2})$, (b) $\sum_{t=[Tr]}^{[Ts]} z_t = O_p(T^{3/2})$, (c) $\sum_{t=[Tr]}^{[Ts]} z_t z_t' = O_p(T^2)$, (d) $\sum_{t=[Tr]}^{[Ts]} z_t \xi_t' = O_p(T)$.

Lemma A.3. *Under* A1, $\sup_{T_1,\ldots,T_m} (D_{1T} G' M_{\bar{W}} G D_{1T})^{-1} = O_p(1)$, *where the supremum is taken over all possible partitions such that* $|T_i - T_{i-1}| \geq q_b + p_b$ $(i = 1, \ldots, m+1)$.

Proof. We have the identity $G' M_{\bar{W}} G = G_1' M_{W_1} G_1 + \cdots + G_{m+1}' M_{W_{m+1}} G_{m+1}$. Each partition leaves at least one true regime intact. That is, there exists an i such that (G_i, W_i) contains (G_i^0, W_i^0) as a sub-matrix. Since $G_i' M_{W_i} G_i \geq G_i^{0\prime} M_{W_i^0} G_i^0$ (using Lemma A.1 of [BP]), we have $(D_{1T} G' M_{\bar{W}} G D_{1T})^{-1} \leq (D_{1T} G_i^{0\prime} M_{W_i^0} G_i^0 D_{1T})^{-1}$. This implies $\left\|(D_{1T} G' M_{\bar{W}} G D_{1T})^{-1}\right\| \leq \max_i \|(D_{1T} G_i^{0\prime} M_{W_i^0} G_i^0 D_{1T})^{-1}\|$ for all partitions. The result follows from Assumption A1. □

Lemma A.4. *Under* A1, $\sup_{T_1,\ldots,T_m} D_{1T} G' M_{\bar{W}} \bar{W}^0 = O_p(T)$.

Proof. Since $M_{\bar{Z}}$ is a projection matrix, we have $\left\|D_{1T} G' M_{\bar{W}} \bar{W}^0\right\| \leq \left\|D_{1T} G'\right\| \left\|\bar{W}^0\right\|$ uniformly over all partitions. The result then follows from the facts that $\left\|D_{1T} G'\right\| = O_p(1)$ and $\|\bar{W}^0\| = O_p(T)$, from Lemma A.2. □

Lemma A.5. *Under* A4 *and* A8, $\sup_{T_1,\ldots,T_m} \left\|P_{\bar{W}} U\right\| = O_p(1)$.

Proof. We shall prove $\left|U' P_{\bar{W}} U\right| = O_p(1)$ uniformly in $T_1, \ldots, T_m$. Note that $U' P_{\bar{W}} U$ is the summation of $m+1$ terms $(\sum_{t=T_i+1}^{T_{i+1}} w_t u_t)'$ $(\sum_{t=T_i+1}^{T_{i+1}} w_t w_t')^{-1} (\sum_{t=T_i+1}^{T_{i+1}} w_t u_t)$, for $i = 0, \ldots, m$. From Lemma A.2, $D_{2T} (\sum_{t=T_i+1}^{T_{i+1}} w_t w_t') D_{2T} = O_p(1)$. Also, $D_{2T} \sum_{t=T_i+1}^{T_{i+1}} w_t u_t = O_p(1)$. Accordingly, $\left|U' P_{\bar{W}} U\right| = O_p(1)$ uniformly in $T_1, \ldots, T_m$. □

Lemma A.6. *Under* A1–A4, (a) $\sup_{T_1,\ldots,T_m} D_{1T} G' P_{\bar{W}} U = O_p(1)$; (b) $\sup_{T_1,\ldots,T_m} \bar{W}^0 P_{\bar{W}} U = O_p(T)$.

Proof. This follows from Lemma A.4, $\left\|D_{1T} G'\right\| = O_p(T)$ and $\left\|D_{1T} G' P_{\bar{W}} U\right\| \leq \left\|D_{1T} G'\right\| \left\|P_{\bar{W}} U\right\|$. Similar arguments can be used to prove part (b). □

Proof of Lemma 1. We have $\sum_{t=1}^{T} u_t d_t = U' G(\hat{\alpha} - \alpha^0) + U' \bar{W} \hat{\gamma} - U' \bar{W}^0 \gamma^0$ where $\bar{W}$ is the diagonal partition of W at some arbitrary partition $(T_1, \ldots, T_m)$. We have

$$D_{1T}^{-1} \left(\hat{\alpha}(\{T_j\}) - \alpha^0\right) = (D_{1T} G' M_{\bar{W}} G D_{1T})^{-1} D_{1T} G' M_{\bar{W}} \bar{W}^0 \gamma^0 + (D_{1T} G' M_{\bar{W}} G D_{1T})^{-1} D_{1T} G' M_{\bar{W}} U. \quad (A.1)$$

Now $D_{1T} G' M_{\bar{W}} U = D_{1T} G' U - D_{1T} G' P_{\bar{W}} U$. Since $D_{1T} G' U = O_p(1)$ and $\left\|D_{1T} G' P_{\bar{W}} U\right\| \leq \left\|D_{1T} G'\right\| \left\|P_{\bar{W}} U\right\| = O_p(1)$ so that the second term of (A.1) is $O_p(1)$. From $\gamma_{T,i+1}^0 - \gamma_{T,i}^0 = D_{3T} O(v_T)$ under Assumption A7, we have $\gamma_{T,i}^0 - \gamma_{T,j}^0 = D_{3T} O(v_T)$ for all i and j. As in [BP], we use the fact that $(\bar{W}^0 - \bar{W})\gamma^0$ depends on changes in the parameters, i.e., $\gamma_i^0 - \gamma_j^0$. This implies that $D_{1T} G' M_{\bar{W}} (\bar{W}^0 - \bar{W})\gamma^0$ is at most $O_p(T^{1/2}) v_T$. By Lemma A.3,

$$(D_{1T} G' M_{\bar{W}} G D_{1T})^{-1} D_{1T} G' M_{\bar{W}} (\bar{W}^0 - \bar{W})\gamma^0 = O_p(T^{1/2} v_T)$$

which implies

$$D_{1T}^{-1}(\hat{\alpha}(\{T_j\}) - \alpha^0) = O_p(T^{1/2} v_T) + O_p(1).$$

Thus,

$$U' G(\hat{\alpha}(\{T_j\}) - \alpha^0) = O_p(T^{1/2} v_T) + O_p(1) \quad (A.2)$$

over all partitions. Next, we have

$$U' \bar{W} \hat{\gamma}(\{T_j\}) - U' \bar{W}^0 \gamma^0 = U' \bar{W} (\bar{W}' M_G \bar{W})^{-1} \bar{W}' M_G \bar{W}^0 \gamma^0 + U' \bar{W} (\bar{W}' M_G \bar{W})^{-1} \bar{W}' M_G U - U' \bar{W}^0 \gamma^0. \quad (A.3)$$

Combine the first and third terms of (A.3) and rewrite them as

$$U' \bar{W} (\bar{W}' M_G \bar{W})^{-1} \bar{W}' M_G (\bar{W}^0 - \bar{W})\gamma^0 + U'(\bar{W} - \bar{W}^0)\gamma^0 \quad (A.4)$$

which can be shown to be $O_p(T^{1/2} v_T)$. Using Lemmas A.3, A.5 and A.6 the second term of (A.3) can be shown to be $O_p(1)$. Thus,

$$U' \bar{W} \hat{\gamma}(\{T_j\}) - U' \bar{W}^0 \gamma^0 = O_p(T^{1/2} v_T) + O_p(1). \quad (A.5)$$

Thus, from (A.2) and (A.5), we get $\sum_{t=1}^{T} u_t d_t = O_p(T^{1/2} v_T)$. □

Proof of Lemma 2. Following [BP], we have

$$\sum_{t=1}^{T} d_t^2 \geq \sum_1 d_t^2 + \sum_2 d_t^2$$
$$= \begin{pmatrix} \hat{\alpha} - \alpha^0 \\ \hat{\gamma}_k - \gamma_j^0 \end{pmatrix}' \begin{pmatrix} \sum_1 g_t g_t' & \sum_1 g_t w_t' \\ \sum_1 w_t g_t' & \sum_1 w_t w_t' \end{pmatrix} \begin{pmatrix} \hat{\alpha} - \alpha^0 \\ \hat{\gamma}_k - \gamma_j^0 \end{pmatrix} + \begin{pmatrix} \hat{\alpha} - \alpha^0 \\ \hat{\gamma}_k - \gamma_{j+1}^0 \end{pmatrix}' \begin{pmatrix} \sum_2 g_t g_t' & \sum_2 g_t w_t' \\ \sum_2 w_t g_t' & \sum_2 w_t w_t' \end{pmatrix} \begin{pmatrix} \hat{\alpha} - \alpha^0 \\ \hat{\gamma}_k - \gamma_{j+1}^0 \end{pmatrix}$$

where $\sum_1$ extends over the set $T(\lambda_j^0 - \eta) \leq t \leq T\lambda_j^0$ and $\sum_2$ extends over the set $T\lambda_j^0 + 1 \leq t \leq T(\lambda_j^0 + \eta)$ and $\eta > 0$. Then, with $D_{4T} = \text{diag}((T\eta)^{-1} I_{q_f}, (T\eta)^{-1/2} I_{p_f}, (T\eta)^{-1} I_{q_b}, (T\eta)^{-1/2} I_{p_b})$,

$$\sum_{t=1}^{T} d_t^2 \geq \begin{pmatrix} \hat{\alpha} - \alpha^0 \\ \hat{\gamma}_k - \gamma_j^0 \end{pmatrix}' D_{4T}^{-1} A_T^* D_{4T}^{-1} \begin{pmatrix} \hat{\alpha} - \alpha^0 \\ \hat{\gamma}_k - \gamma_j^0 \end{pmatrix} + \begin{pmatrix} \hat{\alpha} - \alpha^0 \\ \hat{\gamma}_k - \gamma_{j+1}^0 \end{pmatrix}' D_{4T}^{-1} A_T D_{4T}^{-1} \begin{pmatrix} \hat{\alpha} - \alpha^0 \\ \hat{\gamma}_k - \gamma_{j+1}^0 \end{pmatrix} + o_p(1) \quad (A.6)$$

where

$$A_T^* = D_{4T} \begin{pmatrix} \sum_1 g_t g_t' & \sum_1 g_t w_t' \\ \sum_1 w_t g_t' & \sum_1 w_t w_t' \end{pmatrix} D_{4T}$$
$$A_T = D_{4T} \begin{pmatrix} \sum_2 g_t g_t' & \sum_2 g_t w_t' \\ \sum_2 w_t g_t' & \sum_2 w_t w_t' \end{pmatrix} D_{4T}.$$

Let ρ_T^* and ρ_T be the smallest eigenvalue of A_T^* and A_T. Then,

$$\sum_1 d_t^2 + \sum_2 d_t^2 \geq \rho_T^* \left[(T^2\eta^2) \left\|\hat{\delta}_f - \delta_f^0\right\|^2 + (T\eta) \left\|\hat{\beta}_f - \beta_f^0\right\|^2 + (T^2\eta^2) \left\|\hat{\delta}_{bk} - \delta_{bj}^0\right\|^2 + (T\eta) \left\|\hat{\beta}_{bk} - \beta_{bj}^0\right\|^2 \right] + \rho_T \left[(T^2\eta^2) \left\|\hat{\delta}_f - \delta_f^0\right\|^2 + (T\eta) \left\|\hat{\beta}_f - \beta_f^0\right\|^2 + (T^2\eta^2) \left\|\hat{\delta}_{bk} - \delta_{bj+1}^0\right\|^2 + (T\eta) \left\|\hat{\beta}_{bk} - \beta_{bj+1}^0\right\|^2 \right]$$

$$\geq T^2\eta^2 \min\left\{\rho_T, \rho_T^*\right\}\left(\left\|\hat{\delta}_{bk}-\delta_{bj}^0\right\|^2+\left\|\hat{\delta}_{bk}-\delta_{bj+1}^0\right\|^2\right)$$

$$\geq (1/2)T^2\eta^2 \min\left\{\rho_T, \rho_T^*\right\}\left\|\delta_{bj}^0-\delta_{bj+1}^0\right\|^2.$$

By Assumption A2, ρ_T and ρ_T^* are bounded away from zero. Hence, $\sum_{t=1}^T d_t^2 > T^2 C\|\delta_{bj}^0-\delta_{bj+1}^0\|^2$ for some $C > 0$ with probability no less than $\epsilon_0 > 0$. □

Proof of Theorem 2. The basic idea of the proof is same as that of [BP]. We analyze the case with three breaks ($m = 3$). Without loss of generality, we will consider the case $T_2 < T_2^0$. For each $C > 0$, define

$$V_\epsilon(C) = \left\{(T_1, T_2, T_3);\ \left|T_i - T_i^0\right| < \epsilon T, 1 \leq i \leq 3,\ T_2 - T_2^0 < -C/v_T^2\right\}.$$

We investigate the behavior of the sum of squared residuals $S_T(T_1, T_2, T_3)$ on the set $V_\epsilon(C)$. To prove Theorem 2, it is enough to show that for each η, there exists $C > 0$ and $\epsilon > 0$ such that for large T,

$$P(\min\{[S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)]/(T_2^0 - T_2)\} \leq 0) < \eta$$

where the minimum is taken over the set $V_\epsilon(C)$. This would imply that with large probability, $|\hat{T}_2 - T_2^0| \leq C/v_T^2$. We use the following additional notation: $\hat{\gamma}_2^*$ = estimate of γ_2^0 associated with the regressor $(0, \ldots, 0, w_{T_1+1}, \ldots, w_{T_2}, 0, \ldots, 0)'$; $\hat{\gamma}_\Delta$ = estimate of γ_2^0 associated with the regressor $W_\Delta = (0, \ldots, 0, w_{T_2+1}, \ldots, w_{T_2^0}, 0, \ldots, 0)'$; $\hat{\gamma}_3^*$ = estimate of γ_3^0 associated with the regressor $(0, \ldots, 0, w_{T_2^0+1}, \ldots, w_{T_3}, 0, \ldots, 0)'$; $\mathrm{SSR}_1 = S_T(T_1, T_2, T_3)$, $\mathrm{SSR}_2 = S_T(T_1, T_2^0, T_3)$, and $\bar{F} = (G, \bar{W})$. Then,

$$(\mathrm{SSR}_1 - \mathrm{SSR}_2)/(T_2^0 - T_2) \geq \frac{(\hat{\gamma}_3^* - \hat{\gamma}_\Delta)'\left[W_\Delta' W_\Delta\right](\hat{\gamma}_3^* - \hat{\gamma}_\Delta)}{(T_2^0 - T_2)}$$
$$- \frac{(\hat{\gamma}_3^* - \hat{\gamma}_\Delta)'\left[W_\Delta'\bar{F}\right]\left[\bar{F}'\bar{F}\right]^{-1}\left[\bar{F}'W_\Delta\right](\hat{\gamma}_3^* - \hat{\gamma}_\Delta)}{(T_2^0 - T_2)}$$
$$- \frac{(\hat{\gamma}_2^* - \hat{\gamma}_\Delta)'\left[W_\Delta' W_\Delta/\right](\hat{\gamma}_2^* - \hat{\gamma}_\Delta)}{(T_2^0 - T_2)}.$$

We have, uniformly on the set $V_\epsilon(C)$,

$$\sum_{T_i+1}^{T_i^0} z_{bt}z_{bt}' = \left|T_i - T_i^0\right| O_p(T),$$

$$\sum_{T_i+1}^{T_i^0} z_{bt}x_{bt}' = \left|T_i - T_i^0\right| O_p(T^{1/2}),$$

$$\sum_{T_i+1}^{T_i^0} x_{bt}x_{bt}' = \left|T_i - T_i^0\right| O_p(1),$$

$$D_{2T}^{-1}(\hat{\gamma}_i^* - \gamma_i^0) = \epsilon O_p(T^{1/2}v_T)$$

and

$$D_{2T}^{-1}(\hat{\gamma}_\Delta^* - \gamma_2^0) = (W_\Delta' W_\Delta)^{-1}W_\Delta' U + \epsilon O_p(T^{1/2}v_T).$$

Using these results, we can show that

$$(\mathrm{SSR}_1 - \mathrm{SSR}_2)/(T_2^0 - T_2)$$
$$\geq Av_T^2 - \frac{(W_\Delta' U)'(W_\Delta' W_\Delta)^{-1}(W_\Delta' U)}{(T_2^0 - T_2)} - \epsilon O_p(v_T^2)$$
$$\geq Av_T^2 - \frac{(q_2+p_2)}{T_2^0 - T_2}\sum_{l=1}^{q_2+p_2}\left(\sum_{t=T_2+1}^{T_2^0} W_{lt}u_t\right)^2 \Bigg/ \left(\sum_{t=T_2+1}^{T_2^0} W_{lt}^2\right)$$
$$- \epsilon O_p(v_T^2)$$

where A is a positive constant and W_{lt} is the l-th component of W_t. For every $\eta > 0$, we can choose a small $\epsilon > 0$ such that $P(\epsilon O_p(v_T^2) > Av_T^2/2) < \eta$. Then $P(\min\{(\mathrm{SSR}_1 - \mathrm{SSR}_2)/(T_2^0 - T_2)\} \leq 0)$ is bounded by

$$\eta + P\left(\max\left\{\frac{(q_2+p_2)}{T_2^0 - T_2}\sum_{l=1}^{q_2+p_2}\left(\sum_{t=T_2+1}^{T_2^0} W_{lt}u_t\right)^2 \Bigg/ \left(\sum_{t=T_2+1}^{T_2^0} W_{lt}^2\right)\right\} > Av_T^2/2\right).$$

To prove that the latter probability is less than η, it is sufficient to show that for each l,

$$P\left(\max_{T_2 < T_2^0 - Cv_T^{-2}} \frac{1}{T_2^0 - T_2}\left(\sum_{t=T_2+1}^{T_2^0} W_{lt}u_t\right)^2 \Bigg/ \left(\sum_{t=T_2+1}^{T_2^0} W_{lt}^2\right) > Av_T^2/2\right) < \eta. \tag{A.7}$$

If W_{lt} is an $I(1)$, we have

$$\sum_{t=T_2+1}^{T_2^0} W_{lt}^2 \geq (T_2^0 - T_2)\min_{t\in[T_2+1, T_2^0]} W_{lt}^2 = (T_2^0 - T_2)O_p(T).$$

Then (A.7) is bounded by

$$P\left(\max_{T_2 < T_2^0 - Cv_T^{-2}} O_p(1)\left(\frac{1}{T_2^0 - T_2}\sum_{t=T_2+1}^{T_2^0} T^{-1/2}W_{lt}u_t\right)^2 > Av_T^2/2\right).$$

We can choose a $B < \infty$ such that $P(O_p(1) > B) < \eta$ so that this probability is bounded by

$$\eta + P\left(\max_{T_2 < T_2^0 - Cv_T^{-2}}\left|\frac{1}{T_2^0 - T_2}\sum_{t=T_2+1}^{T_2^0} T^{-1/2}W_{lt}u_t\right| > (A/2B)^{1/2}v_T\right)$$
$$= \eta + P\left(\max_{T_2 < T_2^0 - Cv_T^{-2}}\left|\frac{T^{-1/2}W_{lT_2^0}}{T_2^0 - T_2}\sum_{t=T_2+1}^{T_2^0} u_t + \frac{T^{-1/2}}{T_2^0 - T_2}\sum_{t=T_2+1}^{T_2^0}(W_{lt} - W_{lT_2^0})u_t\right| > (A/2B)^{1/2}v_T\right)$$
$$\leq \eta + P\left(\max_{T_2 < T_2^0 - Cv_T^{-2}}|O_p(1)|\left|\frac{1}{T_2^0 - T_2}\sum_{t=T_2+1}^{T_2^0} u_t\right| > (A/8B)^{1/2}v_T\right) + P\left(\max_{T_2 < T_2^0 - Cv_T^{-2}}\left|\frac{T^{-1/2}}{T_2^0 - T_2}\times\sum_{t=T_2+1}^{T_2^0}(W_{lt} - W_{lT_2^0})u_t\right| > (A/8B)^{1/2}v_T\right).$$

Now, we can choose a $B_1 < \infty$ such that $P(|O_p(1)| > B_1) < \eta$ so that this probability is bounded by

$$2\eta + P\left(\max_{T_2<T_2^0-Cv_T^{-2}}\left|\frac{1}{T_2^0-T_2}\sum_{t=T_2+1}^{T_2^0}u_t\right| > (A/8BB_1^2)^{1/2}v_T\right) + P\left(\max_{T_2<T_2^0-Cv_T^{-2}}\left|\frac{T^{-1/2}}{T_2^0-T_2}\sum_{t=T_2+1}^{T_2^0}(W_{lt}-W_{lT_2^0})u_t\right| > (A/8B)^{1/2}v_T\right).$$

The first probability is bounded by η using Lemma A.1(a). Consider now the second one. We have

$$P\left(\max_{T_2<T_2^0-Cv_T^{-2}}\left|\frac{T^{-1/2}}{T_2^0-T_2}\sum_{t=T_2+1}^{T_2^0}(W_{lt}-W_{lT_2^0})u_t\right| > (A/8B)^{1/2}v_T\right) = P\left(\frac{1}{T^{1/2}v_T}\max_{T_2<T_2^0-Cv_T^{-2}}\left|\frac{1}{T_2^0-T_2}\times\sum_{t=T_2+1}^{T_2^0}(W_{lt}-W_{lT_2^0})u_t\right| > (A/8B)^{1/2}\right). \quad (A.8)$$

Now,

$$\max_{T_2<T_2^0-Cv_T^{-2}}\left|\frac{1}{T_2^0-T_2}\sum_{t=T_2+1}^{T_2^0}(W_{lt}-W_{lT_2^0})u_t\right| = O_p(1)$$

since $\sum_{t=T_2+1}^{T_2^0}(W_{lt}-W_{lT_2^0})u_t = \sum_{t=1}^{T_2^0-T_2}(W_{lt+T_2}-W_{lT_2^0})u_{t+T_2} = \sum_{t=1}^{T_2^0-T_2}W_t^*u_{t+T_2}$, say, where W_t^* is an $I(1)$ process. Then using the fact that $T^{1/2}v_T\to\infty$,

$$\frac{1}{T^{1/2}v_T}\max_{T_2<T_2^0-Cv_T^{-2}}\left|\frac{1}{T_2^0-T_2}\sum_{t=T_2+1}^{T_2^0}(W_{lt}-W_{lT_2^0})u_t\right| = o_p(1).$$

Given that A and B are constants, the probability (A.8) is also bounded by η, which shows that (A.7) holds when W_{lt} is an $I(1)$ variable. If W_{lt} is an $I(0)$ variable, then we have

$$P\left(\max_{T_2<T_2^0-Cv_T^{-2}}\frac{1}{T_2^0-T_2}\left(\sum_{t=T_2+1}^{T_2^0}W_{lt}u_t\right)^2 \Big/ \left(\sum_{t=T_2+1}^{T_2^0}W_{lt}^2\right) > Av_T^2/2\right)$$
$$= P\left(\max_{T_2<T_2^0-Cv_T^{-2}}\frac{1}{(T_2^0-T_2)^2}\left(\sum_{t=T_2+1}^{T_2^0}W_{lt}u_t\right)^2 \Big/ \left(\frac{1}{(T_2^0-T_2)}\sum_{t=T_2+1}^{T_2^0}W_{lt}^2\right) > Av_T^2/2\right)$$
$$= P\left(\max_{T_2<T_2^0-Cv_T^{-2}}\frac{O_p(1)}{(T_2^0-T_2)}\left(\sum_{t=T_2+1}^{T_2^0}W_{lt}u_t\right) > (A/2)^{1/2}v_T\right).$$

We can choose $B < \infty$ such that $P(O_p(1) > B) < \eta$ so that the last probability is bounded by

$$\eta + P\left(\max_{T_2<T_2^0-Cv_T^{-2}}\frac{B}{(T_2^0-T_2)}\left(\sum_{t=T_2+1}^{T_2^0}W_{lt}u_t\right) > (A/2)^{1/2}v_T\right) < 2\eta.$$

The last inequality follows applying Lemma A.1(a). This completes the proof. □

Proof of Proposition 1. Let $\bar{W}^*$ be the matrix $\bar{W}$ evaluated at the estimated break points $(\hat{T}_1,\ldots,\hat{T}_m)$. The true model can be written as $Y = G\alpha + \bar{W}^*\gamma + U^*$, where $U^* = U + (\bar{W}^0 - \bar{W}^*)\gamma$. Thus, we have

$$\tilde{D}_T^{-1}(\hat{\theta}-\theta^0) = \begin{pmatrix} D_{1T}G'GD_{1T} & D_{1T}G'\bar{W}^*\tilde{D}_{m+1} \\ \tilde{D}_{m+1}\bar{W}^{*\prime}GD_{1T} & \tilde{D}_{m+1}\bar{W}^{*\prime}\bar{W}^*\tilde{D}_{m+1}\end{pmatrix}^{-1} \times \begin{pmatrix} D_{1T}G'U + D_{1T}G'(\bar{W}^0-\bar{W}^*)\gamma^0 \\ \tilde{D}_{m+1}\bar{W}^{*\prime}U + \tilde{D}_{m+1}\bar{W}^{*\prime}(\bar{W}^0-\bar{W}^*)\gamma^0\end{pmatrix}. \quad (A.9)$$

We need to show that the limit of the right hand side of (A.9) is the same as the limit when $\bar{W}^*$ is replaced by $\bar{W}^0$. Suppose $T_i < T_i^0$ for all $i = 1,\ldots,m$. Consider the term $\tilde{D}_{m+1}\bar{W}^{*\prime}(\bar{W}^0-\bar{W}^*)\gamma^0$. This involves terms like $T^{-3/2}v_T\sum_{t=T_i}^{T_i^0}z_{bt}z_{bt}'$, $T^{-1/2}v_T\sum_{t=T_i}^{T_i^0}x_{bt}x_{bt}'$ etc. We have

$$T^{-3/2}v_T\sum_{t=T_i}^{T_i^0}z_{bt}z_{bt}' = T^{-1/2}v_T^{-1}\left(T^{-1}v_T^2\sum_{t=T_i^0+sv_T^{-2}}^{T_i^0}z_{bt}z_{bt}'\right) = o_p(1)\cdot O_p(1) = o_p(1)$$
$$T^{-1/2}v_T\sum_{t=T_i}^{T_i^0}x_{bt}x_{bt}' = T^{-1/2}v_T^{-1}\left(v_T^2\sum_{t=T_i^0+sv_T^{-2}}^{T_i^0}x_{bt}x_{bt}'\right) = o_p(1)\cdot O_p(1) = o_p(1).$$

The other terms can be handled similarly, and the result follows. □

Proof of Theorem 3. We provide a detailed proof for the case of two breaks. The extension to m breaks is straightforward. We have $(\hat{T}_1,\hat{T}_2) = \arg\min_{(T_1,T_2)}\mathrm{SSR}(T_1,T_2)$ or

$$(\hat{T}_1,\hat{T}_2) = \arg\max_{(T_1,T_2)}\left\{\mathrm{SSR}(T_1^0,T_2^0)-\mathrm{SSR}(T_1,T_2)\right\}.$$

Following Perron and Zhu (2005), we have

$$\mathrm{SSR}(T_1^0,T_2^0)-\mathrm{SSR}(T_1,T_2) = -\theta^{0\prime}(\bar{F}^0-\bar{F})'(I-P_{\bar{F}})(\bar{F}^0-\bar{F})\theta^0 - 2\theta^{0\prime}(\bar{F}^0-\bar{F})'(I-P_{\bar{F}})U - U'(P_{\bar{F}^0}-P_{\bar{F}})U$$

where $\bar{F} = (G,\bar{W})$, $\bar{F}^0 = (G,\bar{W}^0)$, $\theta^0 = (\alpha^{0\prime},\gamma^{0\prime})'$, $\gamma^0 = (c^0,\delta_b^{0\prime},\beta_b^{0\prime})'$. Let $T_1 = T_1^0 + \left[s_1v_T^{-2}\right]$, $T_2 = T_2^0 + \left[s_2v_T^{-2}\right]$. We then have the following four possible cases: (1) $s_1 < 0, s_2 < 0$, (2) $s_1 < 0, s_2 > 0$, (3) $s_1 > 0, s_2 > 0$, (4) $s_1 > 0, s_2 < 0$. We give a detailed proof for Case 1 only, the proof for the other cases being similar. Consider the term

$$\theta^{0\prime}(\bar{F}^0-\bar{F})'(\bar{F}^0-\bar{F})\theta^0 = \bar{\gamma}_{T,1}'\left(\sum_{T_1+1}^{T_1^0}w_tw_t'\right)\bar{\gamma}_{T,1} + \bar{\gamma}_{T,2}'\left(\sum_{T_1+1}^{T_1^0}w_tw_t'\right)\bar{\gamma}_{T,2}$$

$$\theta^{0\prime}(\bar F^0-\bar F)'(\bar F^0-\bar F)\theta^0$$
$$\Rightarrow |s_1|\,\bar\gamma_1'\begin{pmatrix}(\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_1^0)W_z^b(\lambda_1^0)'(\Omega_{zz}^{bb})^{1/2} & (\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_1^0) & (\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_1^0)\mu_b'\\ W_z^b(\lambda_1^0)'(\Omega_{zz}^{bb})^{1/2} & 1 & \mu_b'\\ \mu_bW_z^b(\lambda_1^0)'(\Omega_{zz}^{bb})^{1/2} & \mu_b & Q_{xx}^{bb}\end{pmatrix}\bar\gamma_1$$
$$+\,|s_2|\,\bar\gamma_2'\begin{pmatrix}(\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_2^0)W_z^b(\lambda_2^0)'(\Omega_{zz}^{bb})^{1/2} & (\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_2^0) & (\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_2^0)\mu_b'\\ W_z^b(\lambda_2^0)'(\Omega_{zz}^{bb})^{1/2} & 1 & \mu_b'\\ \mu_bW_z^b(\lambda_2^0)'(\Omega_{zz}^{bb})^{1/2} & \mu_b & Q_{xx}^{bb}\end{pmatrix}\bar\gamma_2$$
$$=\tilde A(s_1,s_2)$$

Box III.

$$=\sum_{i=1}^{2}\bar\gamma_i'\begin{pmatrix}T^{-1}v_T^2\sum_{T_i+1}^{T_i^0}z_{bt}z_{bt}' & T^{-1/2}v_T^2\sum_{T_i+1}^{T_i^0}z_{bt} & T^{-1/2}v_T^2\sum_{T_i+1}^{T_i^0}z_{bt}x_{bt}'\\ T^{-1/2}v_T^2\sum_{T_i+1}^{T_i^0}z_{bt}' & (T_i^0-T_i)v_T^2 & v_T^2\sum_{T_i+1}^{T_i^0}x_{bt}'\\ T^{-1/2}v_T^2\sum_{T_i+1}^{T_i^0}x_{bt}z_{bt}' & v_T^2\sum_{T_i+1}^{T_i^0}x_{bt} & v_T^2\sum_{T_i+1}^{T_i^0}x_{bt}x_{bt}'\end{pmatrix}\bar\gamma_i.$$

We consider each of the terms in the above matrices. For $i=1,2$, we have, (i) $T^{-1}v_T^2\sum_{T_i+1}^{T_i^0}z_{bt}z_{bt}' = |s_i|\,(T^{-1/2}z_{bT_i^0})(T^{-1/2}z_{bT_i^0}')-T^{-1}v_T^2\sum_{T_i^0+s_iv_T^{-2}}^{T_i^0}(z_{bT_1^0}z_{bT_1^0}'-z_{bt}z_{bt}')$. Since the second term is $o_p(1)$, we have $T^{-1}v_T^2\sum_{T_i+1}^{T_i^0}z_{bt}z_{bt}'\Rightarrow|s_i|\,(\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_i^0)W_z^b(\lambda_i^0)'(\Omega_{zz}^{bb})^{1/2}$; (ii) $T^{-1/2}v_T^2\sum_{T_i+1}^{T_i^0}z_{bt} = |s_i|\,T^{-1/2}z_{bT_i^0}+o_p(1)\Rightarrow |s_i|\,(\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_i^0)$; (iii) $T^{-1/2}v_T^2\sum_{T_i+1}^{T_i^0}z_{bt}x_{bt}'\Rightarrow|s_i|\,(\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_i^0)\mu_b'$; (iv) $v_T^2\sum_{T_i+1}^{T_i^0}x_{bt}\Rightarrow|s_i|\,\mu_b$; (v) $v_T^2\sum_{T_i+1}^{T_i^0}x_{bt}x_{bt}'\Rightarrow|s_i|\,Q_{xx}^{bb}$. Thus, we get the equation in Box III.

Next, consider

$$\theta^{0\prime}(\bar F^0-\bar F)'U = -\bar\gamma_1'\begin{pmatrix}T^{-1/2}v_T\sum_{T_1+1}^{T_1^0}z_{bt}u_t\\ v_T\sum_{T_1+1}^{T_1^0}u_t\\ v_T\sum_{T_1+1}^{T_1^0}x_{bt}u_t\end{pmatrix}-\bar\gamma_2'\begin{pmatrix}T^{-1/2}v_T\sum_{T_2+1}^{T_2^0}z_{bt}u_t\\ v_T\sum_{T_2+1}^{T_2^0}u_t\\ v_T\sum_{T_2+1}^{T_2^0}x_{bt}u_t\end{pmatrix}.$$

We have (i) $v_T\sum_{T_i+1}^{T_i^0}u_t\Rightarrow\sigma\eta_i(s_i)$; (ii) $T^{-1/2}v_T\sum_{T_i+1}^{T_i^0}z_{bt}u_t\Rightarrow(\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_i^0)\sigma\eta_i(s_i)$; (iii) $v_T\sum_{T_i+1}^{T_i^0}x_{bt}u_t\Rightarrow(Q_{xu}^{bb})^{1/2}W_{xb,i}^{(i)}(s_i)$. Hence, it follows that

$$\theta^{0\prime}(\bar F^0-\bar F)'U \Rightarrow -\bar\gamma_1'\begin{pmatrix}(\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_1^0)\sigma\eta_1(s_1)\\ \sigma\eta_1(s_1)\\ (Q_{xu}^{bb})^{1/2}W_{xb,1}^{(1)}(s_1)\end{pmatrix}-\bar\gamma_2'\begin{pmatrix}(\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_2^0)\sigma\eta_2(s_2)\\ \sigma\eta_2(s_2)\\ (Q_{xu}^{bb})^{1/2}W_{xb,2}^{(2)}(s_2)\end{pmatrix} = -\tilde B_1(s_1,s_2).$$

Next, we need to establish that $\theta^{0\prime}(\bar F^0-\bar F)'P_{\bar F}(\bar F^0-\bar F)\theta^0 = o_p(1)$ and $\theta^{0\prime}(\bar F^0-\bar F)'P_{\bar F}U = o_p(1)$. We will show the former, the latter can be shown using similar arguments. Define $\tilde D^* = \mathrm{diag}(T^{-1}I_{q_b},T^{-1/2}I_{p_b},\dots,T^{-1}I_{q_b},T^{-1/2}I_{p_b})$ and $\tilde D_T = \mathrm{diag}(T^{-1}I_{q_f},T^{-1/2}I_{p_f},\tilde D^*)$. Then,

$$\theta^{0\prime}(\bar F^0-\bar F)'P_{\bar F}(\bar F^0-\bar F)\theta^0 = \theta^{0\prime}(\bar F^0-\bar F)'\bar F\tilde D_T(\tilde D_T\bar F'\bar F\tilde D_T)^{-1}\times\tilde D_T\bar F'(\bar F^0-\bar F)\theta^0.$$

We will show that $\tilde D_T\bar F'(\bar F^0-\bar F)\theta^0 = o_p(1)$, which together with $(\tilde D_T\bar F'\bar F\tilde D_T)^{-1} = O_p(1)$ gives the desired result. Now $\tilde D_T\bar F'(\bar F^0-\bar F)\theta^0$ involves the following terms: (i) $T^{-3/2}v_T\sum_{T_i+1}^{T_i^0}z_{ft}z_{bt}' = O_p(T^{-1/2}v_T^{-1}) = o_p(1)$; (ii) $T^{-1}v_T\sum_{T_i+1}^{T_i^0}z_{ft}x_{bt}' = O_p(T^{-1}v_T^{-2}) = o_p(1)$; (iii) $T^{-1}v_T\sum_{T_i+1}^{T_i^0}z_{ft} = O_p(T^{-1/2}v_T^{-1}) = o_p(1)$, $T^{-1}v_T\sum_{T_i+1}^{T_i^0}z_{bt} = O_p(T^{-1/2}v_T^{-1}) = o_p(1)$; (iv) $T^{-1}v_T\sum_{T_i+1}^{T_i^0}x_{ft}z_{bt}' = O_p(T^{-1}v_T^{-2}) = o_p(1)$, $T^{-1}v_T\sum_{T_i+1}^{T_i^0}x_{bt}z_{bt}' = O_p(T^{-1}v_T^{-2}) = o_p(1)$; (v) $T^{-1}v_T\sum_{T_i+1}^{T_i^0}x_{ft}x_{bt}' = O_p(T^{-1}v_T^{-1}) = o_p(1)$; (vi) $T^{-1/2}v_T\sum_{T_i+1}^{T_i^0}x_{ft} = O_p(T^{-1/2}) = o_p(1)$, $T^{-1/2}v_T\sum_{T_i+1}^{T_i^0}x_{bt} = O_p(T^{-1/2}) = o_p(1)$; (vii) $T^{-3/2}v_T\sum_{T_i+1}^{T_i^0}z_{bt}z_{bt}' = O_p(T^{-1/2}v_T^{-1}) = o_p(1)$. Hence, $\tilde D_T\bar F'(\bar F^0-\bar F)\theta^0 = o_p(1)$. Finally,

$$U'(P_{\bar F^0}-P_{\bar F})U = U'(\bar F^0-\bar F)\tilde D_T(\tilde D_T\bar F^{0\prime}\bar F^0\tilde D_T)^{-1}\tilde D_T\bar F^{0\prime}U + U'\bar F\tilde D_T(\tilde D_T\bar F'\bar F\tilde D_T)^{-1}(\tilde D_T\bar F'\bar F\tilde D_T-\tilde D_T\bar F^{0\prime}\bar F^0\tilde D_T)(\tilde D_T\bar F^{0\prime}\bar F^0\tilde D_T)^{-1}\times(\tilde D_T\bar F^{0\prime}U)+U'\bar F\tilde D_T(\tilde D\bar F'\bar F\tilde D_T)^{-1}\tilde D_T(\bar F^0-\bar F)'U.$$

We have $U'\bar F\tilde D_T = O_p(1)$, $U'\bar F^0\tilde D_T = O_p(1)$, $(\tilde D_T\bar F'\bar F\tilde D_T)^{-1} = O_p(1)$, $(\tilde D_T\bar F'^0\bar F^0\tilde D_T)^{-1} = O_p(1)$. Also, for $i=1,2$, we have $T^{-1}\sum_{T_i+1}^{T_i^0}z_{bt}'u_t = O_p(T^{-1/2}v_T^{-1}) = o_p(1)$, $T^{-1/2}\sum_{T_i+1}^{T_i^0}u_t = O_p(T^{-1/2}v_T^{-1}) = o_p(1)$, $T^{-1/2}\sum_{T_i+1}^{T_i^0}x_{bt}'u_t = O_p(T^{-1/2}v_T^{-1}) = o_p(1)$. Hence, $U'(\bar F^0-\bar F)\tilde D_T = o_p(1)$. Similarly, we can show that $(\tilde D_T\bar F'\bar F\tilde D_T-\tilde D_T\bar F^{0\prime}\bar F^0\tilde D_T) = o_p(1)$. Hence, $U'(P_{\bar F^0}-P_{\bar F})U = o_p(1)$. We thus obtain

$$\mathrm{SSR}(T_1^0,T_2^0)-\mathrm{SSR}(T_1,T_2) = \tilde H_{T,1}\left(\left[s_1v_T^{-2}\right],\left[s_2v_T^{-2}\right]\right)\Rightarrow\tilde H_1(s_1,s_2)$$

with $\tilde H_1(s_1,s_2) = -(1/2)\tilde A(s_1,s_2)+\tilde B_1(s_1,s_2)$. For Case 2 with $s_1<0, s_2>0$, we have

$$\tilde B_2(s_1,s_2) = \bar\gamma_1'\begin{pmatrix}(\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_1^0)\sigma\eta_1(s_1)\\ \sigma\eta_1(s_1)\\ (Q_{xu}^{bb})^{1/2}W_{xb,1}^{(1)}(s_1)\end{pmatrix}+\bar\gamma_2'\begin{pmatrix}(\Omega_{zz}^{bb})^{1/2}W_z^b(\lambda_2^0)\sigma\eta_3(s_2)\\ \sigma\eta_3(s_2)\\ (Q_{xu}^{bb})^{1/2}W_{xb,2}^{(2)}(s_2)\end{pmatrix}$$

and

$$\mathrm{SSR}(T_1^0,T_2^0)-\mathrm{SSR}(T_1,T_2) = \tilde H_{T,2}\left(\left[s_1v_T^{-2}\right],\left[s_2v_T^{-2}\right]\right)\Rightarrow\tilde H_2(s_1,s_2),$$

with $\tilde{H}_2(s_1,s_2) = -(1/2)\tilde{A}(s_1,s_2) + \tilde{B}_2(s_1,s_2)$. For Case 3 with $s_1 > 0, s_2 > 0$, we have

$$\tilde{B}_3(s_1,s_2) = \bar{\gamma}_1' \begin{pmatrix} (\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_1^0)\sigma\eta_4(s_1) \\ \sigma\eta_4(s_1) \\ (Q_{xu}^{bb})^{1/2} W_{xb,2}^{(1)}(s_1) \end{pmatrix} + \bar{\gamma}_2' \begin{pmatrix} (\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_2^0)\sigma\eta_3(s_2) \\ \sigma\eta_3(s_2) \\ (Q_{xu}^{bb})^{1/2} W_{xb,2}^{(2)}(s_2) \end{pmatrix}$$

and

$$\mathrm{SSR}(T_1^0,T_2^0) - \mathrm{SSR}(T_1,T_2) = \tilde{H}_{T,3}\left(\left[s_1 v_T^{-2}\right],\left[s_2 v_T^{-2}\right]\right) \Rightarrow \tilde{H}_3(s_1,s_2)$$

with $H_3(s_1,s_2) = -(1/2)A(s_1,s_2) + B_3(s_1,s_2)$. For Case 4 with $s_1 > 0, s_2 < 0$, we have

$$\tilde{B}_4(s_1,s_2) = \bar{\gamma}_1' \begin{pmatrix} (\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_1^0)\sigma\eta_4(s_1) \\ \sigma\eta_4(s_1) \\ (Q_{xu}^{bb})^{1/2} W_{xb,2}^{(1)}(s_1) \end{pmatrix} + \bar{\gamma}_2' \begin{pmatrix} (\Omega_{zz}^{bb})^{1/2} W_z^b(\lambda_2^0)\sigma\eta_2(s_2) \\ \sigma\eta_2(s_2) \\ (Q_{xu}^{bb})^{1/2} W_{xb,1}^{(2)}(s_2) \end{pmatrix}$$

and

$$\mathrm{SSR}(T_1^0,T_2^0) - \mathrm{SSR}(T_1,T_2) = \tilde{H}_{T,4}\left(\left[s_1 v_T^{-2}\right],\left[s_2 v_T^{-2}\right]\right) \Rightarrow \tilde{H}_4(s_1,s_2)$$

with with $\tilde{H}_4(s_1,s_2) = -(1/2)\tilde{A}(s_1,s_2) + \tilde{B}_4(s_1,s_2)$. The result follows using the continuous mapping theorem and a change of variable as in Bai (1997). □

Proof of Theorem 4. As a matter of notation, let $\eta_t = (\Delta z_{t-k_T}', \ldots, \Delta z_{t+k_T}')'$. Following Perron and Zhu (2005), we have

$$\mathrm{SSR}(T_1^0,T_2^0) - \mathrm{SSR}(T_1,T_2) = -\theta^{0\prime}(\bar{F}^0 - \bar{F})'(I - P_{\bar{F}})(\bar{F}^0 - \bar{F})\theta^0 - 2\theta^{0\prime}(\bar{F}^0 - \bar{F})'(I - P_{\bar{F}})V^* - V^{*\prime}(P_{\bar{F}^0} - P_{\bar{F}})V^*$$

where $\bar{F} = (G, \bar{W}), \bar{F}^0 = (G, \bar{W}^0), \theta^0 = (\alpha^{0\prime}, \gamma^{0\prime})', \gamma^0 = (c^0, \delta_b^{0\prime}, \beta_b^{0\prime})'$. Let $T_1 = T_1^0 + \left[s_1 v_T^{-2}\right], T_2 = T_2^0 + \left[s_2 v_T^{-2}\right]$. We will show that, under the stated conditions, $\theta^{0\prime}(\bar{F}^0 - \bar{F})'P_{\bar{F}}(\bar{F}^0 - \bar{F})\theta^0 = o_p(1)$, $\theta^{0\prime}(\bar{F}^0 - \bar{F})'E = o_p(1)$, $\theta^{0\prime}(\bar{F}^0 - \bar{F})'P_{\bar{F}}V^* = o_p(1)$, $V^{*\prime}(P_{\bar{F}^0} - P_{\bar{F}})V^* = o_p(1)$, which implies that the limit distribution of the estimates of the break dates is the same as when the number of $I(0)$ regressors is fixed.

As before, let $\tilde{D}^* = \mathrm{diag}(T^{-1}I_{q_b}, T^{-1/2}I_{p_b}, \ldots, T^{-1}I_{q_b}, T^{-1/2}I_{p_b})$ and additionally define $\tilde{D}_T = \mathrm{diag}(T^{-1}I_{q_f}, T^{-1/2}I_{p_f}, T^{-1/2}I_{(q_f+q_b)(2k_T+1)}, \tilde{D}^*)$. Consider first the term

$$\theta^{0\prime}(\bar{F}^0 - \bar{F})'P_{\bar{F}}(\bar{F}^0 - \bar{F})\theta^0 = \theta^{0\prime}(\bar{F}^0 - \bar{F})'\bar{F}\tilde{D}_T(\tilde{D}_T\bar{F}'\bar{F}\tilde{D}_T)^{-1} \times \tilde{D}_T\bar{F}'(\bar{F}^0 - \bar{F})\theta^0.$$

We have $\left\|(\tilde{D}_T\bar{F}'\bar{F}\tilde{D}_T)^{-1}\right\| = O_p(1)$. Now $\|\tilde{D}_T\bar{F}'(\bar{F}^0 - \bar{F})\theta^0\|$ involves the following additional terms (compared with the case with no leads and lags of the first-differences of the $I(1)$ regressors): (i) $T^{-3/2}v_T\left\|\sum_{T_i+1}^{T_i^0}\eta_t z_{bt}'\right\| = O_p(k_T^{1/2}T^{-1}) = o_p(1)$, (ii) $T^{-1/2}v_T\left\|\sum_{T_i+1}^{T_i^0}\eta_t\right\| = T^{-1/2}O_p(k_T^{1/2}) = o_p(1)$ and (iii) $T^{-1/2}v_T\left\|\sum_{T_i+1}^{T_i^0}\eta_t x_{bt}'\right\| = O_p(T^{-1/2}k_T^{1/2}) = o_p(1)$. Thus we have

$$\theta^{0\prime}(\bar{F}^0 - \bar{F})'P_{\bar{F}}(\bar{F}^0 - \bar{F})\theta^0 \le \left\|\theta^{0\prime}(\bar{F}^0 - \bar{F})'\bar{F}\tilde{D}_T\right\|\left\|(\tilde{D}_T\bar{F}'\bar{F}\tilde{D}_T)^{-1}\right\| \times \|\tilde{D}_T\bar{F}'(\bar{F}^0 - \bar{F})\theta^0\| = o_p(1).$$

Next, consider the term $\theta^{0\prime}(\bar{F}^0 - \bar{F})'E$. This involves the following components: (i) $T^{-1/2}v_T\left\|\sum_{T_i+1}^{T_i^0} z_{bt}e_t\right\| = o_p(v_T^{-1}k_T^{-1}) = o_p(1)$, (ii) $v_T\left\|\sum_{T_i+1}^{T_i^0} e_t\right\| = o_p(v_T^{-1}k_T^{-1}) = o_p(1)$ and (iii) $v_T\left\|\sum_{T_i+1}^{T_i^0} x_{bt}e_t\right\| = o_p(v_T^{-1}k_T^{-1}) = o_p(1)$. Next, consider the term

$$\theta^{0\prime}(\bar{F}^0 - \bar{F})'P_{\bar{F}}V^* = \theta^{0\prime}(\bar{F}^0 - \bar{F})'\bar{F}\tilde{D}_T(\tilde{D}_T\bar{F}'\bar{F}\tilde{D}_T)^{-1}\tilde{D}_T\bar{F}'V^*.$$

The matrix $\tilde{D}_T\bar{F}'V^*$ involves the following terms: (i) $T^{-1}\sum_1^{T_i} z_{bt}v_t^* = O_p(1)$, (ii) $T^{-1/2}\sum_1^{T_i} x_{bt}v_t^* = O_p(1)$ and $T^{-1/2}\left\|\sum_1^T \eta_t v_t^*\right\| = o_p(T^{1/2}k_T^{-1/2})$. Thus we get

$$\begin{aligned}&\left\|\theta^{0\prime}(\bar{F}^0 - \bar{F})'P_{\bar{F}}V^*\right\| \\ &\le \left\|\theta^{0\prime}(\bar{F}^0 - \bar{F})'\bar{F}\tilde{D}_T\right\|\left\|(\tilde{D}_T\bar{F}'\bar{F}\tilde{D}_T)^{-1}\right\|\left\|\tilde{D}_T\bar{F}'V^*\right\| \\ &= O_p(k_T^{1/2}T^{-1/2})O_p(1)o_p(T^{1/2}k_T^{-1/2}) = o_p(1).\end{aligned}$$

Last, consider the term

$$\begin{aligned}V^{*\prime}(P_{\bar{F}^0} - P_{\bar{F}})V^* &= V^{*\prime}(\bar{F}^0 - \bar{F})\tilde{D}_T(\tilde{D}_T\bar{F}^{0\prime}\bar{F}^0\tilde{D}_T)^{-1}\tilde{D}_T\bar{F}^{0\prime}V^* \\ &+ V^{*\prime}\bar{F}\tilde{D}_T(\tilde{D}_T\bar{F}'\bar{F}\tilde{D}_T)^{-1}(\tilde{D}_T\bar{F}'\bar{F}\tilde{D}_T - \tilde{D}_T\bar{F}^{0\prime}\bar{F}^0\tilde{D}_T)(\tilde{D}_T\bar{F}^{0\prime}\bar{F}^0\tilde{D}_T)^{-1} \\ &\times (\tilde{D}_T\bar{F}^{0\prime}V^*) + V^{*\prime}\bar{F}\tilde{D}_T(\tilde{D}\bar{F}'\bar{F}\tilde{D}_T)^{-1}\tilde{D}_T(\bar{F}^0 - \bar{F})'V^*.\end{aligned}$$

We have $\left\|V^{*\prime}\bar{F}\tilde{D}_T\right\| = o_p(T^{1/2}k_T^{-1/2})$, $\left\|V^{*\prime}\bar{F}^0\tilde{D}_T\right\| = o_p(T^{1/2}k_T^{-1/2})$, $\left\|(\tilde{D}_T\bar{F}'\bar{F}\tilde{D}_T)^{-1}\right\| = O_p(1)$, $\left\|(\tilde{D}_T\bar{F}^{0\prime}\bar{F}^0\tilde{D}_T)^{-1}\right\| = O_p(1)$. Also, for $i = 1, 2$, we have $T^{-1}\sum_{T_i+1}^{T_i^0} z_{bt}'v_t^* = O_p(T^{-1/2}v_T^{-1}) = o_p(1)$, $T^{-1/2}\sum_{T_i+1}^{T_i^0} v_t^* = O_p(T^{-1/2}v_T^{-1}) = o_p(1)$, $T^{-1/2}\sum_{T_i+1}^{T_i^0} x_{bt}' v_t^* = O_p(T^{-1/2}v_T^{-1}) = o_p(1)$. Thus we have

$$\begin{aligned}&\left\|V^{*\prime}(\bar{F}^0 - \bar{F})\tilde{D}_T(\tilde{D}_T\bar{F}^{0\prime}\bar{F}^0\tilde{D}_T)^{-1}\tilde{D}_T\bar{F}^{0\prime}V^*\right\| \\ &\le \left\|V^{*\prime}(\bar{F}^0 - \bar{F})\tilde{D}_T\right\|\left\|(\tilde{D}_T\bar{F}^{0\prime}\bar{F}^0\tilde{D}_T)^{-1}\right\|\left\|\tilde{D}_T\bar{F}^{0\prime}V^*\right\| \\ &= O_p(T^{-1/2}v_T^{-1})O_p(1)o_p(T^{1/2}k_T^{-1/2}) = o_p(k_T^{-1/2}v_T^{-1}) = o_p(1).\end{aligned}$$

Using similar arguments, we can show that $\|V^{*\prime}\bar{F}\tilde{D}_T(\tilde{D}\bar{F}'\bar{F}\tilde{D}_T)^{-1}\tilde{D}_T(\bar{F}^0 - \bar{F})'V^*\| = o_p(1)$. Finally, we have $\|(\tilde{D}_T\bar{F}'\bar{F}\tilde{D}_T - \tilde{D}_T\bar{F}^{0\prime}\bar{F}^0\tilde{D}_T)\| = O_p(T^{-1}k_T^{1/2}v_T^{-1})$. Hence, we get

$$\begin{aligned}&\left\|V^{*\prime}\bar{F}\tilde{D}_T(\tilde{D}_T\bar{F}'\bar{F}\tilde{D}_T)^{-1}(\tilde{D}_T\bar{F}'\bar{F}\tilde{D}_T - \tilde{D}_T\bar{F}^{0\prime}\bar{F}^0\tilde{D}_T)(\tilde{D}_T\bar{F}^{0\prime}\bar{F}^0\tilde{D}_T)^{-1}(\tilde{D}_T\bar{F}^{0\prime}V^*)\right\| \\ &\le \left\|V^{*\prime}\bar{F}\tilde{D}_T\right\|\left\|(\tilde{D}_T\bar{F}'\bar{F}\tilde{D}_T)^{-1}\right\|\left\|(\tilde{D}_T\bar{F}'\bar{F}\tilde{D}_T - \tilde{D}_T\bar{F}^{0\prime}\bar{F}^0\tilde{D}_T)\right\| \\ &\times \left\|(\tilde{D}_T\bar{F}^{0\prime}\bar{F}^0\tilde{D}_T)^{-1}\right\|\left\|\tilde{D}_T\bar{F}^{0\prime}V^*\right\| \\ &= o_p(T^{1/2}k_T^{-1/2})O_p(1)O_p(T^{-1}k_T^{1/2}v_T^{-1})O_p(1)o_p(T^{1/2}k_T^{-1/2}) \\ &= o_p(k_T^{-1/2}v_T^{-1}) = o_p(1).\end{aligned}$$

This proves that $V^{*\prime}(P_{\bar{F}^0} - P_{\bar{F}})V^* = o_p(1)$. □

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.

Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383–1414.

Bai, J., 1997. Estimation of a change point in multiple regression models. Review of Economics and Statistics 79, 551–563.

Bai, J., Lumsdaine, R.L., Stock, J.H., 1998. Testing for and dating breaks in multivariate time series. Review of Economic Studies 65, 395–432.
Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.
Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18, 1–22.
Cavaliere, G., Taylor, A.M.R., 2007. Testing for unit roots in time series models with non-stationary volatility. Journal of Econometrics 140, 919–947.
Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press, Oxford.
Dehling, H., Philipp, W., 1982. Almost sure invariance principles for weakly dependent vector-valued random variables. Annals of Probability 10, 689–701.
Gourieroux, C., Monfort, M., 1995. Statistics and Econometric Models, vol. 2. Cambridge University Press, Oxford.
Hájek, J., Rényi, A., 1955. Generalization of an inequality of Kolmogorov. Acta Mathematica Academiae Scientiarum Hungaricae 6, 281–283.
Hansen, B.E., 1992a. Consistent covariance matrix estimation for dependent heterogenous processes. Econometrica 60, 967–972.
Hansen, B.E., 1992b. Tests for parameter instability in regressions with I(1) processes. Journal of Business and Economic Statistics 10, 321–335.
Hao, K., 1996. Testing for structural change in cointegrated regression models: Some comparisons and generalizations. Econometric Reviews 15, 401–429.
Kejriwal, M., Perron, P., 2007. Data dependent rules for the selection of the number of leads and lags in the dynamic OLS cointegrating regression. Econometric Theory (in press).
Kejriwal, M., Perron, P., 2008. Testing for multiple structural changes in cointegrated regression models. Manuscript. Department of Economics, Boston University.
Kuo, B.-S., 1998. Test for partial parameter instability in regressions with I(1) processes. Journal of Econometrics 86, 337–368.
Lavielle, M., Moulines, E., 2000. Least-squares estimation of unknown number of shifts in a time series. Journal of Time Series Analysis 21, 33–59.
Lehmann, E.L., Romano, J.P., 2005. Testing Statistical Hypotheses. Springer, New York.
Ng, S., Perron, P., 1997. Estimation and inference in nearly unbalanced and nearly cointegrated systems. Journal of Econometrics 79, 53–81.
Perron, P., 2006. Dealing with structural breaks. In: Patterson, K., Mills, T.C. (Eds.), Econometric Theory. In: Palgrave Handbook of Econometrics, vol. 1. Palgrave Macmillan, pp. 278–352.
Perron, P., Qu, Z., 2006. Estimating restricted structural change models. Journal of Econometrics 134, 373–399.
Perron, P., Zhu, X., 2005. Structural breaks with deterministic and stochastic trends. Journal of Econometrics 129, 65–119.
Qu, Z., Perron, P., 2007. Estimating and testing structural changes in multivariate regressions. Econometrica 75, 459–502.
Saikkonen, P., 1991. Asymptotically efficient estimation of cointegration regressions. Econometric Theory 7, 1–21.
Seo, B., 1998. Tests for structural change in cointegrated systems. Econometric Theory 14, 222–259.
Stock, J.H., Watson, M.W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. Econometrica 61, 783–820.