Testing for Multiple Structural Changes in Cointegrated Regression Models

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

Pierre PERRON
Economics Department, Boston University, 270 Bay State Rd., Boston, MA 02215 (*perron@bu.edu*)

We consider testing for multiple structural changes in cointegrated systems and derive the limiting distribution of the sup-Wald test under mild conditions on the errors and regressors for a variety of testing problems. We show that even if the coefficients of the integrated regressors are held fixed but the intercept is allowed to change, the limit distributions are not the same as would prevail in a stationary framework. We also propose a sequential procedure that permits consistent estimation of the number of breaks present. We show via simulations that our tests maintain the correct size in finite samples and are much more powerful than the commonly used LM tests, which suffer from important problems of nonmonotonic power in the presence of serial correlation in the errors.

KEY WORDS: Change point; Cointegration; Sequential procedure; Unit root; Wald test.

1. INTRODUCTION

Issues related to structural change have received considerable attention in the statistics and econometrics literature. Andrews (1993) and Andrews and Ploberger (1994) provided a comprehensive treatment of the problem of testing for structural change assuming that the change point is unknown. Bai (1997) studied the least squares estimation of a single change point in regressions involving stationary and/or trending regressors, and derived the consistency, rate of convergence, and limiting distribution of the change point estimator under general conditions on the regressors and the errors. Perron and Zhu (2005) analyzed the properties of parameter estimates in models in which the trend function exhibits a slope change at an unknown date and the errors can either be stationary, $I(0)$, or have a unit root, $I(1)$, where here and throughout the article we refer to an $I(0)$ process as one whose partial sums satisfies a functional central limit theorem with a Brownian motion as the limit random variable, and $I(1)$ is the partial sum of an $I(0)$ series.

With integrated variables, the case of interest is when the variables are cointegrated. Accounting for parameter shifts is crucial in cointegration analysis, which normally involves long spans of data, which are more likely to be affected by structural breaks. Bai, Lumsdaine, and Stock (1998) considered a single break in a multiequation system. They showed consistency of the maximum likelihood estimates and obtained a limit distribution of the break date estimate under a shrinking shifts scenario. Kejriwal and Perron (2008b) studied the properties of the estimates of the break dates and parameters in a linear regression with multiple structural changes involving $I(1)$, $I(0)$, and trending regressors.

With respect to testing, Hansen (1992b) developed tests of the null hypothesis of no change in cointegrated models in which all coefficients are allowed to change. An extension to partial changes was analyzed by Kuo (1998), who considered the Sup and Mean LM tests directed against an alternative of a one-time change in parameters. Hao (1996) also suggested using the exponential LM test. Seo (1998) considered the sup, mean, and exp versions of the LM test within a cointegrated VAR setup; however, these test procedures are based on the fully modified estimation method (Phillips and Hansen 1990), which has been shown to lead to tests with very poor finite-sample properties (Carrion-i-Silvestre and Sansó-i-Rosselló 2006). The results of Quintos and Phillips (1993) also suggest that the LM tests are likely to suffer from the problem of low power in finite samples. Moreover, simulation experiments of Hansen (2000) have shown that the LM test behaves quite poorly in the presence of structural changes in the marginal distribution of the regressors, whereas the sup-Wald test is reasonably robust to such shifts. Hansen (2003) considered multiple structural changes in a cointegrated system, although his analysis was restricted to the case of known break dates. Finally, Qu (2007) proposed a procedure to detect whether cointegration is present when the cointegrating vector changes at some unknown date, possibly multiple dates.

The literature on testing for multiple structural changes is relatively sparse but is practically important, because single-break tests can suffer from nonmonotonic power when the alternative involves more than one break. As stressed by Perron (2006), most tests may exhibit nonmonotonic power functions if the number of breaks present is greater than the number explicitly accounted for in the construction of the tests. The aim of the present work is to provide a comprehensive treatment of issues related to testing for multiple structural changes occurring on unknown dates in cointegrated regression models. Our work builds on that of Bai and Perron (1998), who studied a similar treatment in a stationary context. Our framework is sufficiently general to allow both $I(0)$ and $I(1)$ variables in the regression. The assumptions about the distribution of the error processes

Journal of Business & Economic Statistics
October 2010, Vol. 28, No. 4
DOI: 10.1198/jbes.2009.07220

are sufficiently mild to allow for general forms of serial correlation. Moreover, we analyze both pure and partial structural change models. A partial change model is useful in allowing potential savings in degrees of freedom, a particularly relevant issue for multiple changes. It is also important in empirical work, because it helps isolate the variables responsible for the failure of the null hypothesis. We derive the limiting distribution of the sup-Wald test under the null hypothesis of no structural change against the alternative hypothesis of a given number of cointegrating regimes. We also consider the double-maximum tests proposed by Bai and Perron (1998). We provide critical values for a wide variety of models that are relevant in practice. Finally, our simulation experiments show that with serially correlated errors, the commonly used sup, mean, and exp-LM tests suffer from nonmonotonic power problems. This is true for cases with a single break as well as those with multiple breaks. We propose a modified sup Wald test that exhibits a power function that is monotonic with respect to the magnitude of the break(s) while maintaining reasonable size properties.

The article is organized as follows. Section 2 presents the model and assumptions. Section 3 describes the testing problems and the test statistics used. Section 4 presents our theoretical results regarding the limit distributions of the tests for a wide variety of cases. This is first done for models involving nontrending regressors, no serial correlation in the errors, and exogenous regressors. These restrictions are relaxed in Sections 4.2, 5.1, and 5.2, respectively. Asymptotic critical values are presented in Section 4.3. Section 6 presents simulation experiments that address issues related to the size and power of the tests, including a comparison with the often used LM tests. Section 7 provides some concluding remarks. The Appendix presents technical derivations.

2. THE MODEL AND ASSUMPTIONS

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

$$y_t = c_j + \mathbf{z}_{ft}'\boldsymbol{\delta}_f + \mathbf{z}_{bt}'\boldsymbol{\delta}_{bj} + \mathbf{x}_{ft}'\boldsymbol{\beta}_f + \mathbf{x}_{bt}'\boldsymbol{\beta}_{bj} + u_t \qquad (t = T_{j-1}+1, \ldots, T_j) \quad (1)$$

for $j = 1, \ldots, m+1$, where $T_0 = 0$, $T_{m+1} = T$, and T is the sample size. In this model, y_t is a scalar-dependent $I(1)$ variable, $\mathbf{x}_{ft}$ ($p_f \times 1$) and $\mathbf{x}_{bt}$ ($p_b \times 1$) are vectors of $I(0)$ variables, whereas $\mathbf{z}_{ft}$ ($q_f \times 1$) and $\mathbf{z}_{bt}$ ($q_b \times 1$) are vectors of $I(1)$ variables defined by $\mathbf{z}_{ft} = \mathbf{z}_{f,t-1} + \mathbf{u}_{zt}^f$, $\mathbf{z}_{bt} = \mathbf{z}_{b,t-1} + \mathbf{u}_{zt}^b$, $\mathbf{x}_{ft} = \boldsymbol{\mu}_f + \mathbf{u}_{xt}^f$, and $\mathbf{x}_{bt} = \boldsymbol{\mu}_b + \mathbf{u}_{xt}^b$, where for simplicity, $\mathbf{z}_{f0}$ and $\mathbf{z}_{b0}$ are assumed to be either $O_p(1)$ random variables or fixed finite constants. For ease of reference, the subscript b on the error term represents "break" and the subscript f represents "fixed" (across regimes). 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. When $p_f = q_f = 0$, we have a pure structural change model with all coefficients allowed to change across regimes. It will be useful to express (1) in matrix form as

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

where $\mathbf{Y} = (y_1, \ldots, y_T)'$, $\mathbf{G} = (\mathbf{Z}_f, \mathbf{X}_f)$, $\mathbf{Z}_f = (\mathbf{z}_{f1}, \ldots, \mathbf{z}_{fT})'$, $\mathbf{X}_f = (\mathbf{x}_{f1}, \ldots, \mathbf{x}_{fT})'$, $\mathbf{U} = (u_1, \ldots, u_T)'$, $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_T)'$, $\mathbf{w}_t = (1, \mathbf{z}_{bt}', \mathbf{x}_{bt}')'$, $\boldsymbol{\gamma} = (\boldsymbol{\delta}_{b1}', \boldsymbol{\beta}_{b1}', \ldots, \boldsymbol{\delta}_{b,m+1}', \boldsymbol{\beta}_{b,m+1}')'$, $\boldsymbol{\alpha} = (\boldsymbol{\delta}_f', \boldsymbol{\beta}_f')'$ and $\bar{\mathbf{W}}$ is the matrix which diagonally partitions $\mathbf{W}$ at the m-partition $(T_1, \ldots, T_m)$, that is, $\bar{\mathbf{W}} = \text{diag}(\mathbf{W}_1, \ldots, \mathbf{W}_{m+1})$ with $\mathbf{W}_i = (\mathbf{w}_{T_{i-1}+1}, \ldots, \mathbf{w}_{T_i})'$ for $i = 1, \ldots, m+1$. Kejriwal and Perron (2008b) analyzed the properties of the estimates of the break dates and the other parameters of the model under general conditions on the regressors and the errors. In this article, our interest lies in testing the null hypothesis of no structural change versus the alternative hypothesis of m changes as specified by the model (1). Thus the data-generating process is assumed to be given by (1) with $p_b = q_b = 0$.

As a matter of notation, "$\stackrel{p}{\to}$" denotes convergence in probability, "$\stackrel{d}{\to}$" denotes convergence in distribution, and "$\Rightarrow$" denotes weak convergence in the space $D[0,1]$ under the Skorohod metric. In addition, $\mathbf{x}_t = (\mathbf{x}_{ft}', \mathbf{x}_{bt}')'$, $\mathbf{u}_{xt} = (\mathbf{u}_{xt}^{f\prime}, \mathbf{u}_{xt}^{b\prime})'$, $\mathbf{z}_t = (\mathbf{z}_{ft}', \mathbf{z}_{bt}')'$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_f', \boldsymbol{\mu}_b')'$, and $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_m\}$ is the vector of break fractions defined by $\lambda_i = T_i/T$ for $i = 1, \ldots, m$. We make the following assumptions on $\boldsymbol{\xi}_t = (u_t, \mathbf{u}_{zt}^{f\prime}, \mathbf{u}_{zt}^{b\prime}, \mathbf{u}_{xt}^{f\prime}, \mathbf{u}_{xt}^{b\prime})'$, a vector of dimension $n = q_f + p_f + q_b + p_b + 1$.

Assumption A1. The vector $\boldsymbol{\xi}_t$ satisfies the following multivariate functional central limit theorem (FCLT): $T^{-1/2} \times \sum_{t=1}^{[Tr]} \boldsymbol{\xi}_t \Rightarrow \mathbf{B}(r)$, with $\mathbf{B}(r) = (B_1(r), \mathbf{B}_z^f(r)', \mathbf{B}_z^b(r)', \mathbf{B}_x^f(r)', \mathbf{B}_x^b(r)')'$ is a n vector Brownian motion with symmetric covariance matrix

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

$$= \lim_{T\to\infty} T^{-1} E(\mathbf{S}_T \mathbf{S}_T') = \boldsymbol{\Sigma} + \boldsymbol{\Lambda} + \boldsymbol{\Lambda}',$$

where $\mathbf{S}_T = \sum_{t=1}^T \boldsymbol{\xi}_t$, $\boldsymbol{\Sigma} = \lim_{T\to\infty} T^{-1} \sum_{t=1}^T E(\boldsymbol{\xi}_t \boldsymbol{\xi}_t')$, and $\boldsymbol{\Lambda} = \lim_{T\to\infty} T^{-1} \sum_{j=1}^{T-1} \sum_{t=1}^{T-j} E(\boldsymbol{\xi}_t \boldsymbol{\xi}_{t+j}')$. We also assume $\sigma^2 > 0$ and $p\lim_{T\to\infty} T^{-1} \sum_{t=1}^T u_t^2 = \lim_{T\to\infty} T^{-1} \sum_{t=1}^T E[u_t^2] \equiv \sigma_u^2$.

Assumption A2. The vector $\{\mathbf{x}_t u_t\}$ satisfies assumption A4 of Qu and Perron (2007), so that $T^{-1/2} \sum_{t=1}^{[Tr]} (\mathbf{u}_{xt}^f, \mathbf{u}_{xt}^b) u_t \Rightarrow \sigma \mathbf{Q}^{*1/2} \mathbf{W}_x^*(r)$, where $\mathbf{W}_x^*(r) = (\mathbf{W}_{xf}^*(r)', \mathbf{W}_{xb}^*(r)')'$ is a $(p_f + p_b)$ vector of independent Wiener processes, and

$$\mathbf{Q}^* = \begin{bmatrix} \mathbf{Q}_x^{ff*} & \mathbf{Q}_x^{fb*} \\ \mathbf{Q}_x^{bf*} & \mathbf{Q}_x^{bb*} \end{bmatrix}.$$

Assumption A3. For all t and s, (a) $E(\mathbf{u}_{xt} u_t \mathbf{z}_s') = 0$, (b) $E(\mathbf{u}_{xt} u_t u_s) = 0$, and (c) $E(\mathbf{u}_{xt} u_t \mathbf{u}_{xs}') = 0$.

Assumption A4. The matrix

$$\begin{pmatrix} \boldsymbol{\Omega}_{zz}^{ff} & \boldsymbol{\Omega}_{zz}^{fb} \\ \boldsymbol{\Omega}_{zz}^{bf} & \boldsymbol{\Omega}_{zz}^{bb} \end{pmatrix}$$

is positive definite.

Assumption A5. $T^{-1}\sum_{t=1}^{[Ts]}\mathbf{x}_t\mathbf{x}_t' \xrightarrow{p} s\mathbf{Q}$ and $T^{-1}\sum_{t=1}^{[Ts]}\mathbf{u}_{xt}\times\mathbf{u}_{xt}' \xrightarrow{p} s\mathbf{Q}^*$, uniformly in $s\in[0,1]$, for some positive definite matrices $\mathbf{Q}$ and $\mathbf{Q}^*$.

Assumption A1 requires that the errors satisfy a multivariate FCLT. The conditions required 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. Assumption A2 guarantees that an FCLT also holds for the sequence $\{\mathbf{u}_{xt}u_t\}$. Assumption A3 restricts somewhat the class of models applicable but is quite mild. Sufficient 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 $\mathbf{W}_x^*(\cdot)$ and $\mathbf{B}(\cdot)$ are uncorrelated and, being Gaussian, are thus independent. Without this condition, the analysis would be much more complex. Assumption A4 rules out cointegration among the $I(1)$ regressors. Assumption A5 is standard for $I(0)$ regressors but rules out trending regressors, which we relax in Section 4.2.

Under the alternative hypothesis, the estimates of the parameters are obtained by minimizing the global sum of squared residuals. For each m-partition $(T_1,\dots,T_m)$, denoted by $\{T_j\}$, the associated least squares estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are obtained by minimizing

$$SSR_T(T_1,\dots,T_m)=\sum_{i=1}^{m+1}\sum_{t=T_{i-1}+1}^{T_i}\big[y_t-c_i-\mathbf{z}_{ft}'\boldsymbol{\delta}_f-\mathbf{x}_{ft}'\boldsymbol{\beta}_f-\mathbf{z}_{bt}'\boldsymbol{\delta}_{bi}-\mathbf{x}_{bt}'\boldsymbol{\beta}_{bi}\big]^2. \quad (2)$$

Let $\hat{\boldsymbol{\alpha}}(\{T_j\})$ and $\hat{\boldsymbol{\gamma}}(\{T_j\})$ be the resulting estimates. Substituting these into the objective function and denoting the resulting sum of squared residuals as $S_T(T_1,\dots,T_m)$, the estimate of the break points are $(\hat T_1,\dots,\hat T_m)=\arg\min_{T_1,\dots,T_m}S_T(T_1,\dots,T_m)$, where the minimization is taken over all partitions $(T_1,\dots,T_m)$ such that $T_i-T_{i-1}\ge\epsilon T$ for some $\epsilon>0$. The estimates of the regression coefficients are then $\hat{\boldsymbol{\alpha}}=\hat{\boldsymbol{\alpha}}(\{\hat T_j\})$ and $\hat{\boldsymbol{\gamma}}=\hat{\boldsymbol{\gamma}}(\{\hat T_j\})$. Such estimates can be obtained using the algorithm of Bai and Perron (2003). Finally, consistent estimates of the matrixes $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ (and thus $\boldsymbol{\Omega}$) are $\hat{\boldsymbol{\Sigma}}=T^{-1}\sum_{t=1}^{T}\hat{\boldsymbol{\xi}}_t\hat{\boldsymbol{\xi}}_t'$ and $\hat{\boldsymbol{\Lambda}}=T^{-1}\sum_{j=1}^{T-1}w(j/l)\sum_{t=1}^{T-j}\hat{\boldsymbol{\xi}}_t\hat{\boldsymbol{\xi}}_{t+j}$, where $\hat{\boldsymbol{\xi}}_t=(\hat u_t,\Delta\mathbf{z}_{ft}',\Delta\mathbf{z}_{bt}',(\mathbf{x}_{ft}-\bar{\mathbf{x}}_f)',(\mathbf{x}_{bt}-\bar{\mathbf{x}}_b)')'$, with $\hat u_t$ the ordinary least squares (OLS) residuals from regression (1), $\bar{\mathbf{x}}_i=T^{-1}\sum_{t=1}^{T}\mathbf{x}_{it}$ $(i=f,b)$ and $w(j/l)$ is a kernel function that is continuous and even with $w(0)=1$ and $\int_{-\infty}^{\infty}w^2(x)\,dx<\infty$. In addition, $l\to\infty$ as $T\to\infty$ and $l=o(T^{1/2})$. Hansen (1992c) has demonstrated the consistency of these covariance matrix estimates.

3. THE TESTING PROBLEM AND THE TEST STATISTICS

The data-generating process (1) is the most general, and restricted versions may be used in practice. This gives rise to a variety of possible cases for the testing problems considered. We classify these into two categories: (a) models with only $I(1)$ regressors and (b) models with both $I(1)$ and $I(0)$ regressors. This classification into two categories is useful, because researchers often are faced with only $I(1)$ variables. For category (a), we consider the following testing problems (for ease of reference, we list the relevant regression under the alternative hypothesis):

Testing problems, category (a): models with $I(1)$ variables only ($p_f=p_b=0$, for all cases). Let H_0^a denotes the restrictions $\{c_j=c,\ \boldsymbol{\delta}_{bj}=\boldsymbol{\delta}_b$ for all $j=1,\dots,m+1\}$.

1. $H_0^a(1)=\{H_0^a,\ q_f=0\}$ versus $H_1^a(1)=\{q_f=0\}$ $(y_t=c_j+\mathbf{z}_{bt}'\boldsymbol{\delta}_{bj}+u_t)$;
2. $H_0^a(2)=\{H_0^a,\ q_b=0\}$ versus $H_1^a(2)=\{q_b=0\}$ $(y_t=c_j+\mathbf{z}_{ft}'\boldsymbol{\delta}_f+u_t)$;
3. $H_0^a(3)=\{H_0^a,\ q_f=0\}$ versus $H_1^a(3)=\{c_j=c$ for all $j=1,\dots,m+1,\ q_f=0\}$ $(y_t=c+\mathbf{z}_{bt}'\boldsymbol{\delta}_{bj}+u_t)$;
4. $H_0^a(4)=\{H_0^a\}$ versus $H_1^a(4)=\{$no restriction$\}$ $(y_t=c_j+\mathbf{z}_{ft}'\boldsymbol{\delta}_f+\mathbf{z}_{bt}'\boldsymbol{\delta}_{bj}+u_t)$;
5. $H_0^a(5)=\{H_0^a\}$ versus $H_1^a(5)=\{c_j=c$ for all $j=1,\dots,m+1\}$ $(y_t=c+\mathbf{z}_{ft}'\boldsymbol{\delta}_f+\mathbf{z}_{bt}'\boldsymbol{\delta}_{bj}+u_t)$.

Testing problems, category (b): models with both $I(1)$ and $I(0)$ variables. Let H_0^b denotes the restrictions $\{c_j=c,\ \boldsymbol{\delta}_{bj}=\boldsymbol{\delta}_b,\ \boldsymbol{\beta}_{bj}=\boldsymbol{\beta}_b$ for all $j=1,\dots,m+1\}$.

1. $H_0^b(1)=\{H_0^b,\ p_f=q_b=0\}$ versus $H_1^b(1)=\{c_j=c$ for all $j=1,\dots,m+1,\ p_f=q_b=0\}$ $(y_t=c+\mathbf{z}_{ft}'\boldsymbol{\delta}_f+\mathbf{x}_{bt}'\boldsymbol{\beta}_{bj}+u_t)$;
2. $H_0^b(2)=\{H_0^b,\ p_b=q_f=0\}$ versus $H_1^b(2)=\{c_j=c$ for all $j=1,\dots,m+1,\ p_b=q_f=0\}$ $(y_t=c+\mathbf{z}_{bt}'\boldsymbol{\delta}_{bj}+\mathbf{x}_{ft}'\boldsymbol{\beta}_f+u_t)$;
3. $H_0^b(3)=\{H_0^b,\ p_f=q_f=0\}$ versus $H_1^b(3)=\{c_j=c$ for all $j=1,\dots,m+1,\ p_f=q_f=0\}$ $(y_t=c+\mathbf{z}_{bt}'\boldsymbol{\delta}_{bj}+\mathbf{x}_{bt}'\boldsymbol{\beta}_{bj}+u_t)$;
4. $H_0^b(4)=\{H_0^b,\ p_f=q_f=0\}$ versus $H_1^b(4)=\{p_f=q_f=0\}$ $(y_t=c_j+\mathbf{z}_{bt}'\boldsymbol{\delta}_{bj}+\mathbf{x}_{bt}'\boldsymbol{\beta}_{bj}+u_t)$;
5. $H_0^b(5)=\{H_0^b,\ p_b=q_b=0\}$ versus $H_1^b(5)=\{p_b=q_b=0\}$ $(y_t=c_j+\mathbf{z}_{ft}'\boldsymbol{\delta}_f+\mathbf{x}_{ft}'\boldsymbol{\beta}_f+u_t)$;
6. $H_0^b(6)=\{H_0^b,\ p_b=q_f=0\}$ versus $H_1^b(6)=\{p_b=q_f=0\}$ $(y_t=c_j+\mathbf{z}_{bt}'\boldsymbol{\delta}_{bj}+\mathbf{x}_{ft}'\boldsymbol{\beta}_f+u_t)$;
7. $H_0^b(7)=\{H_0^b,\ p_f=q_b=0\}$ versus $H_1^b(7)=\{p_f=q_b=0\}$ $(y_t=c_j+\mathbf{z}_{ft}'\boldsymbol{\delta}_f+\mathbf{x}_{bt}'\boldsymbol{\beta}_{bj}+u_t)$;
8. $H_0^b(8)=\{H_0^b,\ q_f=0\}$ versus $H_1^b(8)=\{q_f=0\}$ $(y_t=c_j+\mathbf{z}_{bt}'\boldsymbol{\delta}_{bj}+\mathbf{x}_{ft}'\boldsymbol{\beta}_f+\mathbf{x}_{bt}'\boldsymbol{\beta}_{bj}+u_t)$;
9. $H_0^b(9)=\{H_0^b,\ q_b=0\}$ versus $H_1^b(9)=\{q_b=0\}$ $(y_t=c_j+\mathbf{z}_{ft}'\boldsymbol{\delta}_f+\mathbf{x}_{ft}'\boldsymbol{\beta}_f+\mathbf{x}_{bt}'\boldsymbol{\beta}_{bj}+u_t)$;
10. $H_0^b(10)=\{H_0^b\}$ versus $H_1^b(10)=\{$no restriction$\}$ $(y_t=c_j+\mathbf{z}_{ft}'\boldsymbol{\delta}_f+\mathbf{z}_{bt}'\boldsymbol{\delta}_{bj}+\mathbf{x}_{ft}'\boldsymbol{\beta}_f+\mathbf{x}_{bt}'\boldsymbol{\beta}_{bj}+u_t)$;
11. $H_0^b(11)=\{H_0^b\}$ versus $H_1^b(10)=\{c_j=c$ for all $j=1,\dots,m+1\}$ $(y_t=c+\mathbf{z}_{ft}'\boldsymbol{\delta}_f+\mathbf{z}_{bt}'\boldsymbol{\delta}_{bj}+\mathbf{x}_{ft}'\boldsymbol{\beta}_f+\mathbf{x}_{bt}'\boldsymbol{\beta}_{bj}+u_t)$.

We now give a brief description of each of the models in the two categories. In category (a), case 1 is a pure structural change model that allows for a change in the intercept as well. Case 2 is a partial change model in which only the intercept is allowed to change. Case 3 also is a partial change model but in which the intercept is not allowed to change. Cases 4 and 5 are block partial models in which a subset of the $I(1)$ coefficients is allowed to change. In category (b), cases 1–3 are partial change models in which the intercept is not allowed to change across regimes. Case 4 is a pure change model in which all $I(1)$ and $I(0)$ coefficients, as well as the intercept, are allowed to change. Case 5 is a partial change model that involves only an intercept shift. Case 6 is a partial change model in which the $I(0)$ coefficients are not allowed to change. Similarly, case 7 is a partial change model in which the $I(1)$ coefficients are not allowed to change. Cases 8–11 are block partial models in which a subset of coefficients of at least one type of regressor is not allowed to change.

We consider three types of tests. The first type of test applies when the alternative hypothesis involves a fixed value $m = k$ of changes. We consider the Wald test, scaled by the number of regressors whose coefficient are allowed to change, defined by

$$F_T(\boldsymbol{\lambda}, k) = \left(\frac{T-(k+1)(q_b+p_b)-(p_f+q_f)}{k}\right) \times \frac{\hat{\boldsymbol{\gamma}}'\mathbf{R}'(\mathbf{R}(\bar{\mathbf{W}}'\mathbf{M}_G\bar{\mathbf{W}})^{-1}\mathbf{R}')^{-1}\mathbf{R}\hat{\boldsymbol{\gamma}}}{SSR_k}, \quad (3)$$

where $\mathbf{R}$ is the conventional matrix such that $(\mathbf{R}\boldsymbol{\gamma})' = (\boldsymbol{\gamma}_1' - \boldsymbol{\gamma}_2', \ldots, \boldsymbol{\gamma}_k' - \boldsymbol{\gamma}_{k+1}')$ and $\mathbf{M}_G = \mathbf{I} - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$. Here SSR_k is the sum of squared residuals under the alternative hypothesis. Following Bai and Perron (1998), we define the following set for some arbitrary small positive number ϵ, $\Lambda_\epsilon^k = \{\boldsymbol{\lambda} : |\lambda_{i+1} - \lambda_i| \geq \epsilon, \lambda_1 \geq \epsilon, \lambda_k \leq 1-\epsilon\}$. The sup-Wald test is then defined as sup-$F_T(k) = \sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^k} F_T(\boldsymbol{\lambda}, k)$. Because in the current cases the estimates $\hat{\boldsymbol{\lambda}} = \{\hat{\lambda}_1, \ldots, \hat{\lambda}_k\}$ with $\hat{\lambda}_i = \hat{T}_i/T$ (for $i = 1, \ldots, k$) obtained by minimizing the global sum of squared residuals correspond to those that maximize the test $F_T(\boldsymbol{\lambda}, k)$, we have sup-$F_T(k) = F_T(\hat{\boldsymbol{\lambda}}, k)$.

The second type of test applies when the alternative hypothesis involves an unknown number of changes between 1 and some upper bound M. Again following Bai and Perron (1998), we consider a double-maximum test based on the maximum of the individual tests for the null of no break versus m breaks ($m = 1, \ldots, M$), defined by $UD\max F_T(M) = \max_{1\leq m\leq M} \sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^m} F_T(\boldsymbol{\lambda}, m)$. This test is arguably the most useful for determining the presence of structural changes. Simulations of Bai and Perron (2006) showed that with multiple changes, the power of tests for a single break can be quite low in finite samples, especially for certain types of multiple changes (e.g., two breaks with identical first and third regimes). In addition, tests for a particular number of changes may have nonmonotonic power when the number of changes is greater than specified. Finally, their simulations demonstrated that the power of $UD\max$ was nearly as high as that of the sup-F_T test based on the true number of changes.

The third testing procedure is a sequential one based on the estimates of the break dates obtained from a global minimization of sum of squared residuals, as described by Bai and Perron (1998). Consider a model with k breaks, with estimates denoted by $(\hat{T}_1, \ldots, \hat{T}_k)$, which are obtained through global minimization of the sum of squared residuals. Testing the null hypothesis of k breaks versus the alternative hypothesis of $k+1$ breaks involves performing a one-break test for each of the $(k+1)$ segments defined by the partition $(\hat{T}_1, \ldots, \hat{T}_k)$, then assessing whether the maximum of the tests is significant. More precisely, the test is defined by

$$SEQ_T(k+1|k) = \max_{1\leq j\leq k+1} \sup_{\tau\in\Lambda_{j,\varepsilon}} T\big\{SSR_T(\hat{T}_1, \ldots, \hat{T}_k) - SSR_T(\hat{T}_1, \ldots, \hat{T}_{j-1}, \tau, \hat{T}_j, \ldots, \hat{T}_k)\big\}/SSR_{k+1},$$

where $\Lambda_{j,\varepsilon} = \{\tau; \hat{T}_{j-1} + (\hat{T}_j - \hat{T}_{j-1})\varepsilon \leq \tau \leq \hat{T}_j - (\hat{T}_j - \hat{T}_{j-1})\varepsilon\}$. Note that this differs from a purely sequential procedure, because for each value of k, the break dates are reestimated to obtain those corresponding to the global minimizers of the sum of squared residuals.

4. ASYMPTOTIC DISTRIBUTIONS OF THE TESTS

An important issue that arises with integrated regressors is the correlation between the regressors and the errors. We first consider the case in which all $I(1)$ regressors are strictly exogenous. We then examine the case of endogenous regressors and show that if the regression is augmented with leads and lags of the the first differences of the $I(1)$ regressors, then the limiting distribution of the tests is the same as that obtained when all $I(1)$ regressors are strictly exogenous. Thus, for now we assume $\boldsymbol{\Omega}_{1z}^f = \boldsymbol{\Omega}_{1z}^b = \mathbf{0}$, which we later relax in Section 5.2. We also start with the following assumption, which imposes serially uncorrelated errors in the cointegrating regression, which we relax in Section 5.1:

Assumption A6. Let $\boldsymbol{\xi}_t^* = (\mathbf{u}_{zt}^{f\prime}, \mathbf{u}_{zt}^{b\prime}, \mathbf{u}_{xt}^{f\prime}, \mathbf{u}_{xt}^{b\prime})'$, the errors $\{u_t\}$ form an array of martingale differences relative to $\{\mathcal{F}_t\} = \sigma$-field$\{\boldsymbol{\xi}_{t-s}^*, u_{t-1-s}; s > 0\}$.

4.1 Main Theoretical Results

As a matter of notation, we define the following functionals, where $W_1 = \sigma^{-1}B_1$:

$$h(\mathbf{G}, a, b) = \left(\int_a^b \mathbf{G}\,dW_1\right)'\left(\int_a^b \mathbf{G}\mathbf{G}'\right)^{-1}\left(\int_a^b \mathbf{G}\,dW_1\right),$$

$$f(\mathbf{G}) = \left(\sum_{i=1}^{k+1}\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{G}\,dW_1\right)'\left(\sum_{i=1}^{k+1}\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{G}\mathbf{G}'\right)^{-1} \times \left(\sum_{i=1}^{k+1}\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{G}\,dW_1\right),$$

$g(\mathbf{G}, a, b) = (a\mathbf{G}(b) - b\mathbf{G}(a))'(a\mathbf{G}(b) - b\mathbf{G}(a))/ba(b-a)$, and $\mathbf{G}^{(a,b)}(r) = \mathbf{G}(r) - (\lambda_b - \lambda_{a-1})^{-1}\int_{\lambda_{a-1}}^{\lambda_b} \mathbf{G}$. In addition, by convention, $\lambda_0 = 0$ and $\lambda_{k+1} = 1$. The limit distributions of the tests when only $I(1)$ variables are involved are stated in the following theorem.

Theorem 1. Assume that Assumptions A1–A6 hold and that $\boldsymbol{\Omega}_{1z}^f = \boldsymbol{\Omega}_{1z}^b = \mathbf{0}$. For the testing problems in category (a), the limit distribution of $\sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^k} F_T(\boldsymbol{\lambda}, k)$ is $\sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^k} F(\boldsymbol{\lambda}, k)/k$ with

$F(\boldsymbol{\lambda}, k)$ defined as follows for the various cases: For case 1,

$$F(\boldsymbol{\lambda}, k) = \sum_{i=1}^{k}\big[h\big(\mathbf{W}_z^{b(1,i)}, 0, \lambda_i\big) - h\big(\mathbf{W}_z^{b(1,i+1)}, 0, \lambda_{i+1}\big) + h\big(\mathbf{W}_z^{b(i+1,i+1)}, \lambda_i, \lambda_{i+1}\big) + g\big(W_1, \lambda_i, \lambda_{i+1}\big)\big].$$

For case 2, $F(\boldsymbol{\lambda}, k) = f(\mathbf{W}_z^{f(i,i)}) - h(\mathbf{W}_z^{f(1,k+1)}, 0, 1) + \sum_{i=1}^{k} g(W_1, \lambda_i, \lambda_{i+1})$, where $\mathbf{W}_z^f(r) = (\boldsymbol{\Omega}_{zz}^{ff})^{-1/2}\mathbf{B}_z^f(r)$. For case 3,

$$F(\boldsymbol{\lambda}, k) = f(\mathbf{P}_{zi}^b) - h\big(\mathbf{W}_z^{b(1,k+1)}, 0, 1\big) - W_1(1)^2 + \sum_{i=1}^{k+1} h(\mathbf{W}_z^b, \lambda_{i-1}, \lambda_i),$$

where $\mathbf{P}_{zi}^b(r) = 1 - (\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{b\prime})(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^b \mathbf{W}_z^{b\prime})^{-1}\mathbf{W}_z^b(r)$, for $r \in [\lambda_{i-1}, \lambda_i]$. For case 4,

$$F(\boldsymbol{\lambda}, k) = f\big(\mathbf{W}_z^{M(i,i)}\big) - h\big(\mathbf{W}_z^{fb(1,k+1)}, 0, 1\big) + \sum_{i=1}^{k+1} h\big(\mathbf{W}_z^{b(i,i)}, \lambda_{i-1}, \lambda_i\big) + \sum_{i=1}^{k} g(W_1, \lambda_i, \lambda_{i+1}),$$

with $\mathbf{W}_z^{fb}(r) = (\mathbf{W}_z^f(r), \mathbf{W}_z^b(r))$, and where

$$\mathbf{W}_z^{M(i,i)}(r) = \mathbf{W}_z^{f(i,i)}(r) - \int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{f(i,i)}\mathbf{W}_z^{b(i,i)\prime} \times \left(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{b(i,i)}\mathbf{W}_z^{b(i,i)\prime}\right)^{-1} \mathbf{W}_z^{b(i,i)}(r).$$

For case 5, $F(\boldsymbol{\lambda}, k) = f(\mathbf{P}_{zi}) - h(\mathbf{W}_z^{fb(1,k+1)}, 0, 1) - W_1(1)^2 + \sum_{i=1}^{k+1} h(\mathbf{W}_z^b, \lambda_{i-1}, \lambda_i)$, where $\mathbf{P}_{zi}(r)' = (\mathbf{P}_{zi}^b(r)', \mathbf{P}_{zi}^{fb}(r)')$ with $\mathbf{P}_{zi}^{fb}(r) = \mathbf{W}_z^f(r) - (\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^f \mathbf{W}_z^{b\prime})(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^b \mathbf{W}_z^{b\prime})^{-1}\mathbf{W}_z^b(r)$.

Theorem 1 shows that it is possible to make inference in models involving $I(1)$ variables using the sup-Wald test. Moreover, the limiting distributions differ depending on whether the intercept and/or the $I(1)$ coefficients are allowed to change. Note that for cases 2, 4, and 5, the limit distributions depend on the number of $I(1)$ coefficients that are not allowed to change. This is different from a stationary framework, in which the limit distribution is independent of the number of regressors whose coefficients are not allowed to change. We now consider the limit distributions of the test for the various cases in category (b) in which both $I(1)$ and $I(0)$ regressors are present.

Theorem 2. Assume that Assumptions A1–A6 hold and that $\boldsymbol{\Omega}_{1z}^f = \boldsymbol{\Omega}_{1z}^b = \mathbf{0}$ and let $\mathbf{W}_{xb(1)}^* = (\mathbf{W}_{xb}^{*\prime}, W_1)'$. For the cases in category (b), the limiting distributions of $\sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^k} F_T(\boldsymbol{\lambda}, k)$ under the null hypothesis are given by $\sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^k} F(\boldsymbol{\lambda}, k)/k$ with $F(\boldsymbol{\lambda}, k)$ defined as follows: For case 1, $F(\boldsymbol{\lambda}, k) = \sum_{i=1}^{k} g(\mathbf{W}_{xb}^*, \lambda_i, \lambda_{i+1})$. For case 2, the limit distribution is the same as for case 3 in category (a). For case 3,

$$F(\boldsymbol{\lambda}, k) = f(\mathbf{P}_{zi}^b) - h\big(\mathbf{W}_z^{b(1,k+1)}, 0, 1\big) - W_1(1)^2 + \sum_{i=1}^{k+1} h(\mathbf{W}_z^b, \lambda_{i-1}, \lambda_i) + \sum_{i=1}^{k} g(\mathbf{W}_{xb}^*, \lambda_i, \lambda_{i+1}).$$

For cases 4 and 8,

$$F(\boldsymbol{\lambda}, k) = \sum_{i=1}^{k}\big[h\big(\mathbf{W}_z^{b(1,i)}, 0, \lambda_i\big) - h\big(\mathbf{W}_z^{b(1,i+1)}, 0, \lambda_{i+1}\big) + h\big(\mathbf{W}_z^{b(i+1,i+1)}, \lambda_i, \lambda_{i+1}\big) + g\big(\mathbf{W}_{xb(1)}^*, \lambda_i, \lambda_{i+1}\big)\big].$$

For cases 5 and 6, the limit distributions are the same as for cases 1 and 2, respectively, in category (a). For cases 7 and 9,

$$F(\boldsymbol{\lambda}, k) = f\big(\mathbf{W}_z^{f(i,i)}\big) - h\big(\mathbf{W}_z^{f(1,k+1)}, 0, 1\big) + \sum_{i=1}^{k} g\big(\mathbf{W}_{xb(1)}^*, \lambda_i, \lambda_{i+1}\big).$$

For case 10,

$$F(\boldsymbol{\lambda}, k) = f\big(\mathbf{W}_z^{M(i,i)}\big) - h\big(\mathbf{W}_z^{fb(1,k+1)}, 0, 1\big) + \sum_{i=1}^{k+1} h\big(\mathbf{W}_z^{b(i,i)}, \lambda_{i-1}, \lambda_i\big) + \sum_{i=1}^{k} g\big(\mathbf{W}_{xb(1)}^*, \lambda_i, \lambda_{i+1}\big).$$

For case 11,

$$F(\boldsymbol{\lambda}, k) = f(\mathbf{P}_{zi}) - h\big(\mathbf{W}_z^{fb(1,k+1)}, 0, 1\big) - W_1(1)^2 + \sum_{i=1}^{k+1} h(\mathbf{W}_z^b, \lambda_{i-1}, \lambda_i) + \sum_{i=1}^{k} g(\mathbf{W}_{xb}^*, \lambda_i, \lambda_{i+1}).$$

The practical implications of Theorem 2 are as follows. As shown in case 1, if the intercept and the $I(1)$ variables are held fixed and only the coefficients on the $I(0)$ variables are allowed to change, then the same limit distribution as given by Bai and Perron (1998) applies. But this equivalence with the case of stationary regressors holds only if the constant is not allowed to change. As shown in case 7, the limit distribution differs when the intercept is allowed to change and depends on the number of $I(1)$ variables present. The effect of allowing or not allowing the intercept to change can also be seen by comparing cases 3 and 4. The limit distributions are different, and, as expected, both depend on the number of $I(1)$ and $I(0)$ variables whose coefficients are allowed to change. A similar feature also applies when the regression involves $I(1)$ and $I(0)$ variables whose coefficients are not allowed to change, as shown in cases 10 and 11. Comparing these with cases 3 and 4 again shows that having $I(1)$ variables whose coefficients are not allowed to change alters the limit distributions. Finally, comparing cases a1 and b6, a2 and b5, a3 and b2, b4 and b8, and b7 and b9 shows that including $I(0)$ regressors whose coefficients are not allowed to change does not alter the limit distribution.

Remark 1. For case 4 in category (b), the limit distribution of $\sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^k} F_T(\boldsymbol{\lambda}, k)$ is:

$$\sup_{(\lambda_1,\ldots,\lambda_k)\in\Lambda_\epsilon^k}\Bigg\{\sum_{i=1}^{k}(\mathbf{S}^*(\lambda_i, \lambda_{i+1})'\mathbf{V}(\lambda_i, \lambda_{i+1})^{-1}\mathbf{S}^*(\lambda_i, \lambda_{i+1})) + \sum_{i=1}^{k}\big((\lambda_i \mathbf{W}_{xb}^*(\lambda_{i+1}) - \lambda_{i+1}\mathbf{W}_{xb}^*(\lambda_i))' \times (\lambda_i \mathbf{W}_{xb}^*(\lambda_{i+1}) - \lambda_{i+1}\mathbf{W}_{xb}^*(\lambda_i))\big)/(\lambda_{i+1}\lambda_i(\lambda_{i+1} - \lambda_i))\Bigg\}$$

with $\mathbf{S}^*(\lambda_i, \lambda_{i+1}) = \mathbf{S}(\lambda_i) - \mathbf{M}(\lambda_i)\mathbf{M}(\lambda_{i+1})^{-1}\mathbf{S}(\lambda_{i+1})$, $\mathbf{V}(\lambda_i, \lambda_{i+1}) = \mathbf{M}(\lambda_i) - \mathbf{M}(\lambda_i)\mathbf{M}(\lambda_{i+1})^{-1}\mathbf{M}(\lambda_i)$, $\mathbf{S}(\lambda_i) = \int_0^{\lambda_i} \mathbf{Z}^* \, dW_1$, $\mathbf{M}(\lambda_i) = \int_0^{\lambda_i} \mathbf{Z}^*\mathbf{Z}^{*\prime}$, and $\mathbf{Z}^* = (1, \mathbf{W}_z^{b\prime})'$. The first summation corresponds to the distribution in case 1 of category (a), whereas the second summation corresponds to the p_b $I(0)$ regressors whose coefficients are allowed to change.

With these theoretical results for the sup-$F_T(\boldsymbol{\lambda}, k)$, we can obtain the limit distribution of the $UD\max$ and $SEQ_T(k+1|k)$ tests. These are stated in the following corollary.

Corollary 1. Under Assumptions A1–A6 and $\boldsymbol{\Omega}_{1z}^f = \boldsymbol{\Omega}_{1z}^b = \mathbf{0}$, for a particular testing problem denote the limit distribution of the test $\sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^k} F_T(\boldsymbol{\lambda}, k)$ by $\sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^k} F(\boldsymbol{\lambda}, k)/k$, then: (a) $UD\max F_T(M) = \max_{1\le m\le M} \sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^m} F_T(\boldsymbol{\lambda}, m) \Rightarrow \max_{1\le m\le M} \sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^m} F(\boldsymbol{\lambda}, m)/m$, (b) $\lim_{T\to\infty} P(SEQ_T(k+1|k) \le x) = G_\varepsilon(x)^{k+1}$, with $G_\varepsilon(x)$ the distribution function of $\sup_{\lambda\in\Lambda_\varepsilon^1} F(\lambda, 1)$.

4.2 Trends in Regressors

Suppose now that the $I(1)$ regressors have a trending nonstochastic component, that is, are generated by $\mathbf{z}_{ft}^* = \boldsymbol{\rho}_f t + \mathbf{z}_{ft}$ and $\mathbf{z}_{bt}^* = \boldsymbol{\rho}_b t + \mathbf{z}_{bt}$, with $q_b > 1$ and $\rho_b \neq 0$. The limiting distributions of the tests then differ from those in the nontrending case. The derivation of the required modifications follows the treatment of Hansen (1992a). Consider a $q_b \times (q_b - 1)$ matrix $\boldsymbol{\rho}_b^*$ that spans the null space of $\boldsymbol{\rho}_b$ and let $\mathbf{C}_2 = [\mathbf{C}_{12}, \mathbf{C}_{22}] = (\boldsymbol{\rho}_b(\boldsymbol{\rho}_b'\boldsymbol{\rho}_b)^{-1}, \boldsymbol{\rho}_b^*(\boldsymbol{\rho}_b^{*\prime}\boldsymbol{\Omega}_{zz}^{bb}\boldsymbol{\rho}_b^*)^{-1/2})$. Note that $\mathbf{C}_2'\mathbf{z}_{bt}^* = (\mathbf{C}_{12}'\mathbf{z}_{bt} + t, \mathbf{C}_{22}'\mathbf{z}_{bt})'$. With $\bar{\mathbf{W}}_{2T} = \mathrm{diag}(T, \mathbf{I}_{q_b-1}T^{1/2})$, we have

$$\bar{\mathbf{W}}_{2T}^{-1}\mathbf{C}_2'\mathbf{z}_{b[Tr]} = \begin{pmatrix} T^{-1}\mathbf{C}_{12}'\mathbf{z}_{b[Tr]} + T^{-1}[Tr] \\ T^{-1/2}\mathbf{C}_{22}'\mathbf{z}_{b[Tr]} \end{pmatrix} \Rightarrow \begin{pmatrix} r \\ \mathbf{W}_{z(-1)}^b(r) \end{pmatrix} \equiv \mathbf{J}_z^b(r), \quad (4)$$

where $\mathbf{W}_{z(-1)}^b(r)$ is a $(q_b - 1)$-dimensional vector of independent Wiener processes [a linear combination of $\mathbf{W}_z^b(r)$]. Note that when $q_b = 1$, $\mathbf{W}_{z(-1)}^b(r) = r$. It then follows that

$$T^{-1}\bar{\mathbf{W}}_{2T}^{-1}\mathbf{C}_2'\sum_{t=1}^{[Tr]}\mathbf{z}_{bt}^*\mathbf{z}_{bt}^{*\prime}\mathbf{C}_2\bar{\mathbf{W}}_{2T}^{-1} \Rightarrow \int_0^r \mathbf{J}_z^b\mathbf{J}_z^{b\prime}, \quad (5)$$

$$T^{-1/2}\bar{\mathbf{W}}_{2T}^{-1}\mathbf{C}_2'\sum_{t=1}^{[Tr]}\mathbf{z}_{bt}^* u_t \Rightarrow \sigma\int_0^r \mathbf{J}_z^b \, dW_1. \quad (6)$$

Note that (4) through (6) also hold for $\mathbf{z}_{ft}^*$with $\mathbf{W}_{z(-1)}^b(r)$ replaced by $\mathbf{W}_{z(-1)}^f(r)$, a $(q_f - 1)$ dimensional vector of independent Wiener processes [a linear combination of $\mathbf{W}_z^f(r)$]. Here also, when $q_f = 1$, $\mathbf{W}_{z(-1)}^f(r) = r$. Therefore, with trending regressors, the limiting distributions of the tests are not the same as those without trends; however, we can obtain them by simply replacing $\mathbf{W}_z^f$ and $\mathbf{W}_x^b$ by $\mathbf{J}_z^f$ and $\mathbf{J}_z^b$.

4.3 Asymptotic Critical Values

Because the asymptotic distributions are nonstandard, we obtain critical values through simulations for models with and without trends in regressors. We approximate the Wiener processes by partial sums of iid normal random variables with $N = 500$ steps. The number of replications is 2,000. For each replication, the supremum of $F(\boldsymbol{\lambda}, k)$ with respect to $(\lambda_1, \ldots, \lambda_k)$ over the set Λ_ϵ^k is obtained via a dynamic programming algorithm (see Bai and Perron 2003). The $I(0)$ regressors are simulated as independent sequences of iid N(0, 1) variables, and the $I(1)$ regressors as independent random walks with iid N(0, 1) errors [also independent of the $I(0)$ regressors]. The trimming values used are $\epsilon = 0.05, 0.10, 0.15, 0.20$, and 0.25. Critical values are presented for up to nine breaks and four regressors. The maximum number of breaks allowed is eight when $\epsilon = 0.10$, five when $\epsilon = 0.15$, three when $\epsilon = 0.20$ and two when $\epsilon = 0.25$. For the $UD\max$ test, M is set to 5 or the maximum number of breaks possible. For models involving both $I(1)$ and $I(0)$ variables, critical values are provided for all possible permutations up to two regressors of each type. For the limit distributions of the tests with trending regressors and for the sequential tests, we tabulated the critical values for $\epsilon = 0.15, 0.20$, and 0.25. Because of the large number of results, we present critical values only for cases that allow the intercept to change and for $\epsilon = 0.15$ in Tables 1–4. For other cases and trimming values, tables of critical values are available on our website.

5. EXTENSIONS

We now extend the analysis of the previous section to the cases in which we can have either serially correlated errors in the cointegrating regression or endogenous regressors. We show that simple modifications yield tests with the same limit distributions described earlier.

5.1 Serially Correlated Errors: A Modified sup-Wald Test

With serially correlated errors, we use the following robust version of the scaled F test:

$$F_T^*(\boldsymbol{\lambda}, k) = \frac{(T - (k+1)(q_b + p_b) - (q_f + p_f))}{k} \times \hat{\boldsymbol{\gamma}}'\mathbf{R}'(\mathbf{R}T\hat{\boldsymbol{V}}(\hat{\boldsymbol{\gamma}})\mathbf{R}')^{-1}\mathbf{R}\hat{\boldsymbol{\gamma}}, \quad (7)$$

where $\hat{\mathbf{V}}(\hat{\boldsymbol{\gamma}})$ is an estimate of the covariance matrix of $\hat{\boldsymbol{\gamma}}$ that is robust to serial correlation and heteroscedasticity (see Bai and Perron 1998 for details). Note that when testing for the stability of coefficients associated with $I(1)$ variables, whether or not $I(0)$ variables are included, we can simply apply the following transformation to the test in (3): $F_T^*(\boldsymbol{\lambda}; k) = (\hat{\sigma}_u^2/\hat{\sigma}^2)F_T(\boldsymbol{\lambda}, k)$, where $\hat{\sigma}_u^2 = T^{-1}\sum_{t=1}^T \hat{u}_t^2$ and $\hat{\sigma}^2$ is a consistent estimate of σ^2. Because the break fractions are consistent even with serially correlated errors, we can first take the supremum of the original F test to obtain the break points, then obtain the robust version of the test by evaluating $F_T^*(\boldsymbol{\lambda}; k)$ at these estimated break dates. That is, the test considered is $\sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^k} F_T^*(\boldsymbol{\lambda}, k) = F_T^*(\hat{\boldsymbol{\lambda}}, k)$, where $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_k)$ are the estimates of the break fractions obtained by minimizing the global sum of squared residuals (2).

A problem with the sup-Wald test is that with persistent errors, the size distortions can be substantial, due to the estimation of the long-run variance using residuals under the

Table 1. Asymptotic critical values [the entries are quantiles x such that $P(\sup F(\boldsymbol{\lambda}, k)/k \le x) = \alpha$]

		Non trending case						Trending case					
		Number of breaks, k						Number of breaks, k					
q_b	α	1	2	3	4	5	$UD\max$	1	2	3	4	5	$UD\max$
Category (a) case 1, $\epsilon = 0.15$													
1	0.90	10.34	8.85	7.66	6.66	5.30	10.53	11.18	9.25	8.09	6.95	5.53	11.33
	0.95	12.11	9.96	8.60	7.36	5.90	12.25	13.03	10.39	8.94	7.60	6.12	13.07
	0.975	13.85	11.41	9.40	7.99	6.42	13.91	15.08	11.49	9.66	8.28	6.67	15.13
	0.99	17.03	12.41	10.40	8.71	7.08	17.40	16.86	12.73	10.82	8.95	7.32	16.86
2	0.90	12.36	11.01	9.60	8.45	6.96	12.64	11.88	10.31	9.00	7.98	6.62	12.13
	0.95	14.30	12.11	10.41	9.19	7.64	14.47	13.63	11.34	9.94	8.68	7.31	13.99
	0.975	15.72	13.37	11.26	9.75	8.15	15.90	15.51	12.57	10.86	9.37	7.92	15.53
	0.99	17.67	14.73	12.21	10.77	8.82	17.67	17.31	14.63	12.10	10.51	8.73	17.31
3	0.90	14.88	12.84	11.49	10.19	8.53	15.09	14.39	12.14	10.79	9.61	8.22	14.65
	0.95	16.66	14.11	12.38	10.94	9.12	16.71	16.50	13.22	11.66	10.33	8.92	16.61
	0.975	18.32	15.24	13.01	11.52	9.61	18.35	18.08	14.45	12.54	11.04	9.44	18.24
	0.99	20.78	16.29	14.36	12.37	10.23	20.78	20.28	15.55	13.80	12.02	10.10	20.28
4	0.90	16.87	14.72	13.20	11.75	9.90	17.05	16.27	13.80	12.41	11.17	9.62	16.46
	0.95	19.08	15.90	14.15	12.68	10.72	19.16	18.36	15.08	13.38	12.07	10.28	18.46
	0.975	20.81	17.15	15.21	13.38	11.43	20.89	20.52	17.01	14.33	12.98	10.93	20.52
	0.99	22.59	18.85	16.44	14.25	11.98	22.59	23.12	18.71	15.77	13.87	11.72	23.12
q_f													
Category (a) case 2, $\epsilon = 0.15$													
1	0.90	7.52	6.38	5.37	4.54	3.49	7.79	8.67	6.84	6.07	5.31	4.01	8.90
	0.95	9.26	7.30	6.21	5.19	3.98	9.38	10.29	7.89	6.85	5.97	4.49	10.44
	0.975	10.63	8.25	6.98	5.67	4.40	10.87	12.18	8.99	7.57	6.66	5.02	12.18
	0.99	12.57	10.01	7.77	6.42	4.88	12.60	14.21	10.19	8.45	7.10	5.62	14.27
2	0.90	8.48	6.70	5.66	4.77	3.63	8.66	8.32	6.49	5.65	4.98	3.84	8.60
	0.95	10.13	7.66	6.43	5.36	4.10	10.25	10.06	7.45	6.42	5.67	4.36	10.11
	0.975	11.69	8.85	7.34	5.99	4.62	11.82	11.47	8.59	7.21	6.29	5.02	11.52
	0.99	13.66	10.20	8.09	6.91	5.35	13.66	13.21	9.86	8.29	7.01	5.49	13.30
3	0.90	8.47	6.51	5.59	4.77	3.58	8.74	8.40	6.53	5.64	5.03	3.91	8.66
	0.95	10.08	7.61	6.26	5.49	4.07	10.26	10.08	7.48	6.35	5.65	4.35	10.10
	0.975	11.27	8.51	7.21	6.12	4.49	11.43	11.68	8.55	6.90	6.15	4.83	11.68
	0.99	12.88	9.95	7.88	6.70	5.13	12.93	13.72	9.53	7.51	6.72	5.34	13.72
4	0.90	8.56	6.59	5.71	4.87	3.81	8.85	8.57	6.49	5.69	4.94	3.85	8.69
	0.95	10.07	7.66	6.52	5.55	4.30	10.17	10.22	7.34	6.51	5.59	4.46	10.36
	0.975	11.69	8.61	7.10	6.09	4.70	11.69	11.90	8.33	7.22	6.26	4.88	11.95
	0.99	13.88	9.64	7.83	6.58	5.33	13.88	14.53	9.68	8.33	6.97	5.53	14.53
q_f, q_b													
Category (a) case 4, $\epsilon = 0.15$													
1, 1	0.90	10.19	8.77	7.74	6.60	5.26	10.53	10.81	9.18	7.99	6.89	5.48	10.98
	0.95	12.03	9.78	8.53	7.18	5.81	12.30	12.27	10.30	8.87	7.61	6.09	12.34
	0.975	14.05	11.03	9.28	7.92	6.30	14.07	14.43	11.39	9.54	8.28	6.72	14.45
	0.99	16.02	12.33	10.33	8.67	6.99	16.09	16.65	12.56	10.45	9.02	7.14	16.65
1, 2	0.90	12.89	11.03	9.70	8.60	7.02	13.16	12.57	10.62	9.17	8.17	6.80	12.76
	0.95	14.88	12.27	10.76	9.38	7.68	14.97	14.19	11.69	10.12	8.93	7.43	14.27
	0.975	16.72	13.67	11.63	10.03	8.48	16.75	15.86	12.73	10.78	9.51	7.85	15.89
	0.99	18.48	14.72	12.48	10.89	9.06	18.48	17.89	13.79	11.76	10.18	8.39	18.16

Table 1. (Continued)

		Non trending case						Trending case					
		Number of breaks, k						Number of breaks, k					
q_f, q_b	α	1	2	3	4	5	$UD\max$	1	2	3	4	5	$UD\max$
2, 1	0.90	10.99	9.08	7.91	6.82	5.46	11.15	11.33	9.36	8.07	7.04	5.66	11.45
	0.95	13.04	10.09	8.71	7.43	6.02	13.06	13.18	10.46	9.09	7.73	6.21	13.26
	0.975	14.80	10.84	9.46	8.01	6.60	14.80	15.22	11.55	9.80	8.33	6.71	15.22
	0.99	16.46	12.08	10.43	8.87	7.04	16.46	17.85	12.48	10.49	9.08	7.32	17.85
2, 2	0.90	12.87	11.04	9.71	8.58	7.12	13.07	12.58	10.41	9.15	8.15	6.78	12.78
	0.95	14.81	12.25	10.75	9.44	7.74	15.01	14.65	11.78	10.04	8.85	7.48	14.72
	0.975	16.74	13.48	11.57	10.15	8.34	16.74	15.95	12.92	10.94	9.57	8.04	16.12
	0.99	19.36	14.78	12.29	10.83	8.78	19.36	17.94	13.91	11.83	10.32	8.91	18.08

Table 2. Asymptotic critical values [the entries are quantiles x such that $P(\sup F(\boldsymbol{\lambda}, k)/k \leq x) = \alpha$]

		Non trending case						Trending case					
		Number of breaks, k						Number of breaks, k					
q_b, p_b	α	1	2	3	4	5	$UD\max$	1	2	3	4	5	$UD\max$
							Category (b) cases 4 and 8, $\epsilon = 0.15$						
1, 1	0.90	11.69	9.88	8.63	7.52	6.27	11.99	11.98	10.29	8.96	7.83	6.63	12.27
	0.95	13.24	10.96	9.62	8.29	6.87	13.43	13.74	11.64	9.92	8.66	7.28	14.06
	0.975	14.78	12.10	10.54	8.99	7.56	14.87	15.86	12.85	10.87	9.30	7.87	15.91
	0.99	17.28	13.40	11.53	9.75	8.11	17.39	17.99	14.27	11.87	10.20	8.44	17.99
1, 2	0.90	12.88	11.06	9.55	8.53	7.52	13.26	13.24	11.17	9.79	8.85	7.69	13.51
	0.95	15.10	12.13	10.53	9.42	8.16	15.25	15.16	12.19	10.85	9.61	8.29	15.20
	0.975	17.51	13.04	11.30	9.98	8.71	17.60	16.89	13.33	11.59	10.48	8.87	16.89
	0.99	19.10	14.68	12.35	11.07	9.51	19.10	18.95	14.43	12.79	11.23	9.90	18.95
2, 1	0.90	13.85	12.05	10.48	9.35	7.99	14.23	13.42	11.33	10.06	9.00	7.73	13.64
	0.95	15.91	13.45	11.50	10.23	8.64	16.07	15.42	12.76	11.03	9.86	8.44	15.47
	0.975	17.68	14.60	12.44	11.06	9.30	18.06	17.50	13.95	12.05	10.58	8.97	17.50
	0.99	19.89	16.02	13.80	11.88	10.14	20.03	19.61	15.23	13.05	11.38	9.59	19.61
2, 2	0.90	14.82	13.09	11.64	10.40	9.04	15.24	14.91	12.50	11.14	10.06	8.83	15.28
	0.95	17.02	14.49	12.51	11.19	9.73	17.33	17.17	14.02	12.23	10.91	9.59	17.22
	0.975	19.59	15.57	13.39	11.85	10.29	19.59	19.48	15.41	13.18	11.57	10.23	19.48
	0.99	21.66	17.07	14.35	12.81	10.85	21.66	21.46	16.50	14.18	12.60	10.82	21.46
q_f, q_b													
							Category (b) cases 7 and 9, $\epsilon = 0.15$						
1, 1	0.90	8.72	7.48	6.23	5.41	4.52	9.12	8.38	6.72	5.82	5.15	4.29	8.64
	0.95	10.65	8.59	6.97	6.13	5.06	10.87	10.16	7.93	6.82	5.76	4.73	10.34
	0.975	12.13	9.61	7.92	6.68	5.50	12.39	11.95	9.18	7.52	6.32	5.34	11.99
	0.99	14.37	10.75	9.10	7.76	6.32	14.95	13.88	10.40	8.26	6.99	6.09	13.88
1, 2	0.90	9.95	8.17	7.17	6.50	5.63	10.31	9.35	7.38	6.58	5.93	5.31	9.62
	0.95	11.58	9.54	8.25	7.23	6.25	11.93	10.98	8.60	7.32	6.61	5.92	11.07
	0.975	12.99	10.74	9.23	7.83	6.85	13.68	12.76	9.59	8.24	7.35	6.48	12.83
	0.99	15.66	12.19	10.30	8.65	7.71	15.68	15.22	10.92	9.55	8.20	7.16	15.22
2, 1	0.90	9.03	7.51	6.45	5.70	4.66	9.49	8.96	6.80	5.94	5.19	4.41	9.08
	0.95	10.70	8.77	7.34	6.32	5.22	10.85	10.56	7.90	6.84	5.85	5.00	10.73
	0.975	11.98	9.77	7.98	6.98	5.70	12.30	12.50	8.99	7.48	6.53	5.46	12.55
	0.99	15.29	10.80	8.95	7.71	6.32	15.29	14.98	9.87	8.53	7.08	6.03	14.98
2, 2	0.90	10.58	8.52	7.36	6.64	5.78	10.88	9.82	7.95	7.00	6.31	5.50	10.33
	0.95	12.32	9.72	8.23	7.45	6.39	12.53	11.82	9.26	7.88	7.09	6.20	12.09
	0.975	14.09	11.05	9.36	8.23	6.95	14.22	13.76	10.64	8.79	7.87	6.85	13.99
	0.99	16.23	12.04	10.43	9.13	7.67	16.23	15.75	12.06	10.23	8.68	7.70	16.09

Table 2. (Continued)

		Non trending case						Trending case					
		Number of breaks, k						Number of breaks, k					
q_f, q_b, p_b	α	1	2	3	4	5	$UD\max$	1	2	3	4	5	$UD\max$
Category (b) case 10, $\epsilon = 0.15$													
1, 1, 1	0.90	11.83	10.06	8.74	7.79	6.47	12.04	12.30	10.39	9.18	8.10	6.61	12.68
	0.95	13.95	11.26	9.76	8.47	7.15	14.02	14.55	11.71	10.14	8.97	7.32	14.66
	0.975	15.76	12.31	10.61	9.30	7.76	15.79	16.70	12.97	11.17	9.73	7.97	16.70
	0.99	17.98	13.55	11.36	9.85	8.56	17.98	18.68	14.61	12.38	10.45	8.61	18.68
1, 1, 2	0.90	12.87	10.93	9.59	8.68	7.52	13.22	13.45	11.50	10.17	8.88	7.75	13.83
	0.95	15.07	12.24	10.78	9.46	8.28	15.20	15.70	12.78	11.14	9.78	8.38	15.72
	0.975	16.68	13.17	11.62	10.23	8.94	17.10	18.41	14.04	11.86	10.55	8.97	18.41
	0.99	19.17	14.71	12.61	11.03	9.64	19.26	20.75	15.09	12.98	11.23	9.71	20.75
1, 2, 1	0.90	14.06	12.05	10.51	9.48	8.05	14.30	13.80	11.59	10.44	9.08	7.83	14.05
	0.95	15.99	13.20	11.61	10.23	8.77	15.99	15.79	12.99	11.44	9.83	8.56	15.95
	0.975	17.72	14.58	12.38	11.02	9.36	17.78	17.60	14.03	12.25	10.51	9.05	17.67
	0.99	19.77	16.16	13.80	12.00	10.09	19.77	20.69	15.52	13.13	11.66	9.77	20.69
1, 2, 2	0.90	15.06	12.97	11.51	10.40	9.05	15.47	14.61	12.22	11.07	10.17	8.95	15.10
	0.95	17.60	14.32	12.47	11.19	9.62	17.79	16.75	13.64	12.17	10.96	9.63	16.98
	0.975	19.42	15.75	13.55	12.09	10.37	19.57	18.67	15.03	13.34	12.00	10.37	18.88
	0.99	22.29	17.48	14.77	13.10	11.18	22.29	20.94	16.52	14.94	13.02	11.27	20.96
2, 1, 1	0.90	12.06	10.02	8.85	7.81	6.55	12.29	12.39	10.56	9.10	8.06	6.66	12.71
	0.95	13.80	11.36	9.70	8.57	7.21	13.92	14.37	11.87	10.17	8.75	7.31	14.76
	0.975	16.14	12.50	10.57	9.28	7.77	16.16	16.04	13.33	11.18	9.65	7.82	16.36
	0.99	18.68	14.40	11.75	10.21	8.50	18.76	19.23	14.56	12.18	10.48	8.67	19.23
2, 1, 2	0.90	13.13	10.91	9.72	8.72	7.50	13.49	13.56	11.44	10.16	9.06	7.85	13.78
	0.95	15.23	12.41	10.68	9.53	8.24	15.46	15.74	12.62	11.05	9.71	8.43	15.79
	0.975	17.23	13.51	11.56	10.13	8.92	17.36	17.56	13.76	11.97	10.47	8.85	17.62
	0.99	19.37	15.19	12.63	11.23	9.49	19.37	20.26	15.23	12.82	11.26	9.56	20.26
2, 2, 1	0.90	14.50	12.16	10.69	9.58	8.06	14.72	13.78	11.55	10.22	9.25	7.99	14.05
	0.95	16.78	13.46	11.88	10.35	8.74	16.80	15.64	12.81	11.18	9.98	8.62	15.81
	0.975	18.50	14.64	12.76	11.11	9.37	18.50	17.22	14.22	12.07	10.67	9.21	17.24
	0.99	20.83	16.28	13.77	11.82	9.92	20.83	19.20	15.48	13.49	11.61	10.04	19.20
2, 2, 2	0.90	15.29	13.03	11.64	10.49	9.09	15.70	14.82	12.52	11.15	10.17	8.92	15.23
	0.95	17.00	14.47	12.88	11.42	9.75	17.22	16.86	13.94	12.33	11.07	9.70	17.06
	0.975	18.87	15.49	13.72	12.12	10.43	19.08	18.99	15.48	13.30	11.79	10.27	19.24
	0.99	22.03	16.89	14.50	12.96	11.20	22.03	21.22	16.91	14.75	12.67	11.18	21.22

Table 3. Asymptotic critical values of the sequential test $SEQ_T(k+1|k)$

		Non trending case					Trending case				
		k					k				
q_b	α	1	2	3	4	5	1	2	3	4	5
Category (a) case 1, $\epsilon = 0.15$											
1	0.90	12.00	12.94	13.74	14.53	15.23	12.94	13.99	14.93	15.50	15.73
	0.95	13.78	15.25	16.38	17.02	17.70	15.01	15.85	16.53	16.86	17.04
	0.975	16.38	17.70	18.24	18.53	19.18	16.53	17.04	17.17	17.43	18.04
	0.99	18.53	19.33	19.92	20.50	21.34	17.43	18.58	19.11	19.22	19.54
2	0.90	14.26	15.02	15.64	16.02	16.51	13.57	14.78	15.40	15.87	16.12
	0.95	15.65	16.61	17.12	17.66	17.85	15.51	16.18	17.08	17.31	17.50
	0.975	17.12	17.85	18.22	19.04	19.27	17.08	17.50	19.27	19.62	19.70
	0.99	19.04	19.35	19.90	19.99	20.01	19.62	19.79	21.52	22.58	22.75
3	0.90	16.64	17.57	18.28	18.86	19.53	16.38	17.30	17.92	18.40	18.62
	0.95	18.30	19.58	20.21	20.77	21.45	17.99	18.74	19.77	20.28	20.89
	0.975	20.21	21.45	22.67	23.36	23.48	19.77	20.89	21.56	22.11	22.28
	0.99	23.36	23.52	24.13	24.43	25.16	22.11	22.37	22.83	23.98	24.54

Table 3. (Continued)

		Non trending case					Trending case				
		k					k				
q_b	α	1	2	3	4	5	1	2	3	4	5
4	0.90	18.96	19.91	20.68	21.13	21.51	18.29	19.54	20.43	20.97	21.32
	0.95	20.80	21.59	22.36	22.58	23.12	20.51	21.81	22.40	23.12	23.78
	0.975	22.36	23.12	24.10	25.73	26.11	22.40	23.78	25.10	25.75	25.84
	0.99	25.73	27.01	27.43	27.47	27.75	25.75	26.36	26.66	26.86	27.71
q_f											
		Category (a) case 2, $\epsilon = 0.15$									
1	0.90	9.14	10.09	10.61	11.04	11.45	10.22	11.21	12.02	12.33	12.75
	0.95	10.63	11.54	12.09	12.57	12.86	12.15	12.77	13.48	14.21	14.32
	0.975	12.09	12.86	13.25	14.01	14.19	13.48	14.32	14.66	15.41	15.72
	0.99	14.01	14.33	14.80	15.33	16.43	15.41	15.96	16.23	16.48	16.62
2	0.90	10.06	11.18	11.68	12.21	12.52	9.92	10.73	11.41	11.79	12.18
	0.95	11.69	12.62	13.33	13.66	14.07	11.41	12.18	12.80	13.21	13.69
	0.975	13.33	14.07	14.61	15.22	15.31	12.80	13.69	14.19	14.68	14.94
	0.99	15.22	15.40	16.51	17.02	18.13	14.68	15.00	15.96	16.37	17.09
3	0.90	9.97	10.74	11.25	11.73	12.17	9.95	11.05	11.64	11.92	12.76
	0.95	11.27	12.18	12.60	12.88	12.94	11.66	12.77	13.26	13.72	14.15
	0.975	12.60	12.94	13.24	14.33	14.49	13.26	14.15	14.70	14.83	15.71
	0.99	14.33	15.14	15.32	15.56	16.12	14.83	15.86	16.59	16.66	16.91
4	0.90	10.01	10.81	11.55	12.09	12.37	10.19	11.19	11.79	12.67	13.05
	0.95	11.59	12.40	12.80	13.88	14.23	11.90	13.08	13.68	14.53	15.03
	0.975	12.80	14.23	15.59	15.74	16.03	13.68	15.03	15.62	16.08	16.70
	0.99	15.74	16.10	16.61	16.93	17.05	16.08	16.80	17.48	17.48	17.80
q_f, q_b											
		Category (a) case 4, $\epsilon = 0.15$									
1, 1	0.90	11.98	13.02	14.03	14.73	14.94	12.20	13.51	14.26	14.63	15.21
	0.95	14.05	14.94	15.48	16.02	16.50	14.30	15.25	16.28	16.65	17.05
	0.975	15.48	16.50	17.10	17.57	17.92	16.28	17.05	17.85	18.17	18.46
	0.99	17.57	18.68	20.20	20.26	20.63	18.17	18.54	20.88	22.23	22.35
1, 2	0.90	14.77	15.85	16.63	17.17	17.35	14.09	15.20	15.77	16.04	16.38
	0.95	16.64	17.36	18.10	18.48	18.70	15.82	16.44	17.19	17.89	18.19
	0.975	18.10	18.70	19.48	20.38	20.61	17.19	18.19	18.76	19.21	19.61
	0.99	20.38	21.05	21.57	22.36	22.54	19.21	19.69	20.34	20.48	20.66
2, 1	0.90	12.87	13.78	14.72	15.06	15.47	13.11	14.03	15.14	15.73	16.22
	0.95	14.77	15.55	16.14	16.46	16.70	15.22	16.45	17.21	17.85	18.15
	0.975	16.14	16.70	16.99	17.19	18.20	17.21	18.15	18.79	18.96	19.01
	0.99	17.19	18.36	18.55	18.58	18.91	18.96	19.48	20.33	20.49	20.86
2, 2	0.90	14.70	15.67	16.70	17.04	17.56	14.48	15.40	15.93	16.37	16.70
	0.95	16.71	17.65	18.63	19.36	19.49	15.93	16.72	17.56	17.94	18.10
	0.975	18.63	19.49	20.02	20.55	21.07	17.56	18.10	18.78	19.01	19.64
	0.99	20.55	21.38	22.89	23.16	24.18	19.01	20.34	21.05	21.28	21.30

alternative hypothesis. On the other hand, Vogelsang (1999) showed through simulation experiments that estimation of the long-run variance under the null hypothesis leads to the problem of nonmonotonic power in finite samples. In related work, Crainiceanu and Vogelsang (2007) showed that commonly used data-dependent bandwidths for estimation of the long-run variance (based on the misspecified null model) are too large under the alternative hypothesis. This in turn leads to a decrease in power as the magnitude of the change increases. As a solution to this size–power trade-off, we use a new estimator of the long-run variance constructed using a hybrid method that involves residuals computed under both the null and alternative hypotheses. In particular, the data-dependent bandwidth is selected based on the residuals obtained under the alternative hypothesis. With this particular value of the bandwidth, the estimate is computed using residuals obtained under the null hypothesis of no structural change. Specifically, the proposed estimator is

$$\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\tilde{u}_t^2 + 2T^{-1}\sum_{j=1}^{T-1} w(j/\hat{h}) \sum_{t=j+1}^{T} \tilde{u}_t\tilde{u}_{t-j}, \qquad (8)$$

where $\tilde{u}_t$ are the residuals obtained imposing the null hypothesis. The kernel function $w(\cdot)$ is the quadratic spectral, and,

Table 4. Asymptotic critical values of the sequential test $SEQ_T(k+1|k)$

		Non trending case					Trending case				
		k					k				
q_b, p_b	α	1	2	3	4	5	1	2	3	4	5
		Category (b) cases 4 and 8, $\epsilon = 0.15$									
1, 1	0.90	13.18	13.92	14.70	15.08	15.79	13.72	15.14	15.72	16.44	16.75
	0.95	14.72	15.82	16.60	17.28	17.61	15.73	16.83	17.54	17.99	18.17
	0.975	16.60	17.61	19.20	19.43	19.85	17.54	18.17	19.27	19.97	20.53
	0.99	19.43	20.02	21.38	21.43	22.10	19.97	21.13	22.77	23.42	23.98
1, 2	0.90	15.06	16.32	17.39	17.83	18.22	15.09	16.21	16.85	17.33	17.85
	0.95	17.44	18.25	18.65	19.10	19.96	16.86	17.87	18.81	18.95	19.28
	0.975	18.65	19.96	20.06	20.37	20.69	18.81	19.28	19.66	21.10	21.43
	0.99	20.37	20.73	21.96	23.13	23.22	21.10	21.61	22.74	23.70	24.12
2, 1	0.90	15.82	16.69	17.59	18.15	18.39	15.21	16.54	17.44	17.98	18.46
	0.95	17.68	18.63	19.37	19.89	20.39	17.49	18.49	19.26	19.61	20.27
	0.975	19.37	20.39	21.48	22.63	22.84	19.26	20.27	20.76	21.69	22.03
	0.99	22.63	23.82	24.73	25.40	25.62	21.69	22.37	22.94	24.08	24.08
2, 2	0.90	16.95	18.69	19.46	20.06	20.44	17.12	18.56	19.40	19.92	20.75
	0.95	19.48	20.44	21.33	21.66	21.97	19.45	20.42	21.16	21.46	22.33
	0.975	21.33	21.97	22.39	23.52	24.03	21.16	21.86	22.89	23.41	23.85
	0.99	23.52	24.11	24.75	25.05	25.12	23.41	23.85	25.06	25.94	26.32
q_f, p_b											
		Category (b) cases 7 and 9, $\epsilon = 0.15$									
1, 1	0.90	10.56	11.68	12.06	12.70	13.25	10.14	11.07	11.81	12.31	12.90
	0.95	12.08	13.26	14.04	14.37	14.95	11.85	13.01	13.57	13.88	13.99
	0.975	14.04	14.95	15.11	15.68	16.31	13.57	13.99	14.63	15.19	15.95
	0.99	15.68	17.70	18.33	19.01	20.20	15.19	16.15	16.24	16.25	16.34
1, 2	0.90	11.52	12.51	12.96	13.57	14.28	10.95	11.83	12.70	12.92	13.89
	0.95	12.98	14.45	15.30	15.66	15.93	12.70	13.89	14.90	15.22	16.00
	0.975	15.30	15.93	16.30	16.85	16.95	14.90	16.00	16.68	17.33	17.48
	0.99	16.85	17.36	17.77	18.54	19.60	17.33	17.91	18.29	18.71	19.21
2, 1	0.90	10.65	11.45	11.95	12.68	13.14	10.49	11.45	12.34	12.86	13.34
	0.95	11.97	13.47	14.57	15.29	15.85	12.36	13.69	14.55	14.98	15.07
	0.975	14.57	15.85	16.64	17.43	17.92	14.55	15.07	15.29	15.72	15.86
	0.99	17.43	18.13	18.71	19.52	19.64	15.72	15.96	16.44	16.48	17.43
2, 2	0.90	12.22	13.22	14.03	14.56	14.93	11.76	12.88	13.46	14.31	14.75
	0.95	14.03	15.05	15.56	16.23	16.54	13.51	14.97	15.19	15.75	16.10
	0.975	15.56	16.54	17.38	17.82	18.18	15.19	16.10	16.45	17.06	17.27
	0.99	17.82	18.46	19.61	19.65	20.18	17.06	17.40	18.55	19.65	20.08
q_f, q_b, p_b											
		Category (b) case 10, $\epsilon = 0.15$									
1, 1, 1	0.90	13.72	15.13	16.24	16.68	17.11	14.32	15.97	16.59	17.08	17.31
	0.95	15.75	16.74	17.98	18.34	18.44	16.60	17.35	18.07	18.68	18.99
	0.975	17.42	18.34	19.12	19.85	20.15	18.07	18.99	19.73	20.26	20.71
	0.99	19.12	20.15	21.12	21.21	21.21	20.26	21.24	22.52	22.55	22.81
1, 1, 2	0.90	14.85	15.95	17.30	17.86	18.46	15.58	16.90	18.31	18.92	19.14
	0.95	16.64	18.01	19.17	19.55	19.72	18.39	19.14	19.98	20.75	21.50
	0.975	18.69	19.55	21.44	21.63	22.04	19.98	21.50	21.94	22.54	22.86
	0.99	21.44	22.04	23.51	24.20	24.20	22.54	23.07	23.18	23.35	23.85
1, 2, 1	0.90	15.94	16.98	17.99	18.30	18.46	15.72	17.04	17.59	17.75	18.16
	0.95	17.69	18.31	19.77	20.07	20.32	17.59	18.47	19.63	20.69	21.06
	0.975	19.01	20.07	20.93	21.42	21.81	19.63	21.06	21.76	22.59	22.70
	0.99	20.93	21.81	22.83	22.88	22.88	22.59	22.83	23.91	24.31	24.81

Table 4. (Continued)

q_f, q_b, p_b	α	Non trending case k					Trending case k				
		1	2	3	4	5	1	2	3	4	5
1, 2, 2	0.90	17.43	18.53	19.99	20.17	20.75	16.70	17.70	18.60	19.20	19.82
	0.95	19.42	20.21	22.29	22.49	22.57	18.63	19.83	20.60	20.94	21.27
	0.975	21.40	22.49	23.20	24.51	24.63	20.60	21.27	21.71	23.06	23.19
	0.99	23.20	24.63	25.82	26.26	26.26	23.06	23.23	23.52	23.54	25.67
2, 1, 1	0.90	13.75	14.86	16.55	17.03	18.10	14.31	15.26	15.96	16.71	17.40
	0.95	16.09	17.16	18.68	18.90	20.00	16.00	17.60	18.26	19.23	19.96
	0.975	18.12	18.90	20.85	21.25	22.27	18.26	19.96	21.00	22.20	22.30
	0.99	20.85	22.09	22.93	23.14	23.14	22.20	22.61	24.61	24.76	25.10
2, 1, 2	0.90	15.22	16.26	17.60	18.14	18.99	15.68	16.62	17.43	17.99	18.50
	0.95	17.14	18.23	19.37	20.03	20.79	17.52	18.50	19.57	20.26	20.44
	0.975	19.03	20.03	21.76	22.33	23.09	19.57	20.44	21.28	21.79	22.42
	0.99	21.76	22.83	23.18	23.46	23.46	21.79	22.50	22.82	23.61	23.91
2, 2, 1	0.90	16.70	17.48	18.82	19.34	20.49	15.54	16.36	17.08	17.44	17.96
	0.95	18.34	19.38	20.83	21.46	21.70	17.11	18.01	18.62	19.20	19.60
	0.975	20.52	21.46	22.00	23.35	23.69	18.62	19.60	20.74	21.16	22.04
	0.99	22.00	23.59	24.22	25.41	25.41	21.16	22.35	22.92	23.90	24.39
2, 2, 2	0.90	16.93	18.15	19.27	19.87	20.73	16.76	17.94	18.93	19.82	20.09
	0.95	18.87	19.92	22.03	22.30	22.85	18.97	20.11	20.59	21.22	21.83
	0.975	21.29	22.30	23.24	23.62	23.70	20.59	21.83	22.31	22.75	23.49
	0.99	23.24	23.64	24.24	24.36	24.36	22.75	23.58	25.33	25.76	26.04

following Andrews (1991), the estimate of the bandwidth is given by $\hat{h} = 1.3221(\hat{a}(2)T)^{1/5}$, where $\hat{a}(2) = [4\hat{\rho}^2/(1-\hat{\rho})^4]$ and $\hat{\rho} = \sum_{t=2}^{T} \hat{u}_t\hat{u}_{t-1} / \sum_{t=2}^{T} \hat{u}_{t-1}^2$, with $\hat{u}_t$ the residuals from the model estimated under the alternative hypotheses. As we show later, the sup-Wald test based on this estimator is able to bypass the problem of nonmonotonic power while maintaining an exact size close to the nominal size. (For more details on the merits of this approach, see Kejriwal 2009.)

5.2 Endogenous $I(1)$ Regressors

In general, the assumption of strict exogeneity is too restrictive, and the test statistics described in the previous section are not robust to the problem of endogenous regressors. In this section we use the linear leads and lags estimator (dynamic OLS) as proposed by Saikkonen (1991) and Stock and Watson (1993) and prove that the limiting distributions of the tests based on this estimator are the same as those obtained with the static regression under strict exogeneity. The modified regression is given by

$$y_t = \hat{c}_i + \mathbf{z}'_{ft}\hat{\boldsymbol{\delta}}_f + \mathbf{x}'_{ft}\hat{\boldsymbol{\beta}}_f + \mathbf{z}'_{bt}\hat{\boldsymbol{\delta}}_{bi} + \mathbf{x}'_{bt}\hat{\boldsymbol{\beta}}_{bi} + \sum_{j=-\ell_T}^{\ell_T} \Delta\mathbf{z}'_{t-j}\hat{\boldsymbol{\Pi}}_j + \hat{v}_t^*, \quad (9)$$

where $\mathbf{z}_t = (\mathbf{z}'_{ft}, \mathbf{z}'_{bt})'$. Note that the numbers of leads and lags of $\Delta\mathbf{z}_t$ need not be the same, but here we specify the same values for simplicity. Alternatively, ℓ_T could be interpreted as the maximum of the number of leads and lags. To prove our results, we need a few additional assumptions, which are the same as those required to show the consistency of the estimate of the cointegrating vector in the case of a model with no structural change.

Assumption A7. Let $\boldsymbol{\zeta}_t = (u_t, \mathbf{u}_{zt}^{f\prime}, \mathbf{u}_{zt}^{b\prime})'$ and $\boldsymbol{\zeta}_{zt} = (\mathbf{u}_{zt}^{f\prime}, \mathbf{u}_{zt}^{b\prime})'$. The spectral density matrix $\mathbf{f}_{\zeta\zeta}(w)$ is bounded away from 0 so that $\mathbf{f}_{\zeta\zeta}(w) \geq \alpha I_n$ $(n = q_f + q_b + 1)$ for $w \in [0, \pi]$ where $\alpha > 0$. In addition, the covariance function of $\boldsymbol{\zeta}_t$ is absolutely summable; that is, denoting $E(\boldsymbol{\zeta}_t\boldsymbol{\zeta}'_{t+k}) = \boldsymbol{\Gamma}(k)$, we require that $\sum_{k=-\infty}^{\infty} \|\boldsymbol{\Gamma}(k)\| < \infty$, where $\|\cdot\|$ is the standard Euclidean norm. Denoting the fourth-order cumulants of $\boldsymbol{\zeta}_t$ by $\kappa_{ijkl}(m_1, m_2, m_3)$, we assume that $\sum_{m_1}\sum_{m_2}\sum_{m_3} |\kappa_{ijkl}(m_1, m_2, m_3)| < \infty$, where the summations run from $-\infty$ to $+\infty$.

Assumption A7 states the same conditions used by Saikkonen (1991) and allows us to represent the error u_t as $u_t = \sum_{j=-\infty}^{\infty} \boldsymbol{\zeta}'_{zt-j}\boldsymbol{\Pi}_j + v_t$, with $\sum_{k=-\infty}^{\infty} \|\boldsymbol{\Pi}_j\| < \infty$ and where v_t is a stationary process such that $E(\boldsymbol{\zeta}_{zt}v_{t+k}) = 0$, for all k, and $f_{vv}(w) = f_{uu}(w) - \mathbf{f}_{u\zeta_z}(w)\mathbf{f}_{\zeta_z\zeta_z}(w)^{-1}\mathbf{f}_{\zeta_z u}(w)$. The DGP under the null hypothesis is then

$$y_t = c + \mathbf{z}'_{ft}\boldsymbol{\delta}_f + \mathbf{x}'_{ft}\boldsymbol{\beta}_f + \sum_{j=-\ell_T}^{\ell_T} \Delta\mathbf{z}'_{t-j}\boldsymbol{\Pi}_j + v_t^*,$$

where $v_t^* = v_t + \sum_{|j|>\ell_T} \boldsymbol{\zeta}'_{z,t-j}\boldsymbol{\Pi}_j \equiv v_t + e_t$. The last requirements pertain to the possible rate of increase of ℓ_T as T increases. Following Kejriwal and Perron (2008a), these are given by the following:

Assumption A8. As $T \to \infty$, $\ell_T \to \infty$, $\ell_T^2/T \to 0$, and $\ell_T \sum_{|j|>\ell_T} \|\boldsymbol{\Pi}_j\| \to 0$.

Note that Assumption A8 allows the use of information criteria, such as the Akaike information criterion and the Bayes information criterion. Because there can be serial correlation in the errors v_t, we need to apply a correction for its presence. Thus we consider the statistic $\sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^k} F_T^D(\boldsymbol{\lambda}, k) = F_T^D(\hat{\boldsymbol{\lambda}}, k)$, where $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_k)$ are the estimates of the break fractions obtained by minimizing the global sum of squared residuals (2) and $F_T^D(\boldsymbol{\lambda}, k) = T^{-1}(SSR_k/\hat{\sigma}_v^2)F_T(\boldsymbol{\lambda}, k)$ with $F_T(\boldsymbol{\lambda}, k)$ as defined in (3). We consider an estimate $\hat{\sigma}_v^2$ based on a weighted sum of the sample autocovariances of $\tilde{v}_t^*$, the residuals obtained imposing the null, as defined by (8) with $\tilde{v}_t^*$ instead of $\tilde{u}_t$ (and using the unrestricted residuals to obtain the bandwidth as discussed in the previous section). The relevant result is stated in the following proposition.

Theorem 3. Under Assumptions A1–A5, A7, and A8, for all testing problems, the limit distributions of the test $\sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^k} F_T^D(\boldsymbol{\lambda}, k)$, based on regression (9), are the same as those that apply to the test $\sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon^k} F_T(\boldsymbol{\lambda}, k)$ under the added assumption A6 and strict exogeneity with $\boldsymbol{\Omega}_{1z}^f = \boldsymbol{\Omega}_{1z}^b = \mathbf{0}$.

6. SIMULATION EXPERIMENTS

We now present the results of simulation experiments that pertain to the size and power of the tests, including a comparison with the frequently used LM tests. Hansen's (2000) method based on a "fixed regressors bootstrap" is also a possible avenue for providing valid large-sample inference in some of the models considered. In theory, an advantage of Hansen's method is that it remains valid in the presence of changes in the marginal distributions of the regressors. We conducted extensive simulations and found that the Wald tests considered here are very robust to changes in the drift of the $I(1)$ regressors and changes in the variance of the innovations driving them (as in the stationary case, as reported in Hansen 2000). Our asymptotic results provide tests with exact sizes close to nominal size.

6.1 Size of the Tests

We start with the case in which the DGP exhibits no structural change, and analyze the size of the tests. The sample sizes considered are $T = 120$ and $T = 240$. The value of the trimming ϵ is set to 0.20. The maximum number of breaks (M) considered is three. Depending on whether or not we correct for serial correlation and/or endogeneity, we have the following four specifications: (a) S_Corr = 0, C_Corr = 0: no correction for serial correlation or endogeneity; (b) S_Corr = 1, C_Corr = 0: correction for serial correlation but not for endogeneity; (c) S_Corr = 0, C_Corr = 1: correction for endogeneity but not for serial correlation; and (d) S_Corr = 1, C_Corr = 1: correction for both endogeneity and serial correlation. To correct for serial correlation, we use the method discussed in Section 5.1. To correct for endogeneity, we use the dynamic OLS estimator, discussed in Section 5.2, with $\ell_T = 2$. The various DGPs considered include the following basic components: $y_t = z_t + u_t$ with $z_t = z_{t-1} + \eta_t$, where $\eta_t \sim$ iid N(0, 1). The DGPs considered are, where $e_t \sim$ iid N(0, 1) and $\mathrm{Cov}(\eta_t, e_t) = 0$: DGP-1 (iid errors, exogenous regressor): $u_t = e_t$; DGP-2 [AR(1) errors, exogenous regressor]: $u_t = \rho u_{t-1} + e_t$; DGP-3 [MA(1) errors, exogenous regressor]: $u_t = e_t - \theta e_{t-1}$; DGP-4 [iid errors, endogenous regressor]: $u_t = 0.8\eta_t + e_t$; DGP-5 [MA(1) errors, endogenous regressor]: $u_t = 0.5v_t + \eta_t$, $v_t = e_t - 0.5e_{t-1}$.

For each DGP, we consider the case in which the regressors are $\{1, z_t\}$ and both the intercept and the cointegrating coefficient are allowed to change across regimes. In all experiments, 1,000 replications are used. All rejection frequencies are calculated at the nominal 5% level. Table 5 reports the empirical size, with $T = 120$ and 240 and $\rho = \theta = 0.5$. Consider first the base case represented by DGP-1, where the regressor is strictly exogenous and the errors are iid. With S_Corr = 0, C_Corr = 0, the size is adequate for all the tests regardless of the specification used. For DGP-2 with AR(1) errors, all tests show substantial distortions when we do not correct for serial correlation; however, when using our proposed long-run variance estimator, the size distortions are no longer present, and the tests become

Table 5. Empirical size

		$T = 120$					$T = 240$				
Specification	Test\DGP	1	2	3	4	5	1	2	3	4	5
S_Corr = 0, C_Corr = 0	sup-$F_T^*(1)$	0.04	0.55	0.00	0.15	0.20	0.04	0.63	0.00	0.14	0.19
	sup-$F_T^*(2)$	0.05	0.73	0.00	0.19	0.27	0.04	0.82	0.00	0.20	0.31
	sup-$F_T^*(3)$	0.04	0.75	0.00	0.20	0.27	0.04	0.85	0.00	0.22	0.33
	*UD*max	0.04	0.65	0.00	0.16	0.21	0.05	0.72	0.00	0.16	0.21
S_Corr = 1, C_Corr = 0	sup-$F_T^*(1)$	0.04	0.03	0.02	0.14	0.25	0.03	0.03	0.02	0.12	0.28
	sup-$F_T^*(2)$	0.03	0.02	0.05	0.13	0.29	0.03	0.02	0.02	0.17	0.43
	sup-$F_T^*(3)$	0.02	0.01	0.05	0.12	0.29	0.02	0.00	0.02	0.18	0.45
	*UD*max	0.04	0.03	0.03	0.14	0.26	0.03	0.02	0.02	0.14	0.32
S_Corr = 0, C_Corr = 1	sup-$F_T^*(1)$	0.06	0.58	0.00	0.05	0.00	0.05	0.64	0.00	0.05	0.00
	sup-$F_T^*(2)$	0.07	0.76	0.00	0.07	0.00	0.05	0.82	0.00	0.05	0.00
	sup-$F_T^*(3)$	0.06	0.77	0.00	0.06	0.00	0.05	0.86	0.00	0.05	0.00
	*UD*max	0.06	0.67	0.00	0.05	0.00	0.05	0.73	0.00	0.05	0.00
S_Corr = 1, C_Corr = 1	sup-$F_T^*(1)$	0.05	0.04	0.03	0.04	0.04	0.04	0.04	0.01	0.04	0.01
	sup-$F_T^*(2)$	0.03	0.02	0.06	0.03	0.07	0.03	0.02	0.02	0.04	0.03
	sup-$F_T^*(3)$	0.03	0.02	0.07	0.02	0.07	0.02	0.01	0.02	0.04	0.03
	*UD*max	0.05	0.04	0.04	0.04	0.05	0.04	0.04	0.01	0.05	0.02

somewhat conservative. With MA(1) errors (DGP-3), the tests have size 0 when no correction for serial correlation is made. Again, the size is accurate once we use S_Corr = 1. With endogeneity but no serial correlation (DGP-4), we see that all of the tests have good size for S_Corr = 0, C_Corr = 1. Otherwise, size distortions up to 20% may occur. This shows that the correction for endogeneity based on the dynamic OLS estimator is quite effective. When both serial correlation and endogeneity are present (DGP-5), the tests have adequate size when we use S_Corr = 1, C_Corr = 1, although some mild distortions persist when testing for multiple breaks. When $T = 240$, for the DGP-5 and S_Corr = 1, C_Corr = 1, the rejection frequencies are decreased, and even the multiple break tests become conservative.

We also considered the case in which the regressors are $\{1, z_t, x_t\}$, with $x_t \sim$ iid N(1, 1), $\mathrm{Cov}(x_t, u_t) = \mathrm{Cov}(x_t, \eta_t) = 0$, and the model allows the intercept and the cointegrating coefficient to change across regimes but with the coefficient of x_t held fixed. The results were similar to those given in Table 5, indicating that including an irrelevant $I(0)$ regressor does not lead to any size inaccuracies over and above those in the case in which they are not included.

6.2 Power Comparison With the LM Type Tests

In this section we analyze the power of the sup-Wald test and compare it with the sup, mean, and exp-LM tests proposed by Hansen (1992b) and Hao (1996). Vogelsang (1999) and Crainiceanu and Vogelsang (2007) showed that the power function of a wide variety of tests for a shift in the mean of a dynamic time series is nonmonotonic with respect to the magnitude of the break. One reason for this is the behavior of the estimate of the error variance in the presence of a shift in mean. In particular, if the error variance is estimated under the null hypothesis, then nonmonotonic power can result. We show that the LM-type tests have the same problem in the cointegration setup and that in certain cases, the power can go to zero as the magnitude of the break increases. Because the primary issue pertains to the presence of serial correlation in the errors, we consider the case in which the regressor is strictly exogenous and the trimming is set at $\epsilon = 0.15$. (We also simulated the power of our tests with a DGP involving endogenous regressors and found that the power actually was enhanced relative to that in the exogenous regressor case.) For the case with one break, the DGP is $y_t = z_t + u_t$, if $t \leq [T/2]$ and $y_t = (1+\delta)z_t + u_t$, if $t > [T/2]$, where $\eta_t \sim$ iid N(0, 1), $\mathrm{Cov}(u_t, \eta_t) = 0$. The sample size is $T = 240$. We consider DGP 2 [AR(1) errors] and 3 [MA(1) errors] and use the specification S_Corr = 1, C_Corr = 0. We analyze the pure structural change model in which both the intercept and the cointegrating coefficient are allowed to change. The power functions are plotted in Figure 1. Consider first the case with AR(1) errors. The nonmonotonicity of the power function of the LM tests is evident even at moderate values of δ. For very small values of δ, the power of the mean LM test is slightly higher than that of the modified Wald test, because the mean LM test is particularly well suited for detecting small changes (see Andrews and Ploberger 1994). Surprisingly, however, the mean LM test performs better than the exp-LM test even for large changes. The sup-LM test is dominated by all tests regardless of the sample size and degree of persistence. With MA(1) errors, the picture is quite different. All tests

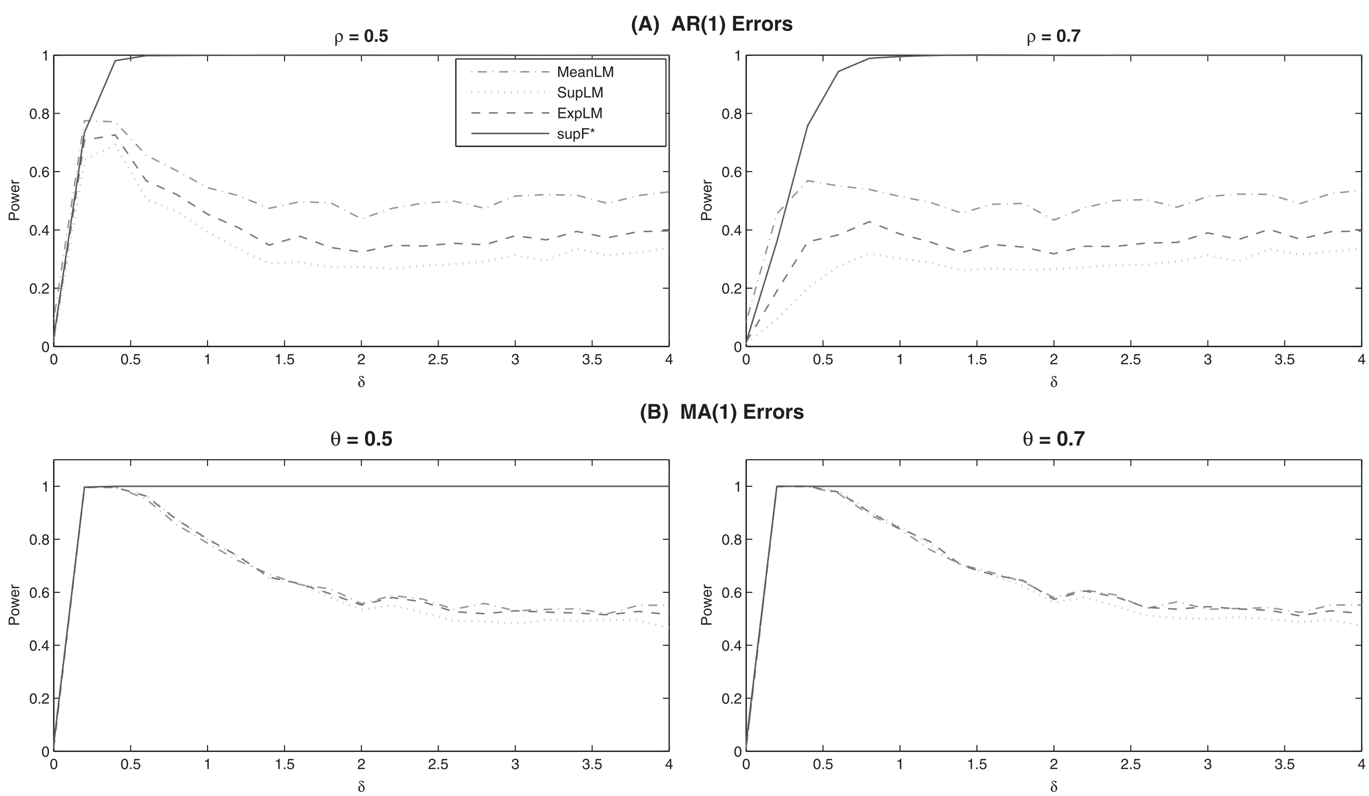

Figure 1. Power functions: the case with one break. The online version of this figure is in color.

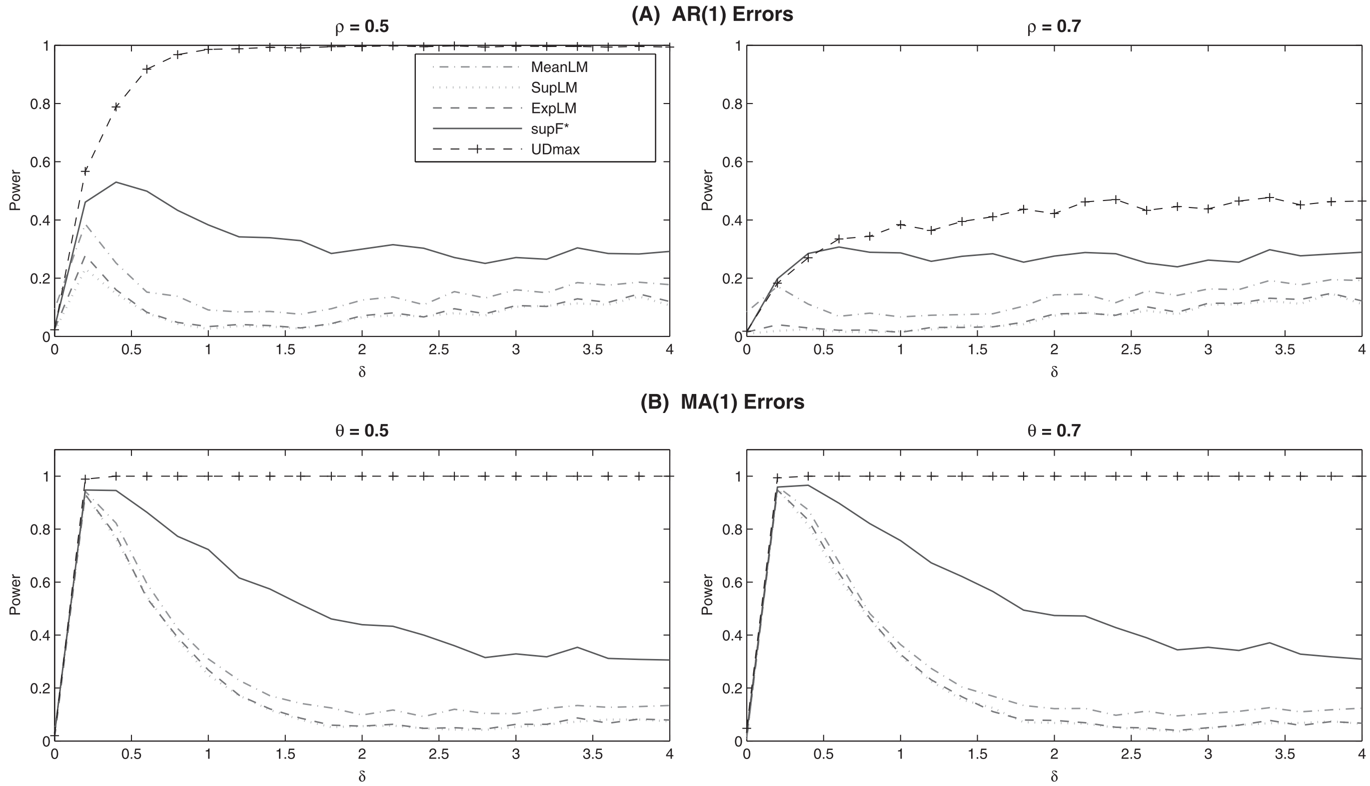

Figure 2. Power functions: the case with two breaks. The online version of this figure is in color.

have higher power than the autoregressive case, although nonmonotonicity is still evident for the LM tests. The performance of the LM tests is quite similar, and no clear ranking emerges between them.

We next consider the case in which the DGP involves two breaks and three regimes, specified by $y_t = z_t + u_t$, if $t \leq [T/3]$, $y_t = (1+\delta)z_t + u_t$ if $[T/3] < t \leq [2T/3]$ and $y_t = z_t + u_t$ if $[2T/3] < t \leq T$, where $z_t = z_{t-1} + \eta_t$, $z_t = z_{t-1} + \eta_t$, $\eta_t \sim$ iid N(0, 1) and $\text{Cov}(u_t, \eta_t) = 0$. The power functions are plotted in Figure 2. Consider first the case with AR(1) errors. Given that single break tests have difficulty detecting such parameter changes, it is not surprising that all tests exhibit nonmonotonic power. The modified sup-Wald test dominates all of the LM tests regardless of the sample size and the degree of persistence. With MA(1) errors, again all tests exhibit nonmonotonicity, although the power function of the modified Wald test is much higher than that of the LM tests. What is quite remarkable is the fact that in all cases, the $UD\max$ test has a much higher monotonic power function than any of the other tests. This provides clear evidence of its usefulness.

Finally, it is useful to comment on what happens when the regression is spurious, that is, there is no cointegration. Hansen (1992b) showed that the LM test designed to detect a martingale specification in the intercept, in the spirit of the test of Nyblom (1989), can be viewed as a test for the null of cointegration against the alternative of no cointegration. Although the sup-Wald test is not specifically targeted to the alternative of random variation in the intercept, it still has power against spurious regressions (i.e., no cointegration). This means that it also will reject when no structural change is present and there is no cointegration [the errors are $I(1)$]. However, we can use the following approach to determine whether the data suggest structural changes in a cointegrating relationship or a spurious regression. Suppose that we are willing to put an upper bound M (say 5) on the number of breaks. Then, if the system is cointegrated with fewer than M breaks, the sequential testing procedure can be used to consistently estimate the number of breaks. In contrast, if the regression is spurious, then the number of breaks selected will always (in large samples) be the maximum number of breaks allowed. Thus selecting the maximum allowable number of breaks can be indicative of the presence of $I(1)$ errors. The same is true when information criteria are used to select the number of breaks. We verified via simulations that this is indeed the case.

7. CONCLUSION

We have presented a comprehensive treatment of issues related to testing in cointegrated regression models with multiple structural changes. We analyzed models with $I(1)$ variables only, as well as models that 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. Our simulation experiments have shown that the commonly used LM tests are plagued with the problem of nonmonotonic power in finite samples, but the sup-Wald test can avoid such nonmonotonicity while maintaining adequate size. Our asymptotic results have allowed us to devise a sequential procedure for selecting the number of breaks. Finally, we have provided the asymptotic critical values of our tests for a wide range of models that we expect to be useful in practice. Our simulation experiments demonstrate the favorable properties of our test and our proposed long-run variance estimator. It is important to note that the idea of constructing the estimate of the long-run variance using information under both the null and alternative hypothesis is quite general and applies even in regression models that do not involve structural change.

APPENDIX

We use $\|\cdot\|$ to denote the Euclidean norm, i.e., $\|\mathbf{x}\| = (\sum_{i=1}^{p} x_i^2)^{1/2}$ for $\mathbf{x} \in R^p$. For a matrix $\mathbf{A}$, we use the vector-induced norm, i.e., $\|\mathbf{A}\| = \sup_{\mathbf{x}\neq\mathbf{0}} \|\mathbf{A}\mathbf{x}\|/\|\mathbf{x}\|$. We have $\|\mathbf{A}\| \leq [\mathrm{tr}(\mathbf{A}'\mathbf{A})]^{1/2}$. Also, for a projection matrix $\mathbf{P}$, $\|\mathbf{P}\mathbf{A}\| \leq \|\mathbf{A}\|$. We use the notation $\tilde{\mathbf{A}}_{i,j} = \mathbf{A}_{(i,j)} - \bar{\mathbf{A}}_{(i,j)}$, where $\mathbf{A}_{(i,j)}$ is the matrix of observations from regime i to regime j (both inclusive), i.e., $\mathbf{A}_{(i,j)} = (\mathbf{a}_{T_{i-1}+1}, \ldots, \mathbf{a}_{T_j})'$ while $\bar{\mathbf{A}}_{(i,j)}$ is the matrix (conformable to $\mathbf{A}_{i,j}$) of means, i.e., $\bar{\mathbf{A}}_{(i,j)} = (\bar{\mathbf{a}}_{i,j}, \ldots, \bar{\mathbf{a}}_{i,j})'$ where $\bar{\mathbf{a}}_{i,j} = (T_j - T_{i-1})^{-1} \sum_{t=T_{i-1}+1}^{T_j} \mathbf{a}_t$. Also, we use $\mathbf{A}^*_{(i,j)} = \mathbf{A}_{(i,j)} - \bar{\mathbf{A}}^{(i,j)}$, where $\bar{\mathbf{A}}^{(i,j)}$ is the matrix (conformable to $\mathbf{A}_{(i,j)}$) of sample averages, i.e., $\bar{\mathbf{A}}^{(i,j)} = (\bar{\mathbf{x}}, \ldots, \bar{\mathbf{x}})'$, where $\bar{\mathbf{x}} = T^{-1} \sum_{t=1}^{T} \mathbf{x}_t$. Let $\mathbf{1}_{(i,j)}$ be a $(T_j - T_{i-1}) \times 1$ vector of ones. To ease notation, we will write $\tilde{\mathbf{A}}_{(i,i)}$ as $\tilde{\mathbf{A}}_i$, $\mathbf{A}^*_{(i,i)}$ as $\mathbf{A}^*_i$, $\bar{\mathbf{A}}_{(i,i)}$ as $\bar{\mathbf{A}}_i$, $\bar{\mathbf{A}}^{(i,i)}$ as $\bar{\mathbf{A}}^i$ and $\mathbf{1}_{(i,j)}$ as $\mathbf{1}_i$, $(W_1, \mathbf{W}_z^f, \mathbf{W}_z^b, \mathbf{W}_x^f, \mathbf{W}_x^b)$ are independent Wiener processes with dimensions corresponding to those of $(B_1, \mathbf{B}_z^f, \mathbf{B}_z^b, \mathbf{B}_x^f, \mathbf{B}_x^b)$. We also use the notation $\mathbf{W}_z = (\mathbf{W}_z^{f\prime}, \mathbf{W}_z^{b\prime})'$. We start with a lemma about the weak convergence of various sample moments whose proof is standard given the results in Qu and Perron (2007).

Lemma A.1. Under Assumption A1–A5, the following weak convergence results hold (for $i = 1, \ldots, m+1$): (a) $T^{-3/2} \times \sum_{t=1}^{[T\lambda_i]} \mathbf{z}_{ft} \Rightarrow \int_0^{\lambda_i} \mathbf{B}_z^f$, $T^{-3/2} \sum_{t=1}^{[T\lambda_i]} \mathbf{z}_{bt} \Rightarrow \int_0^{\lambda_i} \mathbf{B}_z^b$, $T^{-1/2} \times \sum_{t=1}^{[T\lambda_i]} \mathbf{u}_{xt}^f \Rightarrow \mathbf{B}_x^f(\lambda_i)$, $T^{-1/2} \sum_{t=1}^{[T\lambda_i]} \mathbf{u}_{xt}^b \Rightarrow \mathbf{B}_x^b(\lambda_i)$, $T^{-1/2} \times \sum_{t=1}^{[T\lambda_i]} u_t \Rightarrow B_1(\lambda_i)$; (b) $T^{-2} \sum_{t=1}^{[T\lambda_i]} \mathbf{z}_{ft}\mathbf{z}_{ft}' \Rightarrow \int_0^{\lambda_i} \mathbf{B}_z^f \mathbf{B}_z^{f\prime}$, $T^{-2} \sum_{t=1}^{[T\lambda_i]} \mathbf{z}_{bt}\mathbf{z}_{bt}' \Rightarrow \int_0^{\lambda_i} \mathbf{B}_z^b \mathbf{B}_z^{b\prime}$; (c) $T^{-1} \sum_{t=1}^{[T\lambda_i]} \mathbf{z}_{ft} u_t \Rightarrow \int_0^{\lambda_i} \mathbf{B}_z^f \, dB_1 + \lambda_i(\mathbf{\Sigma}_{z1}^f + \mathbf{\Lambda}_{z1}^f)$, $T^{-1} \sum_{t=1}^{[T\lambda_i]} \mathbf{z}_{bt} u_t \Rightarrow \int_0^{\lambda_i} \mathbf{B}_z^b \, dB_1 + \lambda_i(\mathbf{\Sigma}_{z1}^b + \mathbf{\Lambda}_{z1}^b)$; (d) $T^{-1} \sum_{t=1}^{[T\lambda_i]} \mathbf{z}_{ft} \mathbf{u}_{xt}^{f\prime} \Rightarrow \int_0^{\lambda_i} \mathbf{B}_z^f \, d\mathbf{B}_x^{f\prime} + \lambda_i(\mathbf{\Sigma}_{zx}^{ff} + \mathbf{\Lambda}_{zx}^{ff})$, $T^{-1} \sum_{t=1}^{[T\lambda_i]} \mathbf{z}_{ft} \mathbf{u}_{xt}^{b\prime} \Rightarrow \int_0^{\lambda_i} \mathbf{B}_z^f \, d\mathbf{B}_x^{b\prime} + \lambda_i(\mathbf{\Sigma}_{zx}^{fb} + \mathbf{\Lambda}_{zx}^{fb})$, $T^{-1} \times \sum_{t=1}^{[T\lambda_i]} \mathbf{z}_{bt} \mathbf{u}_{xt}^{f\prime} \Rightarrow \int_0^{\lambda_i} \mathbf{B}_z^b \, d\mathbf{B}_x^{f\prime} + \lambda_i(\mathbf{\Sigma}_{zx}^{bf} + \mathbf{\Lambda}_{zx}^{bf})$, $T^{-1} \sum_{t=1}^{[T\lambda_i]} \mathbf{z}_{bt} \times \mathbf{u}_{xt}^{b\prime} \Rightarrow \int_0^{\lambda_i} \mathbf{B}_z^b \, d\mathbf{B}_x^{b\prime} + \lambda_i(\mathbf{\Sigma}_{zx}^{bb} + \mathbf{\Lambda}_{zx}^{bb})$.

The next lemma will also be useful in subsequent developments.

Lemma A.2. Let $\bar{\mathbf{X}}_{i((T_i - T_{i-1})\times p)} = (\bar{\mathbf{x}}_i, \ldots, \bar{\mathbf{x}}_i)'$, $\bar{\mathbf{x}}_i = (T_i - T_{i-1})^{-1} \sum_{t=T_{i-1}+1}^{T_i} \mathbf{x}_t$ and $\boldsymbol{\mu}^i_{((T_i - T_{i-1})\times p)} = (\boldsymbol{\mu}, \ldots, \boldsymbol{\mu})'$. Then under Assumptions A1–A4, we have for $i = 1, \ldots, m+1$: (i) $\boldsymbol{\mu}^i - \bar{\mathbf{X}}_i \xrightarrow{p} \mathbf{0}$; (ii) $T^{-1/2}(\mathbf{X}_i - \bar{\mathbf{X}}_i)'\mathbf{U}_i = T^{-1/2}(\mathbf{X}_i - \boldsymbol{\mu}^i)'\mathbf{U}_i + o_p(1)$; (iii) $T^{-1}(\mathbf{X}_i - \bar{\mathbf{X}}_i)'(\mathbf{X}_i - \bar{\mathbf{X}}_i) = T^{-1}(\mathbf{X}_i - \boldsymbol{\mu}^i)'(\mathbf{X}_i - \boldsymbol{\mu}^i) + o_p(1)$; (iv) $T^{-3/2}\mathbf{Z}_i'(\mathbf{X}_i - \bar{\mathbf{X}}_i) = T^{-3/2}\mathbf{Z}_i'(\mathbf{X}_i - \boldsymbol{\mu}^i) + o_p(1)$.

Proof. Part (i) follows trivially. To prove (ii), note that $T^{-1/2}(\mathbf{X}_i - \bar{\mathbf{X}}_i)'\mathbf{U}_i = T^{-1/2}(\mathbf{X}_i - \boldsymbol{\mu}^i)'\mathbf{U}_i + T^{-1/2}(\boldsymbol{\mu}^i - \bar{\mathbf{X}}_i)'\mathbf{U}_i$. We have $T^{-1/2}(\boldsymbol{\mu}^i - \bar{\mathbf{X}}_i)'\mathbf{U}_i = (\boldsymbol{\mu} - \bar{\mathbf{x}}_i)T^{-1/2} \times \sum_{t=T_{i-1}+1}^{T_i} u_t = o_p(1)$, using part (i). For (iii), note that

$$
\begin{aligned}
&T^{-1}(\mathbf{X}_i - \bar{\mathbf{X}}_i)'(\mathbf{X}_i - \bar{\mathbf{X}}_i) \\
&\quad = T^{-1}(\mathbf{X}_i - \boldsymbol{\mu}^i)'(\mathbf{X}_i - \boldsymbol{\mu}^i) + T^{-1}(\mathbf{X}_i - \boldsymbol{\mu}^i)'(\boldsymbol{\mu}^i - \bar{\mathbf{X}}_i) \\
&\qquad + T^{-1}(\boldsymbol{\mu}^i - \bar{\mathbf{X}}_i)'(\mathbf{X}_i - \boldsymbol{\mu}^i) + T^{-1}(\boldsymbol{\mu}^i - \bar{\mathbf{X}}_i)'(\boldsymbol{\mu}^i - \bar{\mathbf{X}}_i).
\end{aligned}
$$

Now $T^{-1}(\mathbf{X}_i - \boldsymbol{\mu}^i)'(\boldsymbol{\mu}^i - \bar{\mathbf{X}}_i) = T^{-1} \sum_{t=T_{i-1}+1}^{T_i} (\mathbf{x}_t - \boldsymbol{\mu})(\boldsymbol{\mu} - \bar{\mathbf{x}}_i)' = -(\lambda_i - \lambda_{i-1})(\boldsymbol{\mu} - \bar{\mathbf{x}}_i)(\boldsymbol{\mu} - \bar{\mathbf{x}}_i)' = o_p(1)$. Similarly, $T^{-1}(\boldsymbol{\mu}^i - \bar{\mathbf{X}}_i)'(\mathbf{X}_i - \boldsymbol{\mu}^i) = o_p(1)$. Finally, $T^{-1}(\boldsymbol{\mu}^i - \bar{\mathbf{X}}_i)'(\boldsymbol{\mu}^i - \bar{\mathbf{X}}_i) = (\lambda_i - \lambda_{i-1})(\boldsymbol{\mu} - \bar{\mathbf{x}}_i)(\boldsymbol{\mu} - \bar{\mathbf{x}}_i)' = o_p(1)$. To prove (iv), note that $T^{-3/2}\mathbf{Z}_i'(\boldsymbol{\mu}^i - \bar{\mathbf{X}}_i) = T^{-3/2}(\sum_{t=T_{i-1}+1}^{T_i} \mathbf{z}_t)(\boldsymbol{\mu} - \bar{\mathbf{x}}_i) = o_p(1)$ and the result follows immediately.

Proof of Theorem 1

We only consider cases 1 and 4. The details for the other cases can be found in the working paper version. We have

$$
F_T(\boldsymbol{\lambda}, k) = \frac{SSR_0 - SSR_k}{k(T - (k+1)(q_b + p_b) - q_f - p_f)^{-1} SSR_k},
$$

where SSR_0 and SSR_k are the sum of squared residuals under the null and alternative hypotheses, respectively. In all cases, we have $k(T - (k+1)(q_b + p_b) - q_f - p_f)^{-1} SSR_k \xrightarrow{p} k\sigma^2$.

Case 1: The regression under H_1 is $y_t = c_i + \mathbf{z}_{bt}'\boldsymbol{\delta}_{bi} + u_t$ and for SSR_0 we have

$$
\begin{aligned}
SSR_0 &= \big(\mathbf{Y}^*_{(1,k+1)} - \mathbf{Z}^*_{b(1,k+1)}\tilde{\boldsymbol{\delta}}_b\big)'\big(\mathbf{Y}^*_{(1,k+1)} - \mathbf{Z}^*_{b(1,k+1)}\tilde{\boldsymbol{\delta}}_b\big) \\
&= \big(\mathbf{Z}^*_{b(1,k+1)}(\boldsymbol{\delta}_b - \tilde{\boldsymbol{\delta}}_b) + \mathbf{U}^*_{(1,k+1)}\big)' \\
&\quad \times \big(\mathbf{Z}^*_{b(1,k+1)}(\boldsymbol{\delta}_b - \tilde{\boldsymbol{\delta}}_b) + \mathbf{U}^*_{(1,k+1)}\big) \\
&= \mathbf{U}^{*\prime}_{(1,k+1)}\mathbf{U}^*_{(1,k+1)} - \big(\mathbf{Z}^{*\prime}_{b(1,k+1)}\mathbf{U}^*_{(1,k+1)}\big)' \\
&\quad \times \big(\mathbf{Z}^{*\prime}_{b(1,k+1)}\mathbf{Z}^*_{b(1,k+1)}\big)^{-1}\big(\mathbf{Z}^{*\prime}_{b(1,k+1)}\mathbf{U}^*_{(1,k+1)}\big),
\end{aligned}
\tag{A.1}
$$

$$
\begin{aligned}
SSR_k &= \sum_{i=1}^{k+1} (\tilde{\mathbf{Y}}_i - \tilde{\mathbf{Z}}_{bi}\hat{\boldsymbol{\delta}}_{bi})'(\tilde{\mathbf{Y}}_i - \tilde{\mathbf{Z}}_{bi}\hat{\boldsymbol{\delta}}_{bi}) \\
&= \sum_{i=1}^{k+1} (\tilde{\mathbf{Z}}_{bi}(\boldsymbol{\delta}_b - \hat{\boldsymbol{\delta}}_{bi}) + \tilde{\mathbf{U}}_i)'(\tilde{\mathbf{Z}}_{bi}(\boldsymbol{\delta}_b - \hat{\boldsymbol{\delta}}_{bi}) + \tilde{\mathbf{U}}_i) \\
&= \sum_{i=1}^{k+1} \{-(\tilde{\mathbf{Z}}_{bi}'\tilde{\mathbf{U}}_i)'(\tilde{\mathbf{Z}}_{bi}'\tilde{\mathbf{Z}}_{bi})^{-1}(\tilde{\mathbf{Z}}_{bi}'\tilde{\mathbf{U}}_i) + \tilde{\mathbf{U}}_i'\tilde{\mathbf{U}}_i\}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&SSR_0 - SSR_k \\
&\quad \Rightarrow -\sigma^2\left(\int_0^1 \mathbf{W}_z^{b(1,k+1)}\, dW_1\right)' \\
&\qquad \times \left(\int_0^1 \mathbf{W}_z^{b(1,k+1)}\mathbf{W}_z^{b(1,k+1)\prime}\right)^{-1}\left(\int_0^1 \mathbf{W}_z^{b(1,k+1)}\, dW_1\right) \\
&\qquad + \sigma^2 \sum_{i=1}^{k+1}\left\{\left(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{b(i,i)}\, dW_1\right)'\right.
\end{aligned}
$$

$$\times \left(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{b(i,i)} \mathbf{W}_z^{b(i,i)\prime} \right)^{-1} \left(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{b(i,i)} \, dW_1 \right) \Bigg\}$$

$$+ \sigma^2 \sum_{i=1}^{k} \frac{(\lambda_i W_1(\lambda_{i+1}) - \lambda_{i+1} W_1(\lambda_i))^2}{\lambda_{i+1}\lambda_i(\lambda_{i+1} - \lambda_i)}$$

and the result stated follows.

Case 4: The regression under the alternative hypothesis is $y_t = c_i + \mathbf{z}_{ft}'\boldsymbol{\delta}_f + \mathbf{z}_{bt}'\boldsymbol{\delta}_{bi} + u_t$. Let $\mathbf{Z}^*_{(1,k+1)} = (\mathbf{Z}^*_{f(1,k+1)}, \mathbf{Z}^*_{b(1,k+1)})$ and $\tilde{\boldsymbol{\delta}} = (\tilde{\boldsymbol{\delta}}_f', \tilde{\boldsymbol{\delta}}_b')'$. We have

$$\begin{aligned} SSR_0 &= \left(\mathbf{Y}^*_{(1,k+1)} - \mathbf{Z}^*_{(1,k+1)}\tilde{\boldsymbol{\delta}}\right)'\left(\mathbf{Y}^*_{(1,k+1)} - \mathbf{Z}^*_{(1,k+1)}\tilde{\boldsymbol{\delta}}\right) \\ &= -\left(\mathbf{Z}^{*\prime}_{(1,k+1)}\mathbf{U}^*_{(1,k+1)}\right)'\left(\mathbf{Z}^{*\prime}_{(1,k+1)}\mathbf{Z}^*_{(1,k+1)}\right)^{-1} \\ &\quad \times \left(\mathbf{Z}^{*\prime}_{(1,k+1)}\mathbf{U}^*_{(1,k+1)}\right) + \mathbf{U}^{*\prime}_{(1,k+1)}\mathbf{U}^*_{(1,k+1)}, \\ SSR_k &= \sum_{i=1}^{k+1} (\tilde{\mathbf{Y}}_i - \tilde{\mathbf{Z}}_{fi}\hat{\boldsymbol{\delta}}_f - \tilde{\mathbf{Z}}_{bi}\hat{\boldsymbol{\delta}}_{bi})'(\tilde{\mathbf{Y}}_i - \tilde{\mathbf{Z}}_{fi}\hat{\boldsymbol{\delta}}_f - \tilde{\mathbf{Z}}_{bi}\hat{\boldsymbol{\delta}}_{bi}) \\ &= \sum_{i=1}^{k+1} (\tilde{\mathbf{Z}}_{fi}(\boldsymbol{\delta}_f - \hat{\boldsymbol{\delta}}_f) + \tilde{\mathbf{Z}}_{bi}(\boldsymbol{\delta}_b - \hat{\boldsymbol{\delta}}_{bi}) + \tilde{\mathbf{U}}_i)' \\ &\quad \times (\tilde{\mathbf{Z}}_{fi}(\boldsymbol{\delta}_f - \hat{\boldsymbol{\delta}}_f) + \tilde{\mathbf{Z}}_{bi}(\boldsymbol{\delta}_b - \hat{\boldsymbol{\delta}}_{bi}) + \tilde{\mathbf{U}}_i). \end{aligned}$$

After considerable algebra, we can show that

$$\begin{aligned} SSR_k = &-\left(\sum_{i=1}^{k+1} \tilde{\mathbf{Z}}_{1i}'\mathbf{M}_{bi}\tilde{\mathbf{U}}_i\right)'\left(\sum_{i=1}^{k+1} \tilde{\mathbf{Z}}_{1i}'\mathbf{M}_{bi}\tilde{\mathbf{Z}}_{1i}\right)^{-1} \\ &\times \left(\sum_{i=1}^{k+1} \tilde{\mathbf{Z}}_{1i}'\mathbf{M}_{bi}\tilde{\mathbf{U}}_i\right) \\ &- \sum_{i=1}^{k+1} (\tilde{\mathbf{Z}}_{bi}'\tilde{\mathbf{U}}_i)'(\tilde{\mathbf{Z}}_{bi}'\tilde{\mathbf{Z}}_{bi})^{-1}(\tilde{\mathbf{Z}}_{bi}'\tilde{\mathbf{U}}_i) + \sum_{i=1}^{k+1} (\tilde{\mathbf{U}}_i'\tilde{\mathbf{U}}_i), \end{aligned}$$

where $\mathbf{M}_{bi} = \mathbf{I}_i - \tilde{\mathbf{Z}}_{bi}(\tilde{\mathbf{Z}}_{bi}'\tilde{\mathbf{Z}}_{bi})^{-1}\tilde{\mathbf{Z}}_{bi}'$ and $\mathbf{I}_i$ the $(T_i - T_{i-1}) \times (T_i - T_{i-1})$ identity matrix. Thus,

$$\begin{aligned} &SSR_0 - SSR_k \\ &= -\left(\mathbf{Z}^{*\prime}_{(1,k+1)}\mathbf{U}^*_{(1,k+1)}\right)'\left(\mathbf{Z}^{*\prime}_{(1,k+1)}\mathbf{Z}^*_{(1,k+1)}\right)^{-1} \\ &\quad \times \left(\mathbf{Z}^{*\prime}_{(1,k+1)}\mathbf{U}^*_{(1,k+1)}\right) \\ &\quad + \left(\sum_{i=1}^{k+1} \tilde{\mathbf{Z}}_{fi}'\mathbf{M}_{bi}\tilde{\mathbf{U}}_i\right)'\left(\sum_{i=1}^{k+1} \tilde{\mathbf{Z}}_{fi}'\mathbf{M}_{bi}\tilde{\mathbf{Z}}_{fi}\right)^{-1}\left(\sum_{i=1}^{k+1} \tilde{\mathbf{Z}}_{fi}'\mathbf{M}_{bi}\tilde{\mathbf{U}}_i\right) \\ &\quad + \sum_{i=1}^{k+1} (\tilde{\mathbf{Z}}_{bi}'\tilde{\mathbf{U}}_i)'(\tilde{\mathbf{Z}}_{bi}'\tilde{\mathbf{Z}}_{bi})^{-1}(\tilde{\mathbf{Z}}_{bi}'\tilde{\mathbf{U}}_i) \\ &\quad + \mathbf{U}^{*\prime}_{(1,k+1)}\mathbf{U}^*_{(1,k+1)} - \sum_{i=1}^{k+1} (\tilde{\mathbf{U}}_i'\tilde{\mathbf{U}}_i) \end{aligned}$$

and, with $\mathbf{B}_z^{fb}(r) = (\mathbf{B}_z^f(r)', \mathbf{B}_z^b(r)')'$,

$$\begin{aligned} &SSR_0 - SSR_k \\ &\Rightarrow -\left(\int_0^1 \mathbf{B}_z^{fb(1,k+1)} \, dB_1\right)'\left(\int_0^1 \mathbf{B}_z^{fb(1,k+1)}\mathbf{B}_z^{fb(1,k+1)\prime}\right)^{-1} \\ &\quad \times \left(\int_0^1 \mathbf{B}_z^{fb(1,k+1)} \, dB_1\right) \\ &\quad + \left(\sum_{i=1}^{k+1} \int_{\lambda_{i-1}}^{\lambda_i} \mathbf{B}_z^{M(i,i)} \, dB_1\right)'\left(\sum_{i=1}^{k+1} \int_{\lambda_{i-1}}^{\lambda_i} \mathbf{B}_z^{M(i,i)}\mathbf{B}_z^{M(i,i)\prime}\right)^{-1} \\ &\quad \times \left(\sum_{i=1}^{k+1} \int_{\lambda_{i-1}}^{\lambda_i} \mathbf{B}_z^{M(i,i)} \, dB_1\right) \\ &\quad + \sum_{i=1}^{k+1}\left(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{B}_z^{b(i,i)} dB_1\right)'\left(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{B}_z^{b(i,i)}\mathbf{B}_z^{b(i,i)\prime}\right)^{-1} \\ &\quad \times \left(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{B}_z^{b(i,i)} \, dB_1\right) \\ &\quad + \sum_{i=1}^{k}\big((\lambda_i B_1(\lambda_{i+1}) - \lambda_{i+1} B_1(\lambda_i))' \\ &\quad \times (\lambda_i B_1(\lambda_{i+1}) - \lambda_{i+1} B_1(\lambda_i))\big)/(\lambda_{i+1}\lambda_i(\lambda_{i+1} - \lambda_i)), \end{aligned}$$

where $\mathbf{B}_z^{M(i,i)}(r) = \mathbf{B}_z^{f(i,i)}(r) - \int_{\lambda_{i-1}}^{\lambda_i} \mathbf{B}_z^{f(i,i)}\mathbf{B}_z^{b(i,i)\prime}(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{B}_z^{b(i,i)} \times \mathbf{B}_z^{b(i,i)\prime})^{-1}\mathbf{B}_z^{b(i,i)}(r)$. Note that each element of $\mathbf{B}_z^{M(i,i)}(r)$ is the residual from the projection of the corresponding element of $\mathbf{B}_z^{f(i,i)}(r)$ onto the space spanned by $\{B_{zj}^{b(i,i)}\}_{j=1}^{q_b}$ for a given realization of these stochastic processes. We also have $\mathbf{B}_z^{M(i,i)}(r) = (\boldsymbol{\Omega}_{zz}^{ff})^{1/2}\mathbf{W}_z^{M(i,i)}(r)$, so that

$$\begin{aligned} &kF_T(\boldsymbol{\lambda}, k) \\ &\Rightarrow -\left(\int_0^1 \mathbf{W}_z^{fb(1,k+1)} \, dW_1\right)' \\ &\quad \times \left(\int_0^1 \mathbf{W}_z^{fb(1,k+1)}\mathbf{W}_z^{fb(1,k+1)\prime}\right)^{-1} \\ &\quad \times \left(\int_0^1 \mathbf{W}_z^{fb(1,k+1)} \, dW_1\right) \\ &\quad + \left(\sum_{i=1}^{k+1} \int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{M(i,i)} \, dW_1\right)' \\ &\quad \times \left(\sum_{i=1}^{k+1} \int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{M(i,i)}\mathbf{W}_z^{M(i,i)\prime}\right)^{-1} \\ &\quad \times \left(\sum_{i=1}^{k+1} \int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{M(i,i)} \, dW_1\right) \\ &\quad + \sum_{i=1}^{k+1}\left(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{b(i,i)} \, dW_1\right)'\left(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{b(i,i)}\mathbf{W}_z^{b(i,i)\prime}\right)^{-1} \\ &\quad \times \left(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{b(i,i)} \, dW_1\right) \\ &\quad + \sum_{i=1}^{k}\big((\lambda_i W_1(\lambda_{i+1}) - \lambda_{i+1} W_1(\lambda_i))' \\ &\quad \times (\lambda_i W_1(\lambda_{i+1}) - \lambda_{i+1} W_1(\lambda_i))\big)/(\lambda_{i+1}\lambda_i(\lambda_{i+1} - \lambda_i)). \end{aligned}$$

Proof of Theorem 2

We give the details only for cases 4 to 6.

Case 4: The regression under H_1 is $y_t = c_i + \mathbf{z}'_{bt}\boldsymbol{\delta}_{bi} + \mathbf{x}'_{bt}\boldsymbol{\beta}_{bi} + u_t$. We have,

$$SSR_0 = \big[\mathbf{Y}^*_{(1,k+1)} - \mathbf{Z}^*_{b(1,k+1)}\tilde{\boldsymbol{\delta}}_b - \mathbf{X}^*_{b(1,k+1)}\tilde{\boldsymbol{\beta}}_b\big]' \times \big[\mathbf{Y}^*_{(1,k+1)} - \mathbf{Z}^*_{b(1,k+1)}\tilde{\boldsymbol{\delta}}_b - \mathbf{X}^*_{b(1,k+1)}\tilde{\boldsymbol{\beta}}_b\big]$$

By Lemmas A.1 and A.2, $T^{-3/2}\mathbf{Z}^{*\prime}_{b(1,k+1)}\mathbf{X}^*_{b(1,k+1)} = o_p(1)$. Thus,

$$\begin{aligned}SSR_0 &= \big[\mathbf{Z}^*_{b(1,k+1)}(\boldsymbol{\delta}_b - \tilde{\boldsymbol{\delta}}_b) + \mathbf{X}^*_{b(1,k+1)}(\boldsymbol{\beta}_b - \tilde{\boldsymbol{\beta}}_b) + \mathbf{U}^*_{(1,k+1)}\big]' \\ &\quad \times \big[\mathbf{Z}^*_{b(1,k+1)}(\boldsymbol{\delta}_b - \tilde{\boldsymbol{\delta}}_b) + \mathbf{X}^*_{b(1,k+1)}(\boldsymbol{\beta}_b - \tilde{\boldsymbol{\beta}}_b) + \mathbf{U}^*_{(1,k+1)}\big] \\ &= (\boldsymbol{\delta}_b - \tilde{\boldsymbol{\delta}}_b)'\mathbf{Z}^{*\prime}_{b(1,k+1)}\mathbf{Z}^*_{b(1,k+1)}(\boldsymbol{\delta}_b - \tilde{\boldsymbol{\delta}}_b) \\ &\quad + 2(\boldsymbol{\delta}_b - \tilde{\boldsymbol{\delta}}_b)'\mathbf{Z}^{*\prime}_{b(1,k+1)}\mathbf{U}^*_{(1,k+1)} + \mathbf{U}^{*\prime}_{(1,k+1)}\mathbf{U}^*_{(1,k+1)} \\ &\quad + (\boldsymbol{\beta}_b - \tilde{\boldsymbol{\beta}}_b)'\mathbf{X}^{*\prime}_{b(1,k+1)}\mathbf{X}^*_{b(1,k+1)}(\boldsymbol{\beta}_b - \tilde{\boldsymbol{\beta}}_b) \\ &\quad + 2(\boldsymbol{\beta}_b - \tilde{\boldsymbol{\beta}}_b)'\mathbf{X}^{*\prime}_{b(1,k+1)}\mathbf{U}^*_{(1,k+1)} + o_p(1) \\ &= -\big(T^{-1}\mathbf{U}^{*\prime}_{(1,k+1)}\mathbf{Z}^*_{b(1,k+1)}\big)\big(T^{-2}\mathbf{Z}^{*\prime}_{b(1,k+1)}\mathbf{Z}^*_{b(1,k+1)}\big)^{-1} \\ &\quad \times \big(T^{-1}\mathbf{Z}^{*\prime}_{b(1,k+1)}\mathbf{U}^*_{(1,k+1)}\big) \\ &\quad - \big(T^{-1/2}\mathbf{U}^{*\prime}_{(1,k+1)}\mathbf{X}^*_{b(1,k+1)}\big) \\ &\quad \times \big(T^{-1}\mathbf{X}^{*\prime}_{b(1,k+1)}\mathbf{X}^*_{b(1,k+1)}\big)^{-1} \\ &\quad \times \big(T^{-1/2}\mathbf{X}^{*\prime}_{b(1,k+1)}\mathbf{U}^*_{(1,k+1)}\big) \\ &\quad + \mathbf{U}^{*\prime}_{(1,k+1)}\mathbf{U}^*_{(1,k+1)} + o_p(1).\end{aligned}$$

We have $SSR_k = \sum_{i=1}^{k+1}[\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_{bi}\hat{\boldsymbol{\beta}}_{bi} - \tilde{\mathbf{Z}}_{bi}\hat{\boldsymbol{\delta}}_{bi}]'[\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_{bi}\hat{\boldsymbol{\beta}}_{bi} - \tilde{\mathbf{Z}}_{bi}\hat{\boldsymbol{\delta}}_{bi}]$. Using Lemmas A.1–A.2, $T^{-3/2}\tilde{\mathbf{Z}}'_{bi}\tilde{\mathbf{X}}_{bi} = o_p(1)$ and under H_0, $\tilde{\mathbf{Y}}_i = \tilde{\mathbf{X}}_{bi}\boldsymbol{\beta}_b + \tilde{\mathbf{Z}}_{bi}\boldsymbol{\delta}_b + \tilde{\mathbf{U}}_i$, so that

$$\begin{aligned}SSR_k &= \sum_{i=1}^{k+1}[\tilde{\mathbf{X}}_{bi}(\boldsymbol{\beta}_b - \hat{\boldsymbol{\beta}}_{bi}) + \tilde{\mathbf{Z}}_{bi}(\boldsymbol{\delta}_b - \hat{\boldsymbol{\delta}}_{bi}) + \tilde{\mathbf{U}}_i]' \\ &\quad \times [\tilde{\mathbf{X}}_{bi}(\boldsymbol{\beta}_b - \hat{\boldsymbol{\beta}}_{bi}) + \tilde{\mathbf{Z}}_{bi}(\boldsymbol{\delta}_b - \hat{\boldsymbol{\delta}}_{bi}) + \tilde{\mathbf{U}}_i] \\ &= \sum_{i=1}^{k+1}\big[-(T^{-1}\tilde{\mathbf{U}}'_i\tilde{\mathbf{Z}}_{bi})(T^{-2}\tilde{\mathbf{Z}}'_{bi}\tilde{\mathbf{Z}}_{bi})^{-1}(T^{-1}\tilde{\mathbf{Z}}'_{bi}\tilde{\mathbf{U}}_i) \\ &\quad - \big(T^{-1/2}\tilde{\mathbf{U}}'_i\tilde{\mathbf{X}}_{bi}\big)(T^{-1}\tilde{\mathbf{X}}'_{bi}\tilde{\mathbf{X}}_{bi})^{-1}\big(T^{-1/2}\tilde{\mathbf{X}}'_{bi}\tilde{\mathbf{U}}_i\big) + \tilde{\mathbf{U}}'_i\tilde{\mathbf{U}}_i\big] \\ &\quad + o_p(1).\end{aligned}$$

Therefore,

$$\begin{aligned}&kF_T(\boldsymbol{\lambda}, k) \\ &\Rightarrow -\left(\int_0^1 \mathbf{W}_z^{b(1,k+1)}\,dW_1\right)' \\ &\quad \times \left(\int_0^1 \mathbf{W}_z^{b(1,k+1)}\mathbf{W}_z^{b(1,k+1)\prime}\right)^{-1} \\ &\quad \times \left(\int_0^1 \mathbf{W}_z^{b(1,k+1)}\,dW_1\right) - \mathbf{W}^*_{xb}(1)'\mathbf{W}^*_{xb}(1) - W_1(1)^2 \\ &\quad + \sum_{i=1}^{k+1}\big\{(\lambda_i - \lambda_{i-1})^{-1}(W_1(\lambda_i) - W_1(\lambda_{i-1}))^2\big\} \\ &\quad + \sum_{i=1}^{k+1}(\lambda_i - \lambda_{i-1})^{-1}(\mathbf{W}^*_{xb}(\lambda_i) - \mathbf{W}^*_{xb}(\lambda_{i-1}))' \\ &\quad \times (\mathbf{W}^*_{xb}(\lambda_i) - \mathbf{W}^*_{xb}(\lambda_{i-1})) \\ &\quad + \sum_{i=1}^{k+1}\left[\left(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{b(i,i)}\,dW_1\right)'\left(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{b(i,i)}\mathbf{W}_z^{b(i,i)\prime}\right)^{-1} \right. \\ &\quad \left. \times \left(\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{b(i,i)}\,dW_1\right)\right]\end{aligned}$$

which reduces to the expression stated in the theorem.

Case 5: The model under H_1 is $y_t = c_i + \mathbf{z}'_{ft}\boldsymbol{\delta}_f + \mathbf{x}'_{ft}\boldsymbol{\beta}_f + u_t$. We have $SSR_k = \sum_{i=1}^{k+1}[\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_{fi}\hat{\boldsymbol{\beta}}_f - \tilde{\mathbf{Z}}_{fi}\hat{\boldsymbol{\delta}}_f]'[\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_{fi}\hat{\boldsymbol{\beta}}_f - \tilde{\mathbf{Z}}_{fi}\hat{\boldsymbol{\delta}}_f]$. Under H_0, $\tilde{\mathbf{Y}}_i = \tilde{\mathbf{X}}_{fi}\boldsymbol{\beta}_f + \tilde{\mathbf{Z}}_{fi}\boldsymbol{\delta}_f + \tilde{\mathbf{U}}_i$, so that

$$SSR_k = \sum_{i=1}^{k+1}[\tilde{\mathbf{X}}_{fi}(\boldsymbol{\beta}_f - \hat{\boldsymbol{\beta}}_f) + \tilde{\mathbf{Z}}_{fi}(\boldsymbol{\delta}_f - \hat{\boldsymbol{\delta}}_f) + \tilde{\mathbf{U}}_i]' \times [\tilde{\mathbf{X}}_{fi}(\boldsymbol{\beta}_f - \hat{\boldsymbol{\beta}}_f) + \tilde{\mathbf{Z}}_{fi}(\boldsymbol{\delta}_f - \hat{\boldsymbol{\delta}}_f) + \tilde{\mathbf{U}}_i].$$

Furthermore, $T(\hat{\boldsymbol{\delta}}_f - \boldsymbol{\delta}_f) = (T^{-2}\sum_{i=1}^{k+1}\tilde{\mathbf{Z}}'_{fi}\tilde{\mathbf{Z}}_{fi})^{-1}(T^{-1} \times \sum_{i=1}^{k+1}\tilde{\mathbf{Z}}'_{fi}\tilde{\mathbf{U}}_i) + o_p(1)$ and

$$T^{1/2}(\hat{\boldsymbol{\beta}}_f - \boldsymbol{\beta}_f) = \left(T^{-1}\sum_{i=1}^{k+1}\tilde{\mathbf{X}}'_{fi}\tilde{\mathbf{X}}_{fi}\right)^{-1}\left(T^{-1/2}\sum_{i=1}^{k+1}\tilde{\mathbf{X}}'_{fi}\tilde{\mathbf{U}}_i\right) + o_p(1).$$

Hence, after some algebra,

$$\begin{aligned}SSR_k &= -\left(T^{-1}\sum_{i=1}^{k+1}\tilde{\mathbf{U}}'_i\tilde{\mathbf{Z}}_{fi}\right)\left(T^{-2}\sum_{i=1}^{k+1}\tilde{\mathbf{Z}}'_{fi}\tilde{\mathbf{Z}}_{fi}\right)^{-1} \\ &\quad \times \left(T^{-1}\sum_{i=1}^{k+1}\tilde{\mathbf{Z}}'_{fi}\tilde{\mathbf{U}}_i\right) \\ &\quad - \left(T^{-1/2}\sum_{i=1}^{k+1}\tilde{\mathbf{U}}'_i\tilde{\mathbf{X}}_{fi}\right)\left(T^{-1}\sum_{i=1}^{k+1}\tilde{\mathbf{X}}'_{fi}\tilde{\mathbf{X}}_{fi}\right)^{-1} \\ &\quad \times \left(T^{-1/2}\sum_{i=1}^{k+1}\tilde{\mathbf{X}}'_{fi}\tilde{\mathbf{U}}_i\right) + \sum_{i=1}^{k+1}\tilde{\mathbf{U}}'_i\tilde{\mathbf{U}}_i + o_p(1)\end{aligned}$$

and

$$\begin{aligned}&kF_T(\boldsymbol{\lambda}, k) \\ &\Rightarrow -\left(\int_0^1 \mathbf{W}_z^{f(1,k+1)}\,dW_1\right)'\left(\int_0^1 \mathbf{W}_z^{f(1,k+1)}\mathbf{W}_z^{f(1,k+1)\prime}\right)^{-1} \\ &\quad \times \left(\int_0^1 \mathbf{W}_z^{f(1,k+1)}\,dW_1\right) \\ &\quad + \left(\sum_{i=1}^{k+1}\int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{f(i,i)}\,dW_1\right)'\end{aligned}$$

$$\times \left(\sum_{i=1}^{k+1} \int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{f(i,i)} \mathbf{W}_z^{f(i,i)\prime} \right)^{-1}$$

$$\times \left(\sum_{i=1}^{k+1} \int_{\lambda_{i-1}}^{\lambda_i} \mathbf{W}_z^{f(i,i)} \, dW_1 \right)$$

$$+ \sum_{i=1}^{k} \frac{(\lambda_i W_1(\lambda_{i+1}) - \lambda_{i+1} W_1(\lambda_i))^2}{\lambda_{i+1}\lambda_i(\lambda_{i+1} - \lambda_i)}.$$

Case 6: The model under H_1 is $y_t = c_i + \mathbf{z}_{bt}'\boldsymbol{\delta}_{bi} + \mathbf{x}_{ft}'\boldsymbol{\beta}_f + u_t$. In this case, $SSR_k = \sum_{i=1}^{k+1}[\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_{fi}\hat{\boldsymbol{\beta}}_f - \tilde{\mathbf{Z}}_{bi}\hat{\boldsymbol{\delta}}_{bi}]'[\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_{fi}\hat{\boldsymbol{\beta}}_f - \tilde{\mathbf{Z}}_{bi}\hat{\boldsymbol{\delta}}_{bi}]$. Under H_0, $\tilde{\mathbf{Y}}_i = \tilde{\mathbf{X}}_{fi}\boldsymbol{\beta}_f + \tilde{\mathbf{Z}}_{bi}\boldsymbol{\delta}_b + \tilde{\mathbf{U}}_i$, so that

$$SSR_k = \sum_{i=1}^{k+1} [\tilde{\mathbf{X}}_{fi}(\boldsymbol{\beta}_f - \hat{\boldsymbol{\beta}}_f) + \tilde{\mathbf{Z}}_{bi}(\boldsymbol{\delta}_b - \hat{\boldsymbol{\delta}}_{bi}) + \tilde{\mathbf{U}}_i]' \times [\tilde{\mathbf{X}}_{fi}(\boldsymbol{\beta}_f - \hat{\boldsymbol{\beta}}_f) + \tilde{\mathbf{Z}}_{bi}(\boldsymbol{\delta}_b - \hat{\boldsymbol{\delta}}_{bi}) + \tilde{\mathbf{U}}_i].$$

We also have $T(\hat{\boldsymbol{\delta}}_{bi} - \boldsymbol{\delta}_b) = (T^{-2}\tilde{\mathbf{Z}}_{bi}'\tilde{\mathbf{Z}}_{bi})^{-1} T^{-1}\tilde{\mathbf{Z}}_{bi}'\tilde{\mathbf{U}}_i + o_p(1)$ and

$$T^{1/2}(\hat{\boldsymbol{\beta}}_f - \boldsymbol{\beta}_f) = \left(T^{-1} \sum_{i=1}^{k+1} \tilde{\mathbf{X}}_{fi}'\tilde{\mathbf{X}}_{fi} \right)^{-1} \times \left(T^{-1/2} \sum_{i=1}^{k+1} \tilde{\mathbf{X}}_{fi}'\tilde{\mathbf{U}}_i \right) + o_p(1).$$

Hence,

$$SSR_k = -\sum_{i=1}^{k+1} (T^{-1}\tilde{\mathbf{U}}_i'\tilde{\mathbf{Z}}_{bi})(T^{-2}\tilde{\mathbf{Z}}_{bi}'\tilde{\mathbf{Z}}_{bi})^{-1}(T^{-1}\tilde{\mathbf{Z}}_{bi}'\tilde{\mathbf{U}}_i)$$
$$- \left(T^{-1/2} \sum_{i=1}^{k+1} \tilde{\mathbf{U}}_i'\tilde{\mathbf{X}}_{fi}' \right) \left(\sum_{i=1}^{k+1} T^{-1}\tilde{\mathbf{X}}_{fi}'\tilde{\mathbf{X}}_{fi} \right)^{-1}$$
$$\times \left(T^{-1/2} \sum_{i=1}^{k+1} \tilde{\mathbf{X}}_{fi}'\tilde{\mathbf{U}}_i \right) + \sum_{i=1}^{k+1} \tilde{\mathbf{U}}_i'\tilde{\mathbf{U}}_i$$

so that

$$kF_T(\boldsymbol{\lambda}, k)$$
$$\Rightarrow \sum_{i=1}^{k+1} \Bigg[-\left(\int_0^{\lambda_{i+1}} \mathbf{W}_z^{b(1,i+1)} \, dW_1 \right)'$$
$$\times \left(\int_0^{\lambda_{i+1}} \mathbf{W}_z^{b(1,i+1)} \mathbf{W}_z^{b(1,i+1)\prime} \right)^{-1}$$
$$\times \left(\int_0^{\lambda_{i+1}} \mathbf{W}_z^{b(1,i+1)} \, dW_1 \right)$$
$$+ \left(\int_0^{\lambda_i} \mathbf{W}_z^{b(1,i)} \, dW_1 \right)' \left(\int_0^{\lambda_i} \mathbf{W}_z^{b(1,i)} \mathbf{W}_z^{b(1,i)\prime} \right)^{-1}$$
$$\times \left(\int_0^{\lambda_i} \mathbf{W}_z^{b(1,i)} \, dW_1 \right) + \left(\int_{\lambda_i}^{\lambda_{i+1}} \mathbf{W}_z^{b(i+1,i+1)} \, dW_1 \right)'$$
$$\times \left(\int_{\lambda_i}^{\lambda_{i+1}} \mathbf{W}_z^{b(i+1,i+1)} \mathbf{W}_z^{b(i+1,i+1)\prime} \right)^{-1}$$
$$\times \left(\int_{\lambda_i}^{\lambda_{i+1}} \mathbf{W}_z^{b(i+1,i+1)} \, dW_1 \right) \Bigg]$$
$$+ \sum_{i=1}^{k} \frac{(\lambda_i W_1(\lambda_{i+1}) - \lambda_{i+1} W_1(\lambda_i))^2}{\lambda_{i+1}\lambda_i(\lambda_{i+1} - \lambda_i)}.$$

Proof of Theorem 3

We provide a proof for the testing problem (2) in category (a), a pure structural change model with only $I(1)$ regressors and a constant. The proofs for the other cases are very similar. We first let $\tilde{\mathbf{B}}_T = T^{-1/2}\sum_{j=1}^{[Tr]} \tilde{\boldsymbol{\zeta}}_j$, where $\tilde{\boldsymbol{\zeta}}_t = (v_t, \mathbf{u}_{zt}^{b\prime})'$. Under the stated conditions, $\tilde{\mathbf{B}}_T \Rightarrow \tilde{\mathbf{B}} \equiv (B_{1.z}, \mathbf{B}_z^b)$ as $T \to \infty$, where $B_{1.z}^b = B_1 - \boldsymbol{\Omega}_{1z}^b(\boldsymbol{\Omega}_{zz}^{bb})^{-1}\mathbf{B}_z^b$. Note that $B_{1.z}^b$ is independent of $\mathbf{B}_z^b$. Thus, $\tilde{\mathbf{B}}$ denotes a vector Brownian motion with block diagonal covariance matrix $\tilde{\boldsymbol{\Omega}} = \text{diag}((\sigma_{1.z}^b)^2, \boldsymbol{\Omega}_{zz}^{bb})$, where $(\sigma_{1.z}^b)^2 = \sigma^2 - \boldsymbol{\Omega}_{1z}^b(\boldsymbol{\Omega}_{zz}^{bb})^{-1}\boldsymbol{\Omega}_{z1}^b$. The relevant regression under the alternative hypothesis is

$$y_t = c_i + \mathbf{z}_{bt}'\hat{\boldsymbol{\delta}}_{bi} + \sum_{j=-\ell_T}^{\ell_T} \Delta\mathbf{z}_{b,t-j}'\hat{\boldsymbol{\Pi}}_j + \hat{v}_t^*.$$

As a matter of notation, let $\boldsymbol{\eta}_{bt}^* = (\Delta\mathbf{z}_{bt-\ell_T}', \ldots, \Delta\mathbf{z}_{bt+\ell_T}')'$, $\boldsymbol{\eta}_b^* = (\boldsymbol{\eta}_{b1}^*, \ldots, \boldsymbol{\eta}_{bT}^*)'$, $\mathbf{E} = (e_1, \ldots, e_T)'$, $\mathbf{V} = (v_1, \ldots, v_T)'$, and $\boldsymbol{\Pi} = (\boldsymbol{\Pi}_{-\ell_T}', \ldots, \boldsymbol{\Pi}_{\ell_T}')'$. Also, define $\mathbf{M}_\eta = \mathbf{I}_T - \boldsymbol{\eta}_b^*(\boldsymbol{\eta}_b'^*\boldsymbol{\eta}_b^*)^{-1}\boldsymbol{\eta}_b'^*$, $\mathbf{z}_t = (1, \mathbf{z}_{bt})'$, $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_T)'$, $\mathbf{Z}_i = (\mathbf{z}_{T_{i-1}+1}, \ldots, \mathbf{z}_{T_i})$, $\bar{\mathbf{Z}} = \text{diag}(\mathbf{Z}_1, \ldots, \mathbf{Z}_{k+1})$, $\boldsymbol{\delta} = (c, \boldsymbol{\delta}_b')'$, and the $[(k+1)(q_b+1) \times 1]$ vector $\bar{\boldsymbol{\delta}} = (\boldsymbol{\delta}, \boldsymbol{\delta}, \ldots, \boldsymbol{\delta})$. The vectors of estimates under the null and the alternative are $\bar{\tilde{\boldsymbol{\delta}}}$ and $\hat{\boldsymbol{\delta}}$, respectively. The vector of residuals is $\tilde{\mathbf{v}}^* = \mathbf{M}_\eta\mathbf{Y} - \mathbf{M}_\eta\bar{\mathbf{Z}}\bar{\tilde{\boldsymbol{\delta}}}$ under the null and $\hat{\mathbf{v}}^* = \mathbf{M}_\eta\mathbf{Y} - \mathbf{M}_\eta\bar{\mathbf{Z}}\hat{\boldsymbol{\delta}}$ under the alternative. We have $\tilde{\mathbf{v}}^* = \hat{\mathbf{v}}^* + \mathbf{M}_\eta\bar{\mathbf{Z}}(\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}})$, so that

$$SSR_0 - SSR_k = \tilde{\mathbf{v}}^{*\prime}\tilde{\mathbf{v}}^* - \hat{\mathbf{v}}^{*\prime}\hat{\mathbf{v}}^* = (\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}})'\bar{\mathbf{Z}}'\mathbf{M}_\eta\bar{\mathbf{Z}}(\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}})$$
$$= (\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}})'\bar{\mathbf{Z}}'\bar{\mathbf{Z}}(\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}})$$
$$- (\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}})'\bar{\mathbf{Z}}'\boldsymbol{\eta}_b^*(\boldsymbol{\eta}_b'^*\boldsymbol{\eta}_b^*)^{-1}\boldsymbol{\eta}_b'^*\bar{\mathbf{Z}}(\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}}).$$

Now note that

$$\|(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}})'\bar{\mathbf{Z}}'\boldsymbol{\eta}_b^*(\boldsymbol{\eta}_b'^*\boldsymbol{\eta}_b^*)^{-1}\boldsymbol{\eta}_b'^*\bar{\mathbf{Z}}(\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}})\|$$
$$\leq \|(\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}})'\mathbf{D}_T\| \|\mathbf{D}_T^{-1}\bar{\mathbf{Z}}'\boldsymbol{\eta}_b^*\|$$
$$\times \|(\boldsymbol{\eta}_b'^*\boldsymbol{\eta}_b^*)^{-1}\| \|\boldsymbol{\eta}_b'^*\bar{\mathbf{Z}}\mathbf{D}_T^{-1}\| \|\mathbf{D}_T(\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}})\|,$$

where the $[(k+1) \times (q_b+1)]$ diagonal matrix $\mathbf{D}_T = \text{diag}(T^{1/2}, T, T, \ldots, T, \ldots, T^{1/2}, T, \ldots, T)$. We have $\|\mathbf{D}_T(\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}})\| = O_p(1)$, $\|(\boldsymbol{\eta}_b'^*\boldsymbol{\eta}_b^*)^{-1}\| = O_p(T^{-1})$, $\|\mathbf{D}_T^{-1}\bar{\mathbf{Z}}'\boldsymbol{\eta}_b^*\| = O_p(l_T^{1/2})$, since $\|T^{-1}\sum_{t=1}^T \mathbf{Z}_{bt}\boldsymbol{\eta}_{bt}^{*\prime}\| = O_p(l_T^{1/2})$, $\|T^{-1/2}\sum_{t=1}^T \boldsymbol{\eta}_{bt}^{*\prime}\| = O_p(l_T^{1/2})$ (Saikkonen 1991; Kejriwal and Perron 2008a). Hence, $\|(\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}})'\bar{\mathbf{Z}}'\boldsymbol{\eta}_b^*(\boldsymbol{\eta}_b'^*\boldsymbol{\eta}_b^*)^{-1}\boldsymbol{\eta}_b'^*\bar{\mathbf{Z}}(\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}})\| = O_p(l_T/T) = o_p(1)$. Next,

$$(\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}})'\bar{\mathbf{Z}}'\bar{\mathbf{Z}}(\hat{\boldsymbol{\delta}} - \bar{\tilde{\boldsymbol{\delta}}})$$
$$= -(\mathbf{Z}'\mathbf{V})'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}$$
$$+ \sum_{i=1}^{k+1} (\mathbf{Z}_i'\mathbf{V}_i)'(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}(\mathbf{Z}_i'\mathbf{V}_i) + o_p(1)$$

$$= -\left(\mathbf{Z}^{*\prime}_{b(1,k+1)}\mathbf{V}^{*}_{(1,k+1)}\right)'\left(\mathbf{Z}^{*\prime}_{b(1,k+1)}\mathbf{Z}^{*}_{b(1,k+1)}\right)^{-1} \times \left(\mathbf{Z}^{*\prime}_{b(1,k+1)}\mathbf{V}^{*}_{(1,k+1)}\right) + \mathbf{V}^{*\prime}_{(1,k+1)}\mathbf{V}^{*}_{(1,k+1)} + \sum_{i=1}^{k+1}\{(\tilde{\mathbf{Z}}'_{bi}\tilde{\mathbf{V}}_i)'(\tilde{\mathbf{Z}}'_{bi}\tilde{\mathbf{Z}}_{bi})^{-1}(\tilde{\mathbf{Z}}'_{bi}\tilde{\mathbf{V}}_i) - \tilde{\mathbf{V}}'_i\tilde{\mathbf{V}}_i\} + o_p(1).$$

Therefore,

$$SSR_0 - SSR_k \Rightarrow -\left(\int_0^1 \mathbf{B}_z^{b(1,k+1)}\,dB_{1.z}^b\right)'\left(\int_0^1 \mathbf{B}_z^{b(1,k+1)}\mathbf{B}_z^{b(1,k+1)\prime}\right)^{-1} \times \left(\int_0^1 \mathbf{B}_z^{b(1,k+1)}\,dB_{1.z}^b\right) + \sum_{i=1}^{k+1}\left\{\left(\int_{\lambda_{i-1}}^{\lambda_i}\mathbf{B}_z^{b(i,i)}\,dB_{1.z}^b\right)'\left(\int_{\lambda_{i-1}}^{\lambda_i}\mathbf{B}_z^{b(i,i)}\mathbf{B}_z^{b(i,i)\prime}\right)^{-1} \times \left(\int_{\lambda_{i-1}}^{\lambda_i}\mathbf{B}_z^{b(i,i)}\,dB_{1.z}^b\right)\right\} + \sum_{i=1}^{k}\frac{(\lambda_i B_{1.z}^b(\lambda_{i+1}) - \lambda_{i+1}B_{1.z}^b(\lambda_i))^2}{\lambda_{i+1}\lambda_i(\lambda_{i+1}-\lambda_i)}.$$

Since $B_{1.z}^b$ and $\mathbf{B}_z$ are independent, $B_{1.z}^b = \sigma_{1.z}^b W_1$ and $\mathbf{B}_z^b = (\mathbf{\Omega}_{zz}^b)^{1/2}\mathbf{W}_z^b$, so that

$$SSR_0 - SSR_k \Rightarrow -(\sigma_{1.z}^b)^2\left(\int_0^1 \mathbf{W}_z^{b(1,k+1)}\,dW_1\right)' \times \left(\int_0^1 \mathbf{W}_z^{b(1,k+1)}\mathbf{W}_z^{b(1,k+1)\prime}\right)^{-1}\left(\int_0^1 \mathbf{W}_z^{b(1,k+1)}\,dW_1\right) + (\sigma_{1.z}^b)^2\sum_{i=1}^{k+1}\left\{\left(\int_{\lambda_{i-1}}^{\lambda_i}\mathbf{W}_z^{b(i,i)}\,dW_1\right)' \times \left(\int_{\lambda_{i-1}}^{\lambda_i}\mathbf{W}_z^{b(i,i)}\mathbf{W}_z^{b(i,i)\prime}\right)^{-1}\left(\int_{\lambda_{i-1}}^{\lambda_i}\mathbf{W}_z^{b(i,i)}\,dW_1\right)\right\} + (\sigma_{1.z}^b)^2\sum_{i=1}^{k}\frac{(\lambda_i W_1(\lambda_{i+1}) - \lambda_{i+1}W_1(\lambda_i))^2}{\lambda_{i+1}\lambda_i(\lambda_{i+1}-\lambda_i)}.$$

It can be shown, using arguments as in Kejriwal and Perron (2008a) that $\hat{\sigma}_v$ is a consistent estimate of $\sigma_{1.z}^b$ under the stated conditions (the proof is quite tedious and omitted). This proves the theorem.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation grant SES-0649350 (to P. P.). The authors thank an associate editor and two referees for useful suggestions. One referee was especially helpful in pointing out an error in a previous version.

[Received September 2007. Revised February 2009.]

REFERENCES

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858. [514]

——— (1993), "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61, 821–856. [503]

Andrews, D. W. K., and Ploberger, W. (1994), "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414. [503,516]

Bai, J. (1997), "Estimation of a Change Point in Multiple Regression Models," *Review of Economics and Statistics*, 79, 551–563. [503]

Bai, J., and Perron, P. (1998), "Estimating and Testing Linear Models With Multiple Structural Changes," *Econometrica*, 66, 47–78. [503,504,506,507]

——— (2003), "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics*, 18, 1–22. [505,508]

——— (2006), "Multiple Structural Change Models: A Simulation Analysis," in *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, eds. D. Corbea, S. Durlauf, and B. E. Hansen, New York: Cambridge University Press, pp. 212–237. [506]

Bai, J., Lumsdaine, R. L., and Stock, J. H. (1998), "Testing for and Dating Breaks in Multivariate Time Series," *Review of Economic Studies*, 65, 395–432. [503,506,508]

Carrion-i-Silvestre, J. L., and Sansó-i-Rosselló, A. S. (2006), "Testing the Null Hypothesis of Cointegration With Structural Breaks," *Oxford Bulletin of Economics and Statistics*, 68, 623–646. [503]

Crainiceanu, C. M., and Vogelsang, T. J. (2007), "Nonmonotonic Power for Tests of a Mean Shift in a Time Series," *Journal of Statistical Computation and Simulation*, 77, 457–476. [512,516]

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford: Oxford University Press. [505]

Hansen, B. E. (1992a), "Efficient Estimation and Testing of Cointegrating Vectors in the Presence of Deterministic Trends," *Journal Econometrics*, 53, 87–121. [508]

——— (1992b), "Tests for Parameter Instability in Regressions With $I(1)$ Processes," *Journal of Business & Economic Statistics*, 10, 321–335. [503,516,517]

——— (1992c), "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes," *Econometrica*, 60, 967–972. [505]

——— (2000), "Testing for Structural Change in Conditional Models," *Journal of Econometrics*, 97, 93–115. [503,515]

Hansen, P. R. (2003), "Structural Changes in the Cointegrated Vector Autoregressive Model," *Journal of Econometrics*, 114, 261–295. [503]

Hao, K. (1996), "Testing for Structural Change in Cointegrated Regression Models: Some Comparisons and Generalizations," *Econometric Reviews*, 15, 401–429. [503,516]

Kejriwal, M. (2009), "Tests for a Mean Shift With Good Size and Monotonic Power," *Economics Letters*, 102, 78–82. [514]

Kejriwal, M., and Perron, P. (2008a), "Data Dependent Rules for the Selection of the Number of Leads and Lags in the Dynamic OLS Cointegrating Regression," *Econometric Theory*, 24, 1425–1441. [514,521,522]

——— (2008b), "The Limit Distribution of the Estimates in Cointegrated Regression Models With Multiple Structural Changes," *Journal of Econometrics*, 146, 59–73. [503,504]

Kuo, B.-S. (1998), "Test for Partial Parameter Instability in Regressions With $I(1)$ Processes," *Journal of Econometrics*, 86, 337–368. [503]

Nyblom, J. (1989), "Testing the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230. [517]

Perron, P. (2006), "Dealing With Structural Breaks," in *Palgrave Handbook of Econometrics*, eds. K. Patterson and T. C. Mills, New York: Palgrave Macmillan, pp. 278–352. [503]

Perron, P., and Zhu, X. (2005), "Structural Breaks With Deterministic and Stochastic Trends," *Journal of Econometrics*, 129, 65–119. [503]

Phillips, P. C. B., and Hansen, B. E. (1990), "Statistical Inference in Instrumental Variables Regression With $I(1)$ Processes," *Review of Economic Studies*, 57, 99–125. [503]

Qu, Z. (2007), "Searching for Cointegration in a Dynamic System," *Econometrics Journal*, 10, 580–604. [503]

Qu, Z., and Perron, P. (2007), "Estimating and Testing Structural Changes in Multivariate Regressions," *Econometrica*, 75, 459–502. [504,518]

Quintos, C. E., and Phillips, P. C. B. (1993), "Parameter Constancy in Cointegrated Regressions," *Empirical Economics*, 18, 675–706. [503]

Saikkonen, P. (1991), "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory*, 7, 1–21. [514,521]

Seo, B. (1998), "Tests for Structural Change in Cointegrated Systems," *Econometric Theory*, 14, 222–259. [503]

Stock, J. H., and Watson, M. W. (1993), "A Simple Estimator of Cointegrated Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783–820. [514]

Vogelsang, T. J. (1999), "Sources of Nonmonotonic Power When Testing for a Shift in Mean of a Dynamic Time Series," *Journal of Econometrics*, 88, 283–299. [512,516]