Econometric Theory, **25**, 2009, 1754–1792. Printed in the United States of America.
doi:10.1017/S0266466609990326

GLS-BASED UNIT ROOT TESTS WITH MULTIPLE STRUCTURAL BREAKS UNDER BOTH THE NULL AND THE ALTERNATIVE HYPOTHESES

JOSEP LLUÍS CARRION-I-SILVESTRE
University of Barcelona

DUKPA KIM
University of Virginia

PIERRE PERRON
Boston University

Perron (1989, *Econometrica* 57, 1361–1401) introduced unit root tests valid when a break at a known date in the trend function of a time series is present. In particular, they allow a break under both the null and alternative hypotheses and are invariant to the magnitude of the shift in level and/or slope. The subsequent literature devised procedures valid in the case of an unknown break date. However, in doing so most research, in particular the commonly used test of Zivot and Andrews (1992, *Journal of Business & Economic Statistics* 10, 251–270), assumed that if a break occurs it does so only under the alternative hypothesis of stationarity. This is undesirable for several reasons. Kim and Perron (2009, *Journal of Econometrics* 148, 1–13) developed a methodology that allows a break at an unknown time under both the null and alternative hypotheses. When a break is present, the limit distribution of the test is the same as in the case of a known break date, allowing increased power while maintaining the correct size. We extend their work in several directions: (1) we allow for an arbitrary number of changes in both the level and slope of the trend function; (2) we adopt the quasi–generalized least squares detrending method advocated by Elliott, Rothenberg, and Stock (1996, *Econometrica* 64, 813–836) that permits tests that have local asymptotic power functions close to the local asymptotic Gaussian power envelope; (3) we consider a variety of tests, in particular the class of *M*-tests introduced in Stock (1999, *Cointegration, Causality, and Forecasting: A Festschrift for Clive W.J. Granger*) and analyzed in Ng and Perron (2001, *Econometrica* 69, 1519–1554).

This paper was prepared for the conference in honor of Professor Paul Newbold at the University of Nottingham, 21–22 September 2007. Perron acknowledges financial support for this work from the National Science Foundation under grant SES-0649350. Carrion-i-Silvestre acknowledges financial support from the Ministerio de Ciencia y Tecnología under grant SEJ2005-08646/ECON. We thank two referees for useful comments. Address correspondence to Josep Lluís Carrion-i-Silvestre, AQR Research Group, Department of Econometrics, Statistics and Spanish Economy, University of Barcelona, Av. Diagonal, 690, Barcelona, Spain 08034; e-mail: carrion@ub.edu.

 © 2009 Cambridge University Press 0266-4666/09 $15.00

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

1. INTRODUCTION

Professor Paul Newbold has made important contributions in econometrics, in particular in the area of the analysis of nonstationary time series. I (Perron) have been honored and privileged to have him participate in a research agenda that I put forward in Perron (1989). This paper is, in part, a product motivated and made possible by some of his work.

It is well known that a break in the deterministic trend affects the outcome of unit root tests. Perron (1989) showed that a standard Dickey–Fuller (1979) (DF) type unit root test is not consistent if the alternative is a stationary noise component with a break in the slope of the deterministic trend. His main point is that the existence of an exogenous shock that has a permanent effect will lead to a nonrejection of the unit root hypothesis even though it is not true. Of interest also is the fact that Leybourne, Mills, and Newbold (1998) and Leybourne and Newbold (2000) analyzed the effect of a break on a standard DF test under the unit root null hypothesis and showed that size distortions can occur, especially when the break is early in the sample. Perron (1989, 1990) proposed alternative unit root tests that allow the possibility of a break under both the null and alternative hypotheses. These tests have less power than a standard DF type test when there is no break. Nonetheless, they have a correct size asymptotically and are consistent whether there is a break or not. Moreover, they are invariant to the break parameters, and thus their performance does not depend on the magnitude of the break. The most controversial assumption, however, is that its timing is known a priori (see Christiano, 1992).

For the Perron (1989, 1990) test procedures to be valid, the break date should be chosen independently of the given data. Whenever a systematic search for a break is done, the limiting distributions in Perron (1989, 1990) are no longer appropriate. Historical facts can often provide good guidance in choosing a break date independently of the given data. Even in that case, it is very likely that an imprecise break date is used. Hecq and Urbain (1993) showed, by simulations, that the use of an incorrect break date in the Perron (1990) test, applicable with nontrending data, causes size distortions and power loss, though as shown by Montañés (1997) this effect disappears asymptotically. Montañés and Olloqui (1999) extended the analysis to the Perron (1989) tests, applicable with trending data, and showed that a loss of power occurs even in large samples. Kim, Leybourne, and Newbold (2000) examined the effect of using a wrong break date under the null hypothesis.

Zivot and Andrews (1992) provided methods that treat the occurrence of the break date as unknown and that have become quite popular (see also Perron and Vogelsang, 1992a, 1992b; and parts of Perron, 1997). However, Professor Newbold and his co-authors have clearly recognized the potential pitfalls of their approach. In this line of work, a break is not allowed under the null hypothesis, only under the alternative, mostly because of the theoretical apparatus adopted. This means, e.g., that under the null hypothesis a level shift must be viewed

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

as coming from the tail of the distribution of the data generating process, and a slope change involves errors with a different mean in some subsamples. This framework is convenient because it allows establishing various unit root testing procedures; e.g., by minimizing the *t*-statistic related to the sum of the autoregressive coefficients over each possible break date. This approach is, however, contrary to the original motivation of Perron (1989), which was to devise procedures that are invariant to the magnitude of the shift in level and/or slope. In particular, if a change is present it is allowed under both the null and alternative hypotheses.

The existence of a structural break in the trend function is a problem of long horizon data; it can happen whether the noise component is stationary or has a unit root. As argued in Nunes, Newbold, and Kuan (1997) and Harvey, Leybourne, and Newbold (2001) (see also Vogelsang and Perron, 1998; Lee and Strazicich, 2001), if the noise component has a unit root and a break occurs in the trend function, the Zivot and Andrews (1992) type test statistics often diverge or are not invariant to break parameters. This is a natural consequence of not permitting a break under the null hypothesis. An added consequence is that these types of tests have substantially less power than the Perron (1989) tests because they do not fully utilize the information about the break when one is present.

Despite these shortcomings, the method of Zivot and Andrews (1992) has remained popular in empirical work, probably in part because of the lack of sound statistical methods that could tackle the problem of allowing for changes in the trend function at unknown times under both the null and alternative hypotheses. Using recent developments on structural change problems related to nonstationary data by Perron and Zhu (2005) and Perron and Yabu (2009), Kim and Perron (2009) developed a methodology to devise new test procedures that allow a break in the trend function at an unknown time under both the null and alternative hypotheses. Also, when a break is present, the limit distribution of the test is the same as in the case of a known break date, thereby allowing increased power while maintaining the correct size. Simulation experiments confirm that it offers an improvement over commonly used methods in small samples.

In this paper, we extend the work of Kim and Perron (2009) in several directions: (1) we allow for an arbitrary number of changes in both the level and slope of the trend function; (2) we adopt the so-called quasi–generalized least squares (quasi-GLS) detrending method advocated by Elliott, Rothenberg, and Stock (1996) that permits tests that have local asymptotic power functions close to the local asymptotic Gaussian power envelope;[1] (3) we consider a variety of tests, in particular the class of *M*-tests introduced in Stock (1999) and analyzed in Ng and Perron (2001). On the other hand, we restrict our analysis to the case of the so-called additive outlier (AO) models as defined in Perron (1989). Hatanaka and Yamada (1999) is related to the approach of Kim and Perron (2009) and deals with two breaks for the so-called innovational outlier (IO) type model. In independent work, Harris, Harvey, Leybourne, and Taylor (2009) consider a similar problem. Their analysis is, however, restricted to a single break, and they propose

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

using a DF type test with different methods to obtain tests valid with or without a break. Our studies are therefore complementary.

The paper is organized as follows. In Section 2, we present the model with multiple structural breaks in the trend function. Section 3 discusses the feasible point optimal test assuming the break dates to be known and provides details on the methods to construct the relevant noncentrality parameter arising in the quasi-differencing procedures used. Section 4 analyzes the *M*-class and related unit root tests allowing for multiple breaks. Section 5 considers the case with unknown break dates. It shows that if these are estimated by minimizing the sum of squared residuals from the appropriate GLS regression, the limit distributions of the tests are the same as in the known break date case, provided breaks are present. It also provides details on how to compute the estimate of the break dates. Section 6 provides preliminary simulations showing that the tests perform well but that they exhibit important size distortions when no break occurs. A solution to this problem is offered in Section 7 based on a pretest for changes in the slope of a trend function allowing the noise component to be stationary or integrated based on the work of Perron and Yabu (2009). We show that the resulting procedure involving the pretest has good size and power superior to that of alternative methods. Section 8 offers brief concluding remarks and an Appendix the proofs of various theoretical results.

2. THE MODEL

Let y_t be a stochastic process generated according to

$$y_t = d_t + u_t, \tag{1}$$

$$u_t = \alpha u_{t-1} + v_t, \qquad t = 0, \ldots, T, \tag{2}$$

where $\{u_t\}$ is an unobserved mean-zero process. We assume that $u_0 = 0$, although the results generally hold for the weaker requirement that $\mathrm{E}\left(u_0^2\right) < \infty$. The disturbance term v_t is defined by $v_t = \sum_{i=0}^{\infty} \gamma_i \eta_{t-i}$ with $\sum_{i=0}^{\infty} i\,|\gamma_i| < \infty$ and $\{\eta_t\}$ a martingale difference sequence adapted to the filtration $F_t = \sigma\text{-}field\{\eta_{t-i}; i \geq 0\}$. We define the long-run and short-run variance as $\sigma^2 = \sigma_\eta^2 \gamma\,(1)^2$ and $\sigma_\eta^2 = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \mathrm{E}\left(\eta_t^2\right)$, respectively.

We consider three models: Model 0 ("level shift" or "crash"), Model I ("slope change" or "changing growth"), and Model II ("mixed change").[2] Let $DU_t(T_j^0) = 1$ and $DT_t^*(T_j^0) = (t - T_j^0)$ for $t > T_j^0$ and 0 elsewhere, with $T_j^0 = [T\lambda_j^0]$ denoting the jth break date, with $[\cdot]$ the integer part and $\lambda_j^0 \equiv T_j^0/T \in (0,1)$ the break fraction parameter. As a matter of notation, the true break fractions and break dates are denoted with a superscript zero, and their estimates are denoted with a hat. We also use the convention that $T_0^0 = 0$ and $T_{m+1}^0 = T$. We collect the m break fraction parameters in the vector $\lambda^0 = (\lambda_1^0, \ldots, \lambda_m^0)'$. For now it is assumed that the break dates are known; this will be relaxed later.

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

The deterministic component in (1) is given by

$$d_t = z_t'(T_0^0)\psi_0 + z_t'(T_1^0)\psi_1 + \cdots + z_t'(T_m^0)\psi_m \equiv z_t'(\lambda^0)\psi, \tag{3}$$

where $z_t(\lambda^0) = [z_t'(T_0^0), \ldots, z_t'(T_m^0)]'$ and $\psi = (\psi_0', \ldots, \psi_m')'$. The deterministic components and associated coefficients are defined by $z_t(T_0^0) \equiv z_t(0) = (1, t)'$ with $\psi_0 = (\mu_0, \beta_0)'$ and

$$z_t(T_j^0) = \begin{cases} DU_t(T_j^0), & \text{in Model 0,} \\ DT_t^*(T_j^0), & \text{in Model I,} \\ (DU_t(T_j^0), DT_t^*(T_j^0))', & \text{in Model II,} \end{cases}$$

for $1 \leq j \leq m$, with $\psi_j = \mu_j$ in Model 0, $\psi_j = \beta_j$ in Model I, and $\psi_j = (\mu_j, \beta_j)'$ in Model II.

For Models 0 and II, we also consider the case where the magnitude of level shifts gets larger as the sample size grows, i.e., $(\mu_1, \ldots, \mu_m) = T^{1/2+\eta}(\kappa_1, \ldots, \kappa_m)$, with $\eta > 0$. The models are then labeled as Models 0b and IIb, respectively. These models are useful to capture important properties of the tests in finite samples. In Models 0 and II, the level shifts belong to the class of "slowly evolving trend" defined by Elliott et al. (1996) and have no effect on the asymptotic size and power of the tests. When the magnitude of the shifts is nonnegligible, this typically implies that the derived asymptotic distribution is a bad approximation to the finite-sample distribution. In Models 0b and IIb, the level shifts do not belong to the class of "slowly evolving trend." As shown in Harvey et al. (2001) this framework provides better approximations. A second feature of importance is related to the estimation of the unknown break fractions. As shown by Perron and Zhu (2005), the rate of convergence increases when the level shifts are modeled to increase as the sample size grows. This has important implications for the properties of the unit root tests, as will be shown in the following sections.

Remark 1. It is possible to assume that there is no trending deterministic component in Model 0. We do not explicitly consider this case in the subsequent analysis, but it should be understood that most of our results pertaining to Models 0 and 0b can readily be applied to the level shift model with no time trend.

The so-called GLS-detrended unit root test statistics are based on the use of the quasi-differenced variables $y_t^{\bar{\alpha}}$ and $z_t^{\bar{\alpha}}(\lambda^0)$ defined by $y_1^{\bar{\alpha}} = y_1, z_1^{\bar{\alpha}}(\lambda^0) = z_1(\lambda^0)$, and

$$y_t^{\bar{\alpha}} = (1 - \bar{\alpha}L)\, y_t,$$

$$z_t^{\bar{\alpha}}(\lambda^0) = (1 - \bar{\alpha}L)\, z_t(\lambda^0)$$

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

for $t = 2, \ldots, T$ with $\bar{\alpha} = 1 + \bar{c}/T$ where $\bar{c}$ is a noncentrality parameter to be defined later. Once the data have been transformed, the parameters ψ, associated with the deterministic components, can be estimated by minimizing the following objective function:

$$S^*\left(\psi, \bar{\alpha}, \lambda^0\right) = \sum_{t=1}^{T} \left(y_t^{\bar{\alpha}} - \psi' z_t^{\bar{\alpha}}(\lambda^0)\right)^2. \tag{4}$$

We denote the minimum of this function by $S\left(\bar{\alpha}, \lambda^0\right)$.

3. FEASIBLE POINT OPTIMAL TEST WITH MULTIPLE STRUCTURAL BREAKS

The choice of the noncentrality parameter $\bar{c}$ is related to the Gaussian point optimal statistic to test the null hypothesis of $\alpha = 1$ in (2) against the alternative hypothesis that $\alpha = \bar{\alpha}$, as suggested by Elliott et al. (1996). Following their analysis and Perron and Rodríguez (2003), the feasible point optimal statistic is given by

$$P_T^{GLS}\left(\lambda^0\right) = \left\{S\left(\bar{\alpha}, \lambda^0\right) - \bar{\alpha} S\left(1, \lambda^0\right)\right\} / s^2(\lambda^0), \tag{5}$$

where $s^2(\lambda^0)$ is an estimate of the spectral density at frequency zero of v_t. Following Ng and Perron (2001) and Perron and Ng (1998), we use an autoregressive estimate defined by

$$s(\lambda^0)^2 = s_{ek}^2 / \left(1 - \sum_{j=1}^{k} \hat{b}_j\right)^2, \tag{6}$$

where $s_{ek}^2 = (T-k)^{-1} \sum_{t=k+1}^{T} \hat{e}_{t,k}^2$ and $\{\hat{b}_j, \hat{e}_{t,k}\}$ obtained from the ordinary least squares (OLS) regression

$$\Delta \tilde{y}_t = b_0 \tilde{y}_{t-1} + \sum_{j=1}^{k} b_j \Delta \tilde{y}_{t-j} + e_{t,k}, \tag{7}$$

with $\tilde{y}_t = y_t - \hat{\psi}' z_t(\lambda^0)$ where $\hat{\psi}$ minimizes (4). The order of the autoregression k is selected using the modified information criteria suggested by Ng and Perron (2001) with the modification proposed by Perron and Qu (2007).

Let $W_c(r)$ be an Ornstein–Uhlenbeck process, i.e., the solution to the stochastic differential equation $dW_c(r) = cW_c(r)\,dr + dW(r)$, with $W_c(0) = 0$ where $W(r)$ is the standard Brownian motion. Denoting by $\Rightarrow$ weak convergence of the associated measure of probability, the limiting distribution of the test $P_T^{GLS}(\lambda^0)$ is given in the following theorem.

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

THEOREM 1. *Let* $\{y_t\}_{t=1}^T$ *be the stochastic process generated according to (1) and (2) with* $\alpha = 1 + c/T$ *and* $s^2(\lambda^0)$ *be a consistent estimate of* σ^2. *Then,*

(i) For Models 0 and 0b:

$$P_T^{GLS}\left(\lambda^0\right) \Rightarrow \bar{c}^2 \int_0^1 V_{c,\bar{c}}^2(r)\,dr + (1-\bar{c})V_{c,\bar{c}}^2(1) \equiv K^{P_T^{GLS}}(c,\bar{c}),$$

where $V_{c,\bar{c}}^2(r) = W_c(r) - r[bW_c(1) + 3(1-b)\int_0^1 sW_c(s)\,ds]$ *and* $b = (1-\bar{c})/(1-\bar{c}+\bar{c}^2/3)$.

(ii) For Models I, II, and IIb:

$$P_T^{GLS}\left(\lambda^0\right) \Rightarrow M\left(c,0,\lambda^0\right) - M\left(c,\bar{c},\lambda^0\right) - 2\bar{c}\int_0^1 W_c(r)\,dW(r)$$

$$+\left(\bar{c}^2 - 2\bar{c}c\right)\int_0^1 W_c(r)^2\,dr - \bar{c} \equiv H^{P_T^{GLS}}\left(c,\bar{c},\lambda^0\right),$$

where $M\left(c,\bar{c},\lambda^0\right) = \bar{V}\left(\lambda^0\right)' A(\lambda^0)^{-1}\bar{V}\left(\lambda^0\right)$, $\bar{V}(\lambda^0) = (V(\lambda_0^0),\ldots, V(\lambda_m^0))'$ *with*

$$V(\lambda_i^0) = (1+\bar{c}\lambda_i^0)[W(1) - W(\lambda_i^0) + (c-\bar{c})\int_{\lambda_i^0}^1 W_c(r)\,dr]$$

$$-\bar{c}\int_{\lambda_i^0}^1 r\,dW(r) - (c-\bar{c})\bar{c}\int_{\lambda_i^0}^1 r\,W_c(r)\,dr$$

and $A(\lambda^0)$ *a symmetric matrix defined by*

$$A(\lambda^0) = \begin{bmatrix} a\left(\lambda_0^0,\lambda_0^0\right) & a\left(\lambda_0^0,\lambda_1^0\right) & \cdots & a\left(\lambda_0^0,\lambda_m^0\right) \\ & a\left(\lambda_1^0,\lambda_1^0\right) & \cdots & a\left(\lambda_1^0,\lambda_m^0\right) \\ & & \ddots & \vdots \\ & & & a\left(\lambda_m^0,\lambda_m^0\right) \end{bmatrix} \tag{8}$$

with

$$a(\lambda_i^0,\lambda_j^0) = (1/6)(1-\lambda_j)\left[6\bar{c}(\lambda_i - 1) + \bar{c}^2\{\lambda_j(3\lambda_i - 1) - 3\lambda_i - \lambda_j^2 + 2\} + 6\right].$$

The proof for Model II is given in the Appendix. Because Models 0 and I can be viewed as a special case of Model II, no separate proof is provided. The limiting distribution in (i) is the same as that of the linear time trend model with no break, which can be found in Elliott et al. (1996). Because the break dates are assumed known here, the test statistic $P_T^{GLS}(\lambda^0)$ is exactly invariant to the value

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

of the coefficients associated with all regressors including those pertaining to the change in the trend. Hence, there is no distinction between Models 0 and 0b and between Models II and IIb. The limiting distributions of the tests for Models I, II, and IIb depend on the number of breaks and the vector of break fractions.

We now turn to the choice of the noncentrality parameter $\bar{c}$. Using the limiting distributions in Theorem 1, we can obtain the local Gaussian power envelope for the various cases. For Models I, II, and IIb, it is defined by $\pi^*(c) = \Pr[H^{P_T^{GLS}}(c, c, \lambda^0) < b^{P_T^{GLS}}(c, \lambda^0)]$, where, with υ the size of the test, $b^{P_T^{GLS}}(c, \lambda^0)$ is such that $\Pr[H^{P_T^{GLS}}(0, c, \lambda^0) < b^{P_T^{GLS}}(c, \lambda^0)] = \upsilon$. This power envelope allows us to obtain the "optimal" noncentrality parameter $\bar{c}$ that Elliott et al. (1996) recommended choosing as the value that yields tests with a 50% asymptotic power, i.e., $\bar{c}$ is such that $\Pr[H^{P_T^{GLS}}(\bar{c}, \bar{c}) < b^{P_T^{GLS}}(\bar{c})] = 0.5$.

For Models I, II, and IIb the parameter $\bar{c}$ depends on the number of structural breaks and their positions. Table 1 presents the relevant values for $m = 1$ and 2 breaks for all possible combinations of break fraction vectors $\lambda^0 = (\lambda_1^0, \lambda_2^0)'$, with $\lambda_i^0 = \{0.1, 0.2, \ldots, 0.9\}$. These were obtained by simulations using 1,000 steps to approximate the Wiener process and 10,000 replications. This should be sufficient for most applications, but we also constructed a response surface analysis that can provide the relevant values for up to 5 breaks and any combinations of break fractions. The details of its construction and the results are available on the corresponding author's web page, www.eco.ub.es/∼carrion. Note that for Models 0 and 0b, a similar analysis holds, though much simpler because the limit distribution is invariant to any break parameters. The results are described in Elliott et al. (1996), and $\bar{c} = -13.5$ (if no trend is present $\bar{c} = -7$).

TABLE 1. Values for the noncentrality parameter $\bar{c}$

$m = 1$									
λ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\bar{c}$	−16.0	−17.6	−18.2	−18.2	−17.8	−17.2	−16.4	−15.6	−14.8
$m = 2$									
$\lambda_1 \backslash \lambda_2$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1		−18.0	−19.7	−20.6	−21.0	−20.9	−20.5	−19.6	−18.1
0.2			−19.5	−20.9	−21.5	−21.6	−21.2	−20.4	−19.2
0.3				−20.3	−21.5	−21.8	−21.5	−20.8	−19.6
0.4					−20.5	−21.3	−21.4	−20.8	−19.6
0.5						−20.1	−20.6	−20.3	−19.3
0.6							−19.2	−19.4	−18.7
0.7								−18.0	−17.8
0.8									−16.5

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

4. THE *M*-CLASS AND RELATED UNIT ROOT TESTS

Following Perron and Rodríguez (2003), we use the *M*-class of tests analyzed in Ng and Perron (2001) allowing for multiple structural breaks, defined by

$$MZ_{\alpha}^{GLS}\left(\lambda^{0}\right)=\left(T^{-1}\tilde{y}_{T}^{2}-s\left(\lambda^{0}\right)^{2}\right)\left(2T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^{2}\right)^{-1}, \tag{9}$$

$$MSB^{GLS}\left(\lambda^{0}\right)=\left(s(\lambda^{0})^{-2}T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^{2}\right)^{1/2}, \tag{10}$$

$$MZ_{t}^{GLS}\left(\lambda^{0}\right)=\left(T^{-1}\tilde{y}_{T}^{2}-s\left(\lambda^{0}\right)^{2}\right)\left(4s\left(\lambda^{0}\right)^{2}T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^{2}\right)^{-1/2}, \tag{11}$$

with $\tilde{y}_t = y_t - \hat{\psi}' z_t(\lambda^0)$, where $\hat{\psi}$ minimizes (4) and $s(\lambda^0)^2$ is defined in (6). Another statistic, following Ng and Perron (2001), is a modified feasible point optimal test defined by

$$MP_{T}^{GLS}\left(\lambda^{0}\right)=\left[\bar{c}^{2}T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^{2}+(1-\bar{c})T^{-1}\tilde{y}_{T}^{2}\right]/s\left(\lambda^{0}\right)^{2}.$$

This test is based on the same motivation that leads to the definition of the *M*-tests in Stock (1999), namely, to provide functionals of sample moments that have the same asymptotic distributions as well-known unit root tests. The $MP_T^{GLS}(\lambda^0)$ is important because its limiting distribution coincides with that of the feasible point optimal test.

The following theorem provides the limit null distribution of the various tests considered.

THEOREM 2. *Let $\{y_t\}_{t=1}^T$ be the stochastic process generated according to (1) and (2) with $\alpha = 1 + c/T$ and $s^2(\lambda^0)$ be a consistent estimate of σ^2. Then,*

(i) For Models 0 and 0b:

$$MZ_{\alpha}^{GLS}(\lambda^{0})\Rightarrow 0.5\left(V_{c,\bar{c}}(1)^{2}-1\right)\left(\int_{0}^{1}V_{c,\bar{c}}(r)^{2}\,dr\right)^{-1},$$

$$MSB^{GLS}(\lambda^{0})\Rightarrow\left(\int_{0}^{1}V_{c,\bar{c}}(r)^{2}\,dr\right)^{1/2},$$

where $V_{c,\bar{c}}^2(r) = W_c(r) - r[bW_c(1) + 3(1-b)\int_0^1 sW_c(s)\,ds]$ and $b = (1-\bar{c})/(1-\bar{c}+\bar{c}^2/3)$.

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

TABLE 2A. Asymptotic critical values for the $P_T^{GLS}(\lambda^0)$ and $MP_T^{GLS}(\lambda^0)$ tests

					$m=1$				
λ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1%	4.354	4.807	5.042	5.079	4.949	4.704	4.423	4.191	4.075
2.5%	5.235	5.732	5.978	6.000	5.837	5.556	5.250	5.016	4.933
5%	6.238	6.759	6.992	6.986	6.789	6.476	6.146	5.910	5.865
10%	7.492	8.076	8.344	8.335	8.105	7.744	7.373	7.118	7.086
$m=2$			1% (top triangle), 2.5% (bottom triangle)						
0.1		4.723	5.213	5.536	5.71	5.724	5.573	5.266	4.810
0.2	5.520		5.134	5.585	5.819	5.866	5.743	5.475	5.087
0.3	6.095	5.948		5.471	5.829	5.932	5.841	5.596	5.241
0.4	6.471	6.496	6.328		5.617	5.848	5.812	5.595	5.252
0.5	6.670	6.768	6.772	6.502		5.524	5.622	5.461	5.135
0.6	6.686	6.817	6.892	6.798	6.403		5.214	5.197	4.923
0.7	6.514	6.678	6.783	6.759	6.548	6.063		4.772	4.654
0.8	6.155	6.384	6.516	6.516	6.373	6.081	5.596		4.328
0.9	5.593	5.954	6.143	6.157	6.025	5.795	5.511	5.172	
$m=2$			5% (top triangle), 10% bottom triangle						
0.1		6.478	7.034	7.398	7.592	7.607	7.434	7.072	6.511
0.2	7.791		6.897	7.431	7.708	7.765	7.624	7.310	6.844
0.3	8.376	8.210		7.264	7.712	7.850	7.749	7.461	7.037
0.4	8.733	8.757	8.568		7.435	7.747	7.727	7.475	7.070
0.5	8.903	9.013	9.012	8.719		7.338	7.498	7.328	6.951
0.6	8.887	9.039	9.123	9.019	8.600		6.997	7.015	6.719
0.7	8.697	8.873	8.997	8.978	8.754	8.245		6.524	6.424
0.8	8.361	8.563	8.706	8.722	8.582	8.276	7.772		6.094
0.9	7.902	8.159	8.322	8.345	8.226	7.998	7.705	7.358	

Note: For 1% and 5% (resp., 2.5% and 10%), the value of λ_1 is on the horizontal (resp., vertical) axis and that of λ_2 on the vertical (resp., horizontal) axis.

(ii) For Models I, II, and IIb:

$$MZ_\alpha^{GLS}(\lambda^0) \Rightarrow 0.5\left(V_{c,\bar{c}}(1,\lambda^0)^2 - 1\right)\left(\int_0^1 V_{c,\bar{c}}(r,\lambda^0)^2\,dr\right)^{-1}$$
$$\equiv H^{MZ_\alpha^{GLS}}(c,\bar{c},\lambda^0),$$
$$MSB^{GLS}(\lambda^0) \Rightarrow \left(\int_0^1 V_{c,\bar{c}}(r,\lambda^0)^2\,dr\right)^{1/2} \equiv H^{MSB^{GLS}}(c,\bar{c},\lambda^0),$$

where $V_{c,\bar{c}}(r,\lambda^0) = W_c(r) - z_2(r)\,A(\lambda^0)^{-1}\bar{V}(\lambda^0)$ *with* $A(\lambda^0)$ *and* $\bar{V}(\lambda^0)$ *as defined in Theorem 1, and* $z_2(r) = (r,\ (r-\lambda_1^0)1(r > \lambda_1^0),\ldots,(r-\lambda_m^0)1(r > \lambda_m^0))$, *where* $1(\cdot)$ *is the indicator function.*

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

TABLE 2B. Asymptotic critical values for the $MZ_t^{GLS}(\lambda^0)$ test

					$m=1$				
λ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1%	−3.800	−3.942	−3.981	−3.969	−3.938	−3.897	−3.843	−3.763	−3.639
2.5%	−3.485	−3.650	−3.702	−3.695	−3.662	−3.618	−3.561	−3.477	−3.344
5%	−3.232	−3.396	−3.448	−3.445	−3.419	−3.382	−3.328	−3.240	−3.094
10%	−2.950	−3.118	−3.171	−3.166	−3.137	−3.096	−3.039	−2.949	−2.801
$m=2$				1% (top triangle), 2.5% (bottom triangle)					
0.1	−4.131	−4.276	−4.338	−4.347	−4.326	−4.285	−4.222	−4.124	
0.2	−3.812		−4.288	−4.384	−4.415	−4.403	−4.363	−4.298	−4.203
0.3	−3.980	−3.965		−4.337	−4.403	−4.415	−4.385	−4.321	−4.226
0.4	−4.048	−4.084	−4.019		−4.325	−4.371	−4.365	−4.313	−4.223
0.5	−4.059	−4.118	−4.105	−4.011		−4.278	−4.304	−4.277	−4.202
0.6	−4.040	−4.105	−4.116	−4.074	−3.964		−4.203	−4.206	−4.158
0.7	−4.001	−4.066	−4.083	−4.066	−4.007	−3.887		−4.100	−4.079
0.8	−3.933	−4.003	−4.022	−4.012	−3.977	−3.905	−3.775		−3.960
0.9	−3.807	−3.902	−3.927	−3.918	−3.892	−3.845	−3.761	−3.619	
$m=2$				5% (top triangle), 10% bottom triangle					
0.1		−3.566	−3.729	−3.797	−3.812	−3.801	−3.768	−3.699	−3.558
0.2	−3.266		−3.723	−3.836	−3.872	−3.865	−3.834	−3.775	−3.665
0.3	−3.425	−3.437		−3.780	−3.861	−3.875	−3.849	−3.792	−3.693
0.4	−3.499	−3.553	−3.502		−3.777	−3.836	−3.832	−3.781	−3.684
0.5	−3.524	−3.598	−3.588	−3.502		−3.737	−3.777	−3.747	−3.657
0.6	−3.519	−3.600	−3.611	−3.566	−3.463		−3.665	−3.678	−3.609
0.7	−3.489	−3.571	−3.589	−3.565	−3.503	−3.388		−3.551	−3.524
0.8	−3.417	−3.507	−3.528	−3.512	−3.470	−3.395	−3.269		−3.377
0.9	−3.272	−3.387	−3.417	−3.405	−3.372	−3.319	−3.232	−3.089	

Note: For 1% and 5% (resp., 2.5% and 10%), the value of λ_1 is on the horizontal (resp., vertical) axis and that of λ_2 on the vertical (resp., horizontal) axis.

(iii) The limiting distribution of $MZ_t^{GLS}(\lambda^0)$ in all models can be obtained in view of the fact that $MZ_t^{GLS}(\lambda^0) = MZ_\alpha^{GLS}(\lambda^0) \cdot MSB^{GLS}(\lambda^0)$, denoted by $H^{MZ_t^{GLS}}(c, \bar{c}, \lambda^0)$.

Again, the limiting distribution in (i) is the same as that of the linear time trend model with no break given in Ng and Perron (2001) because of the invariance of the tests to the break parameters in these models. This is not the case for Models I, II, and IIb, for which the limiting distribution depends on the number and location of the break points. The proof of Model II is given in the Appendix, and the proof for the other models easily follows.

For Models I, II, and IIb, the limit distributions depend on the number of breaks and their positions. Tables 2A–2D present the 1, 2.5, 5, and 10% percentiles

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

TABLE 2C. Asymptotic critical values for the $MSB^{GLS}(\lambda^0)$ test

					$m=1$				
λ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1%	0.130	0.125	0.123	0.124	0.125	0.127	0.129	0.132	0.137
2.5%	0.142	0.135	0.133	0.133	0.135	0.137	0.139	0.143	0.149
5%	0.152	0.144	0.142	0.143	0.144	0.147	0.149	0.154	0.161
10%	0.166	0.157	0.154	0.155	0.157	0.159	0.163	0.169	0.178
$m=2$			1% (top triangle), 2.5% (bottom triangle)						
0.1		0.120	0.115	0.114	0.114	0.115	0.116	0.118	0.121
0.2	0.129		0.115	0.113	0.112	0.113	0.114	0.116	0.119
0.3	0.124	0.124		0.114	0.112	0.112	0.113	0.115	0.118
0.4	0.122	0.121	0.123		0.115	0.114	0.114	0.115	0.118
0.5	0.122	0.120	0.120	0.123		0.116	0.115	0.116	0.119
0.6	0.122	0.121	0.120	0.121	0.125		0.118	0.118	0.120
0.7	0.124	0.122	0.122	0.122	0.124	0.127		0.121	0.122
0.8	0.126	0.124	0.124	0.124	0.125	0.127	0.132		0.126
0.9	0.131	0.128	0.127	0.127	0.128	0.130	0.133	0.138	
$m=2$			5% (top triangle), 10% bottom triangle						
0.1		0.138	0.132	0.130	0.129	0.130	0.131	0.134	0.140
0.2	0.150		0.132	0.128	0.127	0.128	0.129	0.132	0.136
0.3	0.143	0.143		0.131	0.128	0.128	0.129	0.131	0.135
0.4	0.140	0.138	0.140		0.131	0.129	0.130	0.132	0.135
0.5	0.140	0.137	0.137	0.141		0.133	0.131	0.133	0.137
0.6	0.140	0.137	0.137	0.138	0.143		0.135	0.135	0.138
0.7	0.142	0.139	0.138	0.139	0.141	0.146		0.140	0.142
0.8	0.145	0.141	0.141	0.141	0.143	0.146	0.152		0.148
0.9	0.153	0.147	0.146	0.146	0.148	0.150	0.154	0.162	

Note: For 1% and 5% (resp., 2.5% and 10%), the value of λ_1 is on the horizontal (resp., vertical) axis and that of λ_2 on the vertical (resp., horizontal) axis.

of the limit distributions of the tests $m = 1$ and 2 breaks for all possible combinations of break fraction vectors $\lambda^0 = (\lambda_1^0, \lambda_2^0)'$ with $\lambda_i^0 = \{0.1, 0.2, \ldots, 0.9\}$ (again, obtained via simulations using 1,000 steps to approximate the Wiener process and 10,000 replications). We also constructed a response surface analysis that can provide the relevant values for up to 5 breaks. The details of its construction and the results are available on the corresponding author's web page, www.eco.ub.es/∼carrion.

The asymptotic power functions of the tests are defined by $\pi^*_{J^{GLS}}(c, \bar{c}, \lambda^0) = \Pr\left[H^{J^{GLS}}(c, \bar{c}, \lambda^0) < b^{J^{GLS}}(\bar{c}, \lambda^0)\right]$ for $J = MZ_\alpha$, MSB, MZ_t with $H^{J^{GLS}}(c, \bar{c}, \lambda^0)$ defined in Theorem 2. The constants $b^{J^{GLS}}(\bar{c}, \lambda^0)$ are such that $\Pr[H^{J^{GLS}}(0, \bar{c}, \lambda^0) < b^{J^{GLS}}(\bar{c}, \lambda^0)] = \upsilon$, the size of the tests. To assess the efficiency in terms of local asymptotic power of the various tests, we consider the case with a single

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

TABLE 2D. Asymptotic critical values for the $MZ_{\alpha}^{GLS}(\lambda^0)$ test

					$m=1$				
λ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1%	−29.315	−31.442	−32.024	−31.828	−31.333	−30.713	−29.873	−28.539	−26.400
2.5%	−24.730	−26.811	−27.544	−27.518	−27.131	−26.544	−25.698	−24.372	−22.293
5%	−21.244	−23.322	−24.018	−23.969	−23.612	−23.113	−22.393	−21.188	−19.126
10%	−17.815	−19.793	−20.433	−20.350	−19.965	−19.450	−18.743	−17.599	−15.657
$m=2$				1% (top triangle), 2.5% (bottom triangle)					
0.1		−34.600	−36.957	−38.023	−38.147	−37.698	−36.884	−35.727	−34.161
0.2	−29.635		−37.208	−38.881	−39.455	−39.195	−38.368	−37.097	−35.403
0.3	−32.072	−31.932		−38.001	−39.208	−39.418	−38.798	−37.548	−35.792
0.4	−33.127	−33.725	−32.784		−37.720	−38.551	−38.408	−37.405	−35.764
0.5	−33.283	−34.279	−34.104	−32.643		−36.806	−37.238	−36.726	−35.428
0.6	−32.935	−34.038	−34.279	−33.577	−31.806		−35.447	−35.441	−34.674
0.7	−32.231	−33.319	−33.683	−33.400	−32.350	−30.436		−33.666	−33.314
0.8	−31.049	−32.180	−32.551	−32.418	−31.792	−30.539	−28.591		−31.352
0.9	−29.068	−30.458	−30.903	−30.834	−30.415	−29.604	−28.244	−26.244	
$m=2$				5% (top triangle), 10% bottom triangle					
0.1		−25.856	−28.150	−29.163	−29.379	−29.144	−28.555	−27.445	−25.411
0.2	−21.932		−28.123	−29.825	−30.383	−30.243	−29.663	−28.640	−26.924
0.3	−23.932	−24.159		−28.948	−30.194	−30.399	−29.929	−28.935	−27.329
0.4	−24.888	−25.683	−24.966		−28.848	−29.736	−29.612	−28.740	−27.186
0.5	−25.178	−26.270	−26.113	−24.890		−28.176	−28.715	−28.187	−26.768
0.6	−25.044	−26.245	−26.385	−25.715	−24.275		−27.045	−27.150	−26.073
0.7	−24.528	−25.760	−26.014	−25.642	−24.744	−23.210		−25.350	−24.879
0.8	−23.466	−24.763	−25.061	−24.811	−24.210	−23.204	−21.564		−22.848
0.9	−21.514	−23.021	−23.405	−23.218	−22.756	−22.054	−20.930	−19.082	

Note: For 1% and 5% (resp., 2.5% and 10%), the value of λ_1 is on the horizontal (resp., vertical) axis and that of λ_2 on the vertical (resp., horizontal) axis.

break occurring at $\lambda^0 = 0.3$, 0.5, and 0.7. The asymptotic power functions are shown in Figure 1, where the solid line is the power envelope. As can be seen, the local power functions are nearly identical and indeed very close to the Gaussian power envelope. So from this local asymptotic power perspective, all tests are nearly efficient. For the case of Models 0 and 0b, we already know from Ng and Perron (2001) that the local power functions of the various tests considered are very close to the power envelope.

5. THE CASE WITH UNKNOWN BREAK DATES

The analysis so far has assumed that the timing of the structural breaks is known. We need to establish a procedure to estimate them and deduce the effect on the limit distribution of the various unit root tests. Section 5.1 presents the method

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

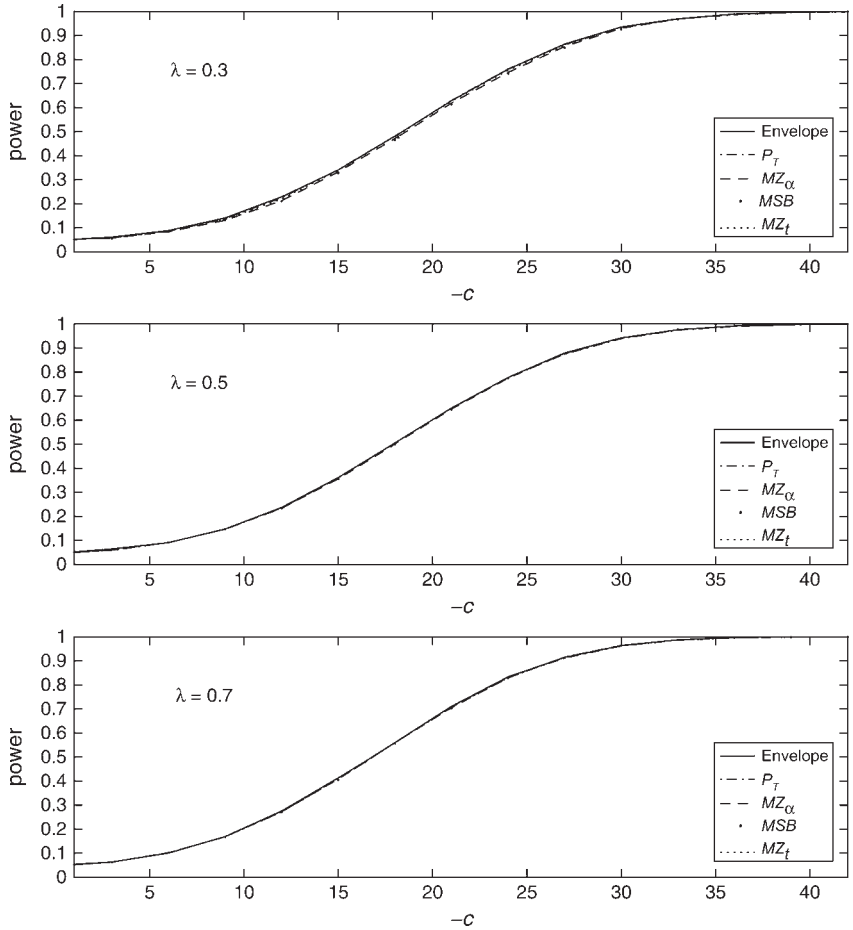


FIGURE 1. Gaussian local power envelope and the asymptotic local power functions.

of estimation, the rate of convergence of the estimates, and the properties of the tests when evaluated at these estimates. Section 5.2 presents an efficient method to compute the estimates when $m > 1$.

5.1. The Limit Distributions of the Tests when Using Estimates of the Break Dates

We propose to estimate the break dates via a global minimization of the sum of squared residuals (SSR) of the GLS-detrended model,[3] i.e., $\hat{\lambda} = \arg\min_{\lambda \in \Lambda(\varepsilon)} S(\bar{\alpha}, \lambda)$, so that

$$S(\bar{\alpha}, \hat{\lambda}) = \min_{\lambda \in \Lambda(\varepsilon)} S(\bar{\alpha}, \lambda), \tag{12}$$

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

where the infimum is taken over all possible break fractions defined on the set

$$\Lambda(\varepsilon)=\{(\lambda_1,\ldots,\lambda_m);\ |\lambda_{i+1}-\lambda_i|\geq\varepsilon\ (i=1,\ldots,m-1),\lambda_1\geq\varepsilon,\lambda_m\leq 1-\varepsilon\},$$

with ε some trimming parameter that dictates the minimal length of a segment. A common value in the related literature is $\varepsilon = 0.15$. The following proposition establishes the consistency and rate of convergence of the estimate of the vector of break fractions λ^0.

Proposition 1. *Let $\{y_t\}_{t=1}^T$ be the stochastic process generated according to (1) and (2) with $\alpha = 1$. Assume that $m > 0$ and $\psi_j \neq 0$ ($j = 1,\ldots,m$), so that there are structural breaks affecting y_t under the null hypothesis. Let $\hat{\lambda} = \arg\min_{\lambda\in\Lambda(\varepsilon)} S(\bar{\alpha},\lambda)$; then, as $T\to\infty$: (i) in Models I and II: $||\hat{\lambda}-\lambda^0|| = O_p\left(T^{-1}\right)$ and (ii) in Models 0b and IIb: $||\hat{\lambda}-\lambda^0|| = o_p\left(T^{-1}\right)$.*

The proof is given in the Appendix. Following Kim and Perron (2009), the next step is to derive the limit distribution of the unit root tests when this estimate of the break fractions is used instead of the true values. The next proposition, proved in the Appendix, demonstrates that the rate of convergence is fast enough to guarantee that we recover the same limit distribution as in the known break date case.

Proposition 2. *Let $\{y_t\}_{t=1}^T$ be the stochastic process generated according to (1) and (2) with $\alpha = 1$. Assume that $m > 0$ and $\psi_j \neq 0$ ($j = 1,\ldots,m$) and that $s(\hat{\lambda})^2$ is a consistent estimate of σ^2. Let $\hat{\lambda} = \arg\min_{\lambda\in\Lambda(\varepsilon)} S(\bar{\alpha},\lambda)$; then the limit distributions of $P_T^{GLS}(\hat{\lambda})$, $MP_T^{GLS}(\hat{\lambda})$, $MZ_\alpha^{GLS}(\hat{\lambda})$, $MSB^{GLS}(\hat{\lambda})$, and $MZ_t^{GLS}(\hat{\lambda})$ are the same as those of $P_T^{GLS}(\lambda^0)$, $MP_T^{GLS}(\lambda^0)$, $MZ_\alpha^{GLS}(\lambda^0)$, $MSB^{GLS}(\lambda^0)$, and $MZ_t^{GLS}(\lambda^0)$, respectively, for all models.*

This result is important because even if the break dates are unknown, the use of the particular estimate $\hat{\lambda}$ considered allows us to obtain unit root tests with the same limit distribution as in the known break date case. Because the latter have a local asymptotic power function close to the power envelope, this implies that whether the break dates are known or not the same optimality properties hold and that in the case of Gaussian errors we cannot do better in terms of local asymptotic power. Note that the result is different from that of Kim and Perron (2009), who considered a framework using OLS based on a regression involving the raw variables (not quasi-detrended). For Models I and II, the rate of convergence was not fast enough to obtain such an equivalence, and to solve the problem they proposed a procedure involving data trimmed around the estimate of the break date. No such device is needed here with the GLS type procedure.

5.2. Computation of the Estimates of the Break Dates

In practice, the computation of the estimates of the break dates defined in (12) is computationally prohibitive using a regular grid search when $m > 2$. Following

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

Bai and Perron (2003), we propose to use a dynamic programming approach. The procedure involves, however, an additional layer of difficulty because the quasi-differencing used in (12) destroys the block-diagonality of the matrix of regressors. We can, nevertheless, recover a block diagonal matrix provided appropriate restrictions are imposed on the coefficients. We provide details for Model II and summarize the relevant conditions for the others at the end of the section. We start by noting that the matrix of regressors is given by (assuming only two breaks):

$$\begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 \\
(1-\bar{\alpha}) & (1-\bar{\alpha}L)2 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
(1-\bar{\alpha}) & (1-\bar{\alpha}L)T_1^0 & 0 & 0 & 0 & 0 \\
(1-\bar{\alpha}) & (1-\bar{\alpha}L)\left(T_1^0+1\right) & 1 & 1 & 0 & 0 \\
(1-\bar{\alpha}) & (1-\bar{\alpha}L)\left(T_1^0+2\right) & (1-\bar{\alpha}) & (1-\bar{\alpha}L)2 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
(1-\bar{\alpha}) & (1-\bar{\alpha}L)T_2^0 & (1-\bar{\alpha}) & (1-\bar{\alpha}L)\left(T_2^0-T_1^0\right) & 0 & 0 \\
(1-\bar{\alpha}) & (1-\bar{\alpha}L)\left(T_2^0+1\right) & (1-\bar{\alpha}) & (1-\bar{\alpha}L)\left(T_2^0-T_1^0+1\right) & 1 & 1 \\
(1-\bar{\alpha}) & (1-\bar{\alpha}L)\left(T_2^0+2\right) & (1-\bar{\alpha}) & (1-\bar{\alpha}L)\left(T_2^0-T_1^0+2\right) & (1-\bar{\alpha}) & (1-\bar{\alpha}L)2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
(1-\bar{\alpha}) & (1-\bar{\alpha}L)T & (1-\bar{\alpha}) & (1-\bar{\alpha}L)\left(T-T_1^0\right) & (1-\bar{\alpha}) & (1-\bar{\alpha}L)\left(T-T_2^0\right)
\end{bmatrix}.$$

For $t=1$, we have

$$y_1^{\bar{\alpha}} = \mu_0 + \beta_0 + u_1,$$

which can be expressed as

$$\begin{aligned}
y_1^{\bar{\alpha}} &= \mu_0 + \beta_0 \pm \mu_0(1-\bar{\alpha}) \pm \beta_0(1-\bar{\alpha}L)1 + u_1 \\
&= \mu_0(1-\bar{\alpha}) + \beta_0(1-\bar{\alpha}L)1 + [\mu_0 + \beta_0 - \mu_0(1-\bar{\alpha}) - \beta_0(1-\bar{\alpha}L)1] + u_1 \\
&= \mu_0(1-\bar{\alpha}) + \beta_0(1-\bar{\alpha}L)1 + [\mu_0 + \beta_0 - \mu_0(1-\bar{\alpha}) - \beta_0(1-\bar{\alpha}L)1]D_1(T_0) \\
&\quad + u_1,
\end{aligned}$$

where $D_t\left(T_j\right) = 1$ for $t = T_j + 1$ and zero otherwise, with the convention that $T_0 = 0$. Therefore, the model for $1 \le t \le T_1^0$ can be written as

$$y_t^{\bar{\alpha}} = \mu_0 1^{\bar{\alpha}} + \beta_0 t^{\bar{\alpha}} + \gamma_0 D_t\left(T_0^0\right) + u_t^{\bar{\alpha}},$$

with $\gamma_0 = \mu_0\bar{\alpha} + \beta_0 - \beta_0(1-\bar{\alpha}L)\cdot 1 = \mu_0\bar{\alpha}$, using the fact that $(1-\bar{\alpha}L)t = t-(1+\bar{c}/T)(t-1) = 1-(\bar{c}/T)(t-1)$. When $t = T_1^0+1$ the quasi-differenced

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

variable is given by

$$
\begin{aligned}
y^{\bar{\alpha}}_{T_1^0+1} &= \mu_0(1-\bar{\alpha}) + \beta_0(1-\bar{\alpha}L)(T_1^0+1) + \mu_1 + \beta_1 + u^{\bar{\alpha}}_{T_1^0+1} \\
&= \mu_0(1-\bar{\alpha}) + \beta_0(1-\bar{\alpha}L)(T_1^0+1) \\
&\quad \pm \mu_1(1-\bar{\alpha}) \pm \beta_1(1-\bar{\alpha}L)(T_1^0+1) + \mu_1 + \beta_1 + u^{\bar{\alpha}}_{T_1^0+1} \\
&= \mu_0(1-\bar{\alpha}) + \beta_0(1-\bar{\alpha}L)(T_1^0+1) + \mu_1(1-\bar{\alpha}) + \beta_1(1-\bar{\alpha}L)(T_1^0+1) \\
&\quad + \left[\mu_1 + \beta_1 - \mu_1(1-\bar{\alpha}) - \beta_1(1-\bar{\alpha}L)(T_1^0+1)\right] D_{T_1^0+1}(T_1^0) + u^{\bar{\alpha}}_{T_1^0+1} \\
&= (\mu_0 + \mu_1 - \beta_1 T_1^0)(1-\bar{\alpha}) + (\beta_0 + \beta_1)(1-\bar{\alpha}L)(T_1^0+1) \\
&\quad + \bar{\alpha}\mu_1 D_{T_1^0+1}(T_1^0) + u^{\bar{\alpha}}_{T_1^0+1},
\end{aligned}
$$

and for $T_1^0 < t \leq T_2^0$ it is given by

$$
y_t^{\bar{\alpha}} = (\mu_0 + \mu_1 - \beta_1 T_1^0)(1-\bar{\alpha}) + (\beta_0 + \beta_1)(1-\bar{\alpha}L)t + \gamma_1 D_t(T_1^0) + u_t^{\bar{\alpha}},
$$

where $\gamma_1 = \bar{\alpha}\mu_1$. A similar expression can be obtained for $T_2^0 < t \leq T$, namely,

$$
\begin{aligned}
y_t^{\bar{\alpha}} &= (\mu_0 + \mu_1 + \mu_2 - \beta_1 T_1^0 - \beta_2 T_2^0)(1-\bar{\alpha}) + (\beta_0 + \beta_1 + \beta_2)(1-\bar{\alpha}L)t \\
&\quad + \gamma_2 D_t(T_2^0) + u_t^{\bar{\alpha}},
\end{aligned}
$$

with $\gamma_2 = \bar{\alpha}\mu_2$. In general, $y_t^{\bar{\alpha}}$ in the jth regime $\left(T_{j-1}^0 < t \leq T_j^0\right)$, $j = 1, \ldots, m+1$, can therefore be written as

$$
y_t^{\bar{\alpha}} = \mu^*(1-\bar{\alpha}) + \beta^*(1-\bar{\alpha}L)t + \gamma_{j-1} D_t(T_{j-1}^0) + u_t^{\bar{\alpha}}, \tag{13}
$$

where the coefficients satisfy the restrictions $\mu^* = \left(\sum_{i=0}^{j-1} \mu_i - \sum_{i=1}^{j-1} \beta_i T_i^0\right)$, $\beta^* = \left(\sum_{i=0}^{j-1} \beta_i\right)$, and $\gamma_{j-1} = \bar{\alpha}\mu_{j-1}$. Then, the moment matrix of the regressors can be expressed via the block diagonal matrix,

$$
\begin{bmatrix}
(1-\bar{\alpha}) & (1-\bar{\alpha}L)1 & 1 & & & & & & & & & 0 \\
\vdots & \vdots & \vdots & & & & & & & & & \\
(1-\bar{\alpha}) & (1-\bar{\alpha}L)T_1^0 & 0 & & & & & & & & & \\
 & & & (1-\bar{\alpha}) & (1-\bar{\alpha}L)\left(T_1^0+1\right) & 1 & & & & & & \\
 & & & \vdots & \vdots & \vdots & & & & & & \\
 & & & (1-\bar{\alpha}) & (1-\bar{\alpha}L)T_2^0 & 0 & & & & & & \\
 & & & & & & (1-\bar{\alpha}) & (1-\bar{\alpha}L)\left(T_2^0+1\right) & 1 & & & \\
 & & & & & & \vdots & \vdots & \vdots & & & \\
0 & & & & & & (1-\bar{\alpha}) & (1-\bar{\alpha}L)T & 0
\end{bmatrix},
$$

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

and the dynamic programming algorithm can be used provided the appropriate restrictions on the coefficients of the system (13) are imposed. With this specification, the estimation of the break dates can be done using an iterative procedure similar to that of Perron and Qu (2006). The exact steps are as follows.

1. Compute initial estimates of the break dates and associated break fractions $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_m)$ and coefficients $\hat{\psi} = (\hat{\psi}_0', \hat{\psi}_1', \ldots, \hat{\psi}_m')'$ using an OLS method applied to (1). This involves a standard application of the algorithm described in Bai and Perron (2003).
2. For a given set of preliminary estimates of the break dates, obtain an initial value of $\bar{c}$.
3. Let $T^*(\psi, r, n) = (T_1^*(\psi, r, n), \ldots, T_r^*(\psi, r, n))$ be the vector of the optimal r break dates using the first n observations for a given vector of coefficients ψ and let $RSSR(T^*(\psi, r, n))$ be the associated restricted sum of squared residuals. Then, compute the restricted sum of squared residuals $RSSR(T^*(\psi, 1, n))$ for $2h \leq n \leq T - (m-1)h$ by

$$RSSR(T^*(\hat{\psi}, 1, n)) = \min_{h \leq j \leq n-h} \left[RSSR^1(1, j) + RSSR^2(j+1, n) \right]$$

and

$$T^*(\hat{\psi}, 1, n) = \arg \min_{h \leq j \leq n-h} \left[RSSR^1(1, j) + RSSR^2(j+1, n) \right],$$

where

$$RSSR^1(1, j) = \sum_{t=1}^{j} (y_t^{\bar{\alpha}} - z_t^{\bar{\alpha}}(T_0)' \hat{\phi}_0 - \hat{\gamma}_0 D_t(T_0))^2,$$

with $\hat{\phi}_0 = \hat{\psi}_0$ and $\hat{\gamma}_0 = \bar{\alpha}\hat{\mu}_0$, and

$$RSSR^2(j+1, n) = \sum_{t=j+1}^{n} \left(y_t^{\bar{\alpha}} - z_t^{\bar{\alpha}}(T_0)' \hat{\phi}_1(j) - \hat{\gamma}_1 D_t(j) \right)^2,$$

with $\hat{\phi}_1(j) = (\hat{\mu}_0 + \hat{\mu}_1 - \hat{\beta}_1 j, \hat{\beta}_0 + \hat{\beta}_1)'$ and $\hat{\gamma}_1 = \bar{\alpha}\hat{\mu}_1$. Then, sequentially compute and store $RSSR(T^*(\hat{\psi}, r, n))$ for $r = 2, \ldots, m-1$, with n ranging from $(r+1)h$ to $T - (m-r)h$. This is done solving

$$RSSR(T^*(\hat{\psi}, r, n)) = \min_{rh \leq j \leq n-h} \left[RSSR(T^*(\hat{\psi}, r-1, n)) + RSSR^{r+1}(j+1, n) \right].$$

The last rth element is

$$T^*(\hat{\psi}, r, n) = \arg \min_{rh \leq j \leq n-h} \left[RSSR(T^*(\hat{\psi}, r-1, n)) + RSSR^{r+1}(j+1, n) \right],$$

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

where

$$RSSR^{r+1}(j+1,n) = \sum_{t=j+1}^{n} (y_t^{\bar{\alpha}} - z_t^{\bar{\alpha}}(T_0)'\hat{\phi}_r(j) - \hat{\gamma}_r D_t(j))^2, \tag{14}$$

with $\hat{\phi}_r(j) = \left(\sum_{i=0}^{r}\hat{\mu}_i - \sum_{i=1}^{r-1}\hat{\beta}_i T_i^*(\hat{\psi}, r-1, j) - \beta_r j, \sum_{i=0}^{r}\hat{\beta}_i\right)'$ and $\hat{\gamma}_r = \bar{\alpha}\mu_r$. Finally compute

$$RSSR(T^*(\hat{\psi}, m-1, n)) = \min_{mh \le j \le T-h} \Big[RSSR(T^*(\hat{\psi}, m-1, n)) + RSSR^{m+1}(j+1, T)\Big],$$

where $RSSR^{m+1}(j+1,T)$ is computed as in (14). Then store the estimated break dates and update $\bar{c}$ accordingly.

4. Repeat steps 2 and 3 until convergence.

Note that for Models 0 and I, the constraints to be imposed are slightly different. For the first regime ($j = 0$) they are the same as for Model II, but for the other regimes, $T_{j-1}^0 < t \le T_j^0$ ($j = 1, \ldots, m$), we have $\mu^* = \sum_{i=1}^{j-1}\mu_i$, $\beta^* = \beta_0$, and $\gamma_{j-1} = \bar{\alpha}\mu_{j-1}$ for Model 0 and $\mu^* = \mu_0 - \sum_{i=1}^{j-1}\beta_i T_i^0$, $\beta^* = \sum_{i=0}^{j-1}\beta_i$, and $\gamma_{j-1} = 0$ for Model I.

6. PRELIMINARY SIMULATIONS

Even though we have assumed the existence of a break in the trend function so far, it is instructive to analyze the properties of the tests. We consider cases with a single break without loss of generality. The data generating process used in the simulations is given by

$$y_t = d_t + u_t; \qquad d_t = \mu_b DU_t(T_1^0) + \beta_b DT_t^*(T_1^0); \qquad u_t = \alpha u_{t-1} + v_t,$$

$v_t \sim i.i.d.\ N(0,1)$, $u_0 = 0$. We specified four values for the magnitude of the level shift $\mu_b = \{0, 0.5, 1, 5\}$. For each value of μ_b, we considered values of β_b ranging from 0 to 3 in increments of 0.2. The sample size is set at $T = \{100, 200, 300\}$, and the results are based on 1,000 replications. The empirical size is analyzed setting $\alpha = 1$, whereas the power is evaluated with $\alpha = \bar{\alpha} = 1 + \bar{c}/T$, where the parameter $\bar{c}$ depends on the break fractions. Hence, in large samples, the power is 50% in all cases. In this paper, we only report results for $\lambda^0 = 0.5$ (the results were similar for $\lambda^0 = 0.3$ and 0.7) and for the tests $P_T^{GLS}(\hat{\lambda})$ (the properties of the other tests are qualitatively similar).

Figure 2 presents the results of the empirical size using the asymptotic critical values at the 5% level of significance drawn from the estimated response surfaces, and Figure 3 presents results for power. For size, the following features are worthy of note. First, the exact size is close to the nominal size as $|\beta_b|$ increases

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

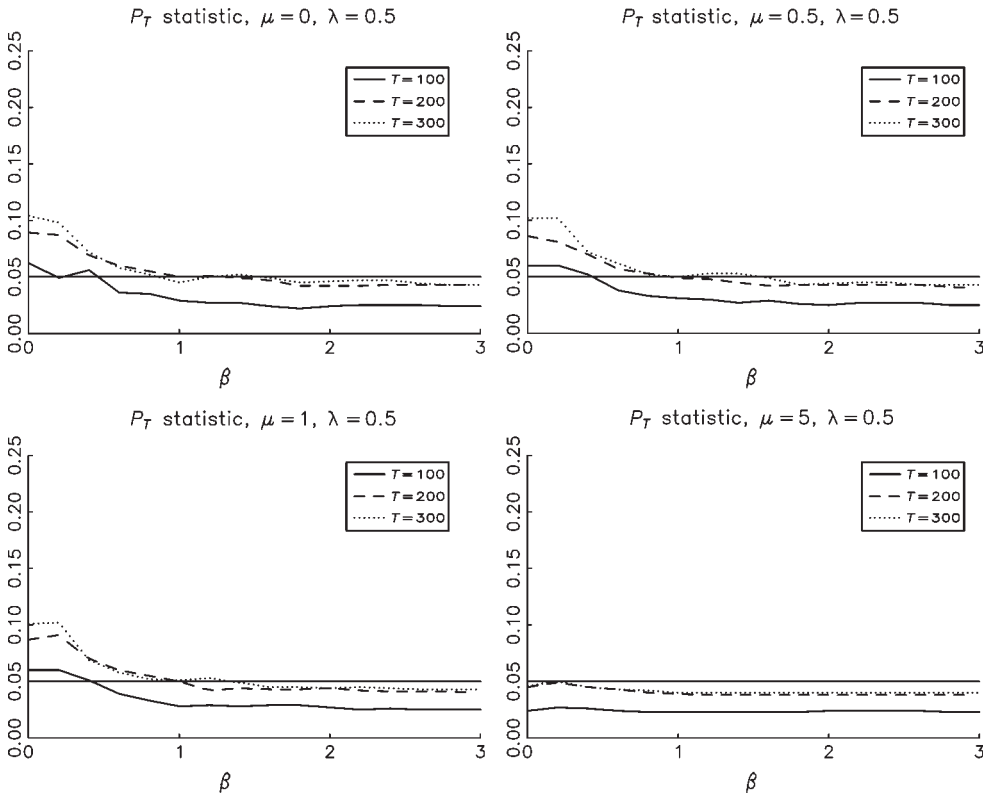


FIGURE 2. Empirical size of the $P_T^{GLS}(\hat{\lambda})$ test, $\lambda^0 = 0.5$.

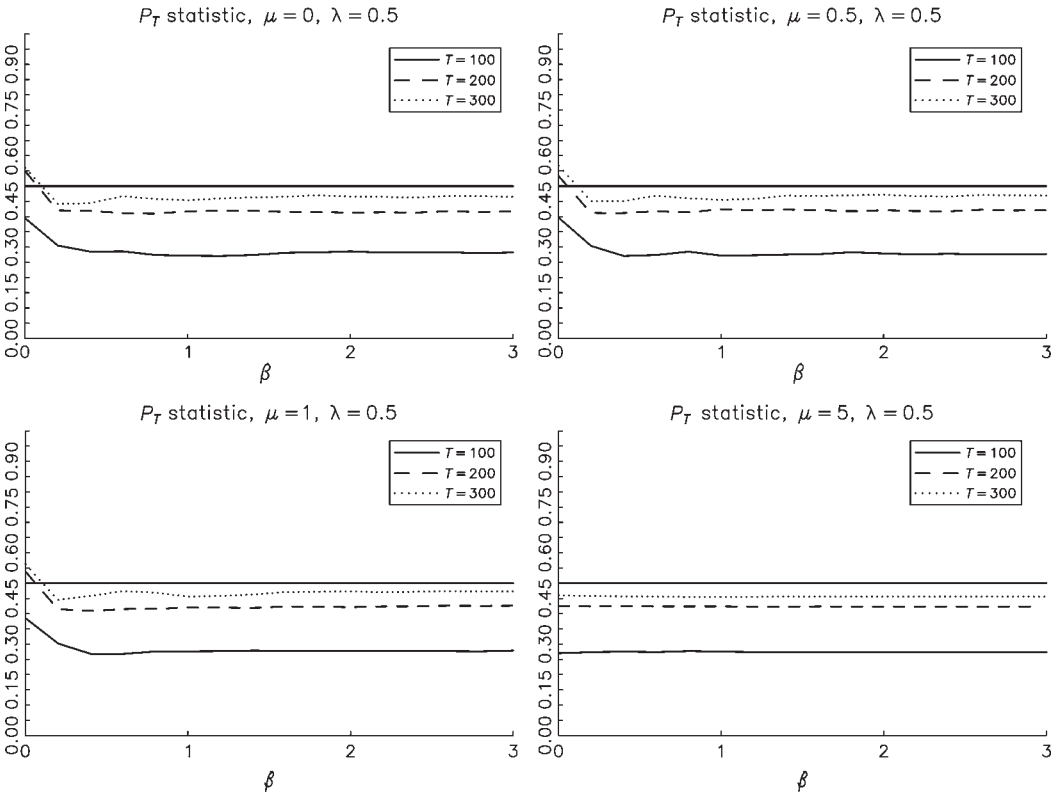


FIGURE 3. Empirical power of the $P_T^{GLS}(\hat{\lambda})$ test, $\lambda^0 = 0.5$.

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

and more so the larger the sample. This is in accordance with our theoretical results. When $|\beta_b|$ and $|\mu_b|$ are small, the tests show liberal size distortions and more so as the sample size increases. This is due to the fact that our theoretical results so far assumed $\beta_b \neq 0$ so that a break is present. The limit null distributions when no break occurs are indeed different because the estimate of the break fraction has a nondegenerate limit distribution on the interval [0, 1] instead of converging to either zero or one (see Nunes, Kuan, and Newbold, 1995; Bai, 1998). What is of special interest is the fact that when $|\mu_b|$ is large, the exact sizes are close to the nominal 5% level irrespective of the value of β_b. This suggests that a correction will be needed to account for cases with both $|\beta_b|$ and $|\mu_b|$ small. With respect to power, the following features emerge. The power quickly approaches the 50% limit value suggested by the local asymptotic power analysis unless T is small. This is so even if $|\beta_b|$ is not very large. For very small values of $|\beta_b|$, the power is above 50% but only because of the presence of important size distortions.

7. EXTENSION TO THE CASE WHERE A BREAK NEED NOT OCCUR

Up to this point, we have assumed that a break occurs under both the null and alternative hypotheses. When no break occurs, the asymptotic results described in the previous sections do not hold because, under the null hypothesis of a unit root, the estimates of the break fractions have a nondegenerate limit distribution on the interval [0, 1] instead of converging to either zero or one. If there is no break in the trend, then the proper unit root test procedure is to simply apply a standard DF type test with no break dummies. Hence, what is needed is a pretest to assess whether a break is present or not. This pretest for a break should have the correct size if the noise is integrated but should also be powerful whether the noise is stationary or has a unit root, to ensure a specification that allows a unit root test procedure with good power. This testing problem has recently been addressed by Vogelsang (2001) and Perron and Yabu (2009). Because the procedure of Perron and Yabu (2009) has better size and power, we shall use it as the pretest. It is based on a quasi-GLS approach using an autoregression for the noise component, with a truncation to one when the sum of the autoregressive coefficients is in some neighborhood of 1, along with a bias correction. For given break dates, one constructs the F-test for the null hypothesis of no structural change in the deterministic components. The final statistic uses the *Exp* functional of Andrews and Ploberger (1994). The test has virtually the same asymptotic size with a stationary or integrated noise component. We label this test as $Exp\text{-}W_{FS}$ and define the alternative estimate of the break fraction, $\tilde{\lambda}$, as $\tilde{\lambda} = \hat{\lambda} \cdot 1(Exp\text{-}W_{FS} > cv)$ where cv is the critical value for a test with, say, nominal size $p\%$. When there is a break, $\tilde{\lambda}$ consistently estimates λ^0, as long as $\hat{\lambda}$ consistently estimates λ^0, given that $Exp\text{-}W_{FS}$ is a consistent test. When there is no break, $\tilde{\lambda}$ will yield a nonzero estimate $p\%$ of times even with infinitely many observations, if cv is fixed. Now, suppose that $Exp\text{-}W_{FS} = O_p(T^{\bar{\omega}})$, $\bar{\omega} > 0$ under the alternative of a break. Let

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

$cv = cT^{\bar{\omega}-\varepsilon}$, $0 < \varepsilon < \bar{\omega}$. Such an increasing sequence of critical values does not harm the consistency of the test, and with the size converging to zero as $T \to \infty$, it ensures the consistency of $\tilde{\lambda}$ for $\lambda^0 \in [0, 1]$. Hence, the modified test procedure is to use a DF type test when $\tilde{\lambda} = 0$ and the procedure described earlier when $\tilde{\lambda} \neq 0$.

We now present simulation evidence about the size and power properties of our testing procedures using the pretest for the presence of breaks. We start with the case of a single break and then consider the case with an unknown number of breaks up to two. This will allow us, among other things, to discuss issues related to the estimation of the number of breaks.

The Case with One Break. We performed simulations with the same data generating process and design as used in Section 6 but this time for the version of our statistics involving the pretest. We only report results for $\lambda^0 = 0.5$ and for the test $P_T^{GLS}(\hat{\lambda})$ (the properties of the other tests are similar). The results are presented in Figures 4 (for size) and 5 (for power). With respect to the size of the test, one can see that the size distortions for values of $|\beta_b|$ near zero disappear to a large extent. The test $P_T^{GLS}(\hat{\lambda})$ with the pretest for break in trend has an exact size close to the nominal level in all cases.

The power functions with the pretest for break approach the target 50% level quickly as T increases when $|\beta_b|$ is moderate to large. When $|\beta_b|$ is near zero, the power is actually higher than 50%. This is due to the fact that with $\beta_b = 0$ the test performed without allowing for a change in trend has a local power function that is actually higher than the test that does allow for breaks. So there is improvement in terms of power also. The only drawback is for values of $|\beta_b|$ neither moderate nor small. Here, the power dips down. This is due to the fact that the pretest for break is not powerful enough for this range of values of $|\beta_b|$ so that a standard test without breaks is applied. Yet, the break is large enough to affect the power of the test, which decreases for the reasons explained in Perron (1989).

Overall, the performance of the $P_T^{GLS}(\hat{\lambda})$ tests (and the others presented in the paper whose performance is similar) is quite satisfactory and offers an improvement in terms of both size and power over existing procedures. It avoids the large size distortions of the Zivot and Andrews (1992) procedure while still allowing for much improved power by taking advantage of the information about the presence or absence of a break whether it be under the null or the alternative hypothesis.

The Case with an Unknown Number of Breaks up to Two. We now consider the case with a maximum of up to two breaks. When dealing with the possibility of multiple breaks, one needs a method to estimate the number of breaks present. The application of the test of Perron and Yabu (2009) for two breaks is not appropriate because it would indicate a rejection when a single break is present; i.e., it does not allow us to estimate the number of breaks. Kejriwal and Perron (2008) have extended the Perron–Yabu test so that it is applicable in a specific to general sequential manner, thereby allowing the estimation of the number of breaks. The

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

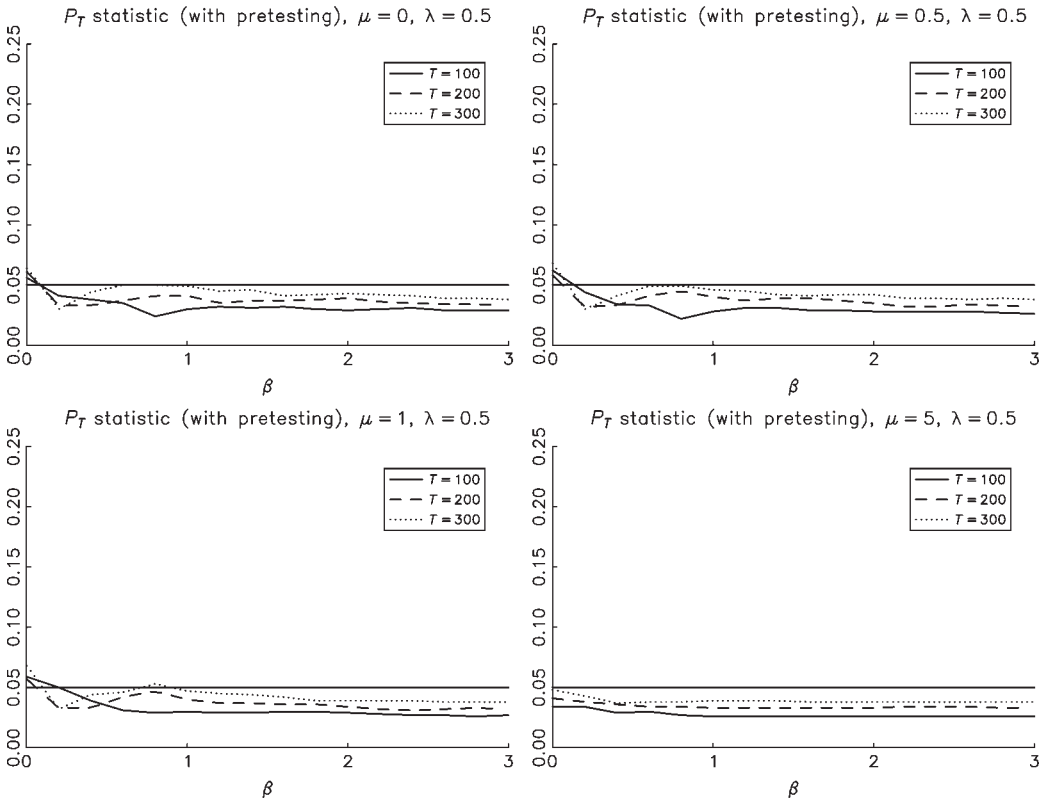


FIGURE 4. Empirical size of the $P_T^{GLS}(\hat{\lambda})$ test, $\lambda^0 = 0.5$, with pretesting.

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

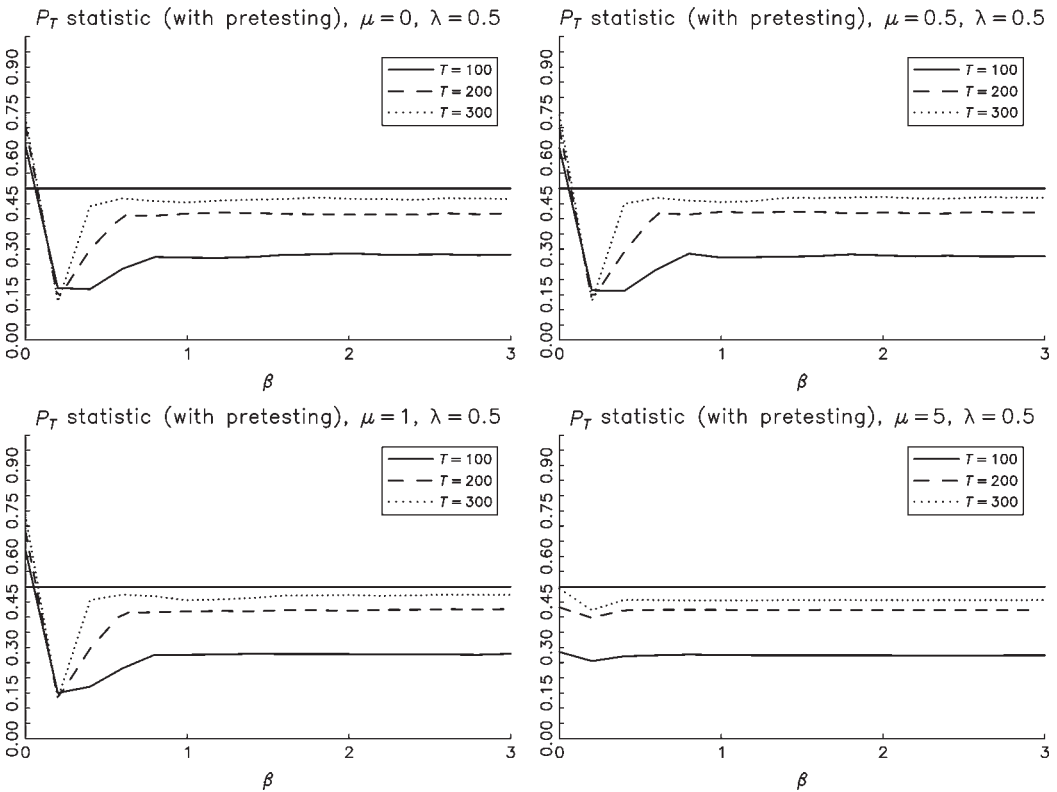


FIGURE 5. Empirical power of the $P_T^{GLS}(\hat{\lambda})$ test, $\lambda^0 = 0.5$, with pretesting.

procedure first applies the Perron–Yabu test for a single break. If the test does not reject, a unit root test with no break is applied. If the test does reject, the sample is split at the estimate of the break date obtained using the method described in Section 5.2. The Perron–Yabu test is then applied to each subsample (with a 15% trimming applied at each end), and the maximum value of the two tests is compared to the critical values derived in Kejriwal and Perron (2008). If no rejection occurs, a unit root test with a single break is used. Upon a rejection, a unit root test with two breaks is applied with the estimates again obtained using the method laid out in Section 5.2.

To illustrate the size and power properties of the unit root tests using this method to determine the number and location of the breaks, we consider simulations using Model I as the data generating process so that

$$y_t = \beta_1 DT_t^*(T_1^0) + \beta_2 DT_t^*(T_2^0) + u_t,$$

$$u_t = \alpha u_{t-1} + v_t,$$

with $v_t \sim i.i.d.\ N(0,1)$, $u_0 = 0$, and $T_i^0 = [T\lambda_i^0]$ for $i = 1, 2$. We set $\lambda_1^0 = 0.3$ and $\lambda_2^0 = 0.7$. We consider three sample sizes, $T = 200$, 300, and 500. The nominal size of the pretest and the unit root test is again 5%, and the results are based on 1,000 replications. For the size simulation, $\alpha = 1$; for the power simulation, $\alpha = \bar{\alpha} = 1 + \bar{c}/T$, with $\bar{c}$ the value of the noncentrality parameter corresponding to the number and location of the breaks detected. Hence, for two nonzero breaks, the power is 50% in large samples. The magnitudes of the break coefficients β_1 and β_2 take values $\{0, 1, 2, 3\}$. When $\beta_1 = \beta_2 = 0$, no break occurs; if either β_1 or β_2 is zero (but not both), we have a one break model, whereas if both β_1 and β_2 are nonzero, we have a two break model. In all cases, the method searches for up to a maximum of two breaks.

The results for size are presented in Figure 6, again only for the test $P_T^{GLS}(\hat{\lambda})$, the results for the other tests being qualitatively similar. One can see that for small values of the break coefficients, the exact size is close to the 5% nominal level. There is, however, a tendency for the unit root test to be slightly conservative when at least one of the breaks is large, though the size generally approaches 5% quickly unless both breaks are large.

The results for power are presented in Figure 7. When $\beta_1 = \beta_2 = 0$, the test has power well above 50%. This indicates that the method to estimate the number of breaks correctly selects no break, in which case power is larger than 50% in large samples. The power is also above 50% when there is a single break of large enough magnitude, as illustrated in the top left panel. When one of the breaks is small we have, as in the single break case, power below the target 50% level. The power, however, increases and approaches the target 50% level as the magnitudes of the breaks or the sample size increase. Hence, we can conclude from this limited set of simulations that the method is adequate in delivering tests with good size and power in the multiple break case with an unknown number of breaks.

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

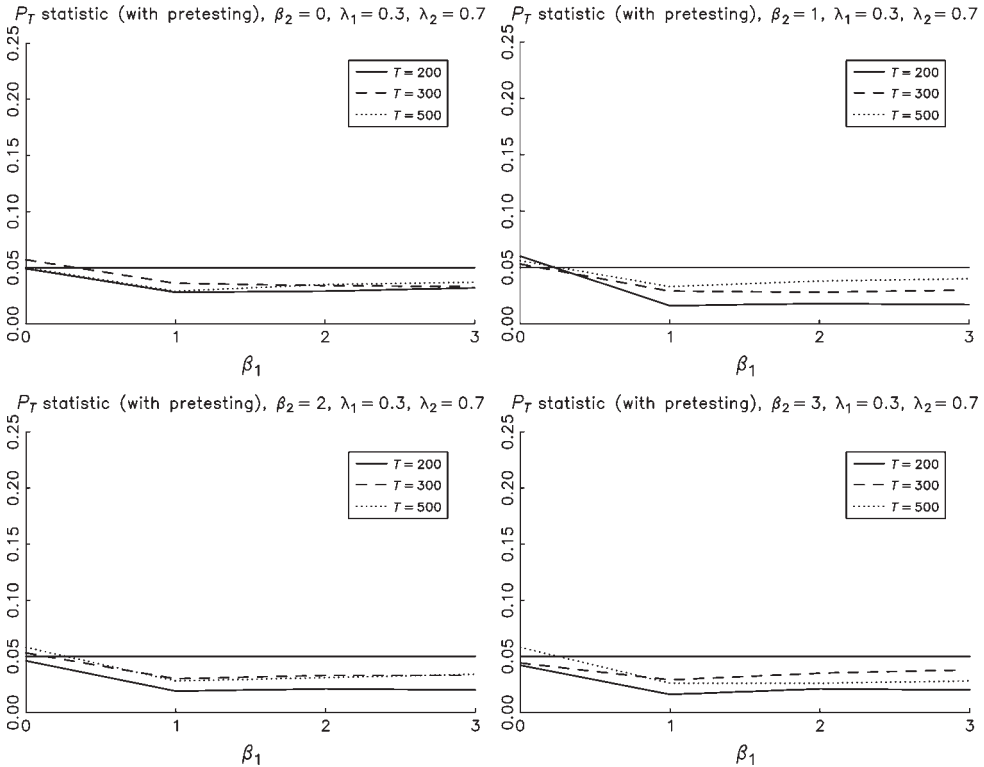


FIGURE 6. Empirical size of the $P_T^{GLS}(\hat{\lambda})$ test, $\lambda^0 = (0.3, 0.7)'$, with pretesting.

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

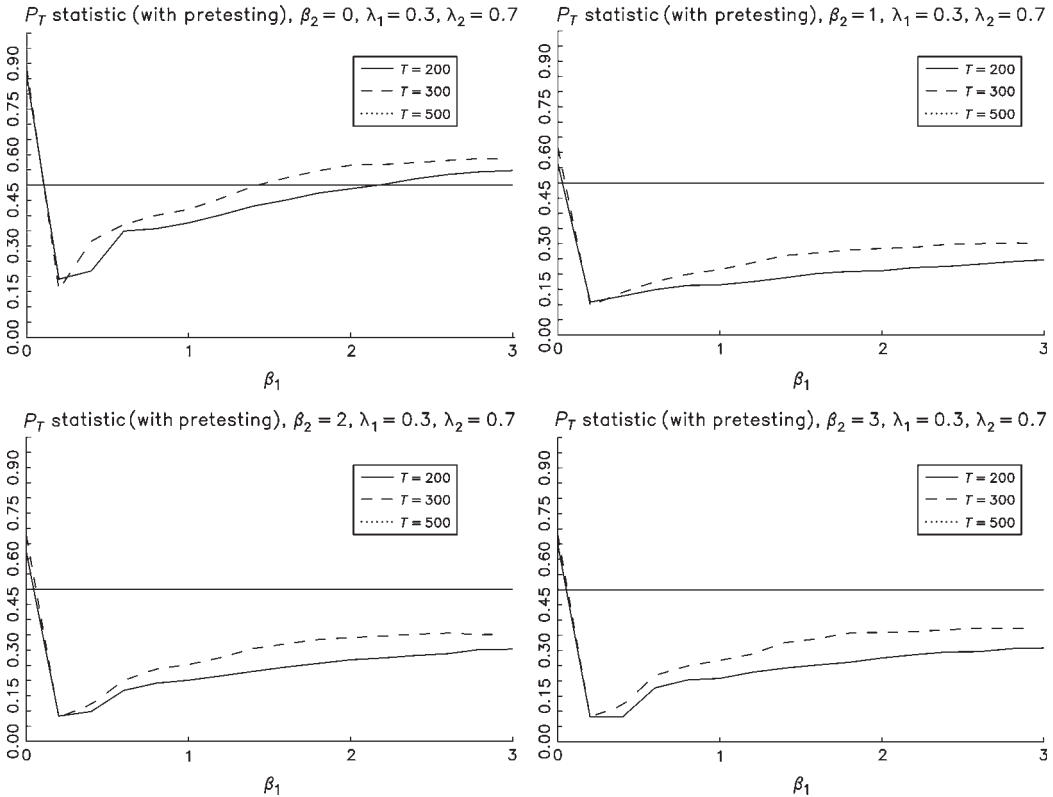


FIGURE 7. Empirical power of the $P_T^{GLS}(\hat{\lambda})$ test, $\lambda^0 = (0.3, 0.7)'$, with pretesting.

8. CONCLUSION

Following the work of Kim and Perron (2009), the procedures suggested in this paper solve many of the problems raised in the Introduction that plague most existing unit root tests designed for series with a breaking trend function. First, the tests allow for a break under both the null and the alternative hypotheses. This is desirable for several reasons. It imposes a symmetric treatment when allowing for a break, so that the tests do not reject when the noise is integrated but the trend is changing. Also, if a break is present, this information is exploited to improve power. Second, when a break is present, the limit distributions of the tests are the same as in the case of a known break date, thereby allowing increased power while maintaining the correct size. Our paper used the quasi-GLS procedure suggested by Elliott et al. (1996) to obtain tests that have a local asymptotic power close to the power envelope, except perhaps in a small neighborhood where the change in slope is small but nonzero. We have also extended the analysis to the multiple break case made possible by a modification of the dynamic programming algorithm as described in Bai and Perron (2003) and Perron and Qu (2006). Simulation experiments confirm that our procedures offer improvements over commonly used methods in small samples. Our tests should therefore be useful in empirical applications.

NOTES

1. Perron and Rodríguez (2003) also consider GLS-detrended type procedures allowing a single break. The approach is, however, quite different. Also, although their treatment yields "optimal" tests when a break is present, this is not the case if a break is absent.
2. Perron and Rodríguez (2003) also considered Models I and II but only with one break.
3. Note that this approach is different from the one adopted by Perron and Rodríguez (2003), who estimated the break date through a minimization of the SSR under the null and alternative hypotheses.

REFERENCES

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

Bai, J. (1998) A note on spurious break. *Econometric Theory* 14, 663–669.

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

Christiano, L.J. (1992) Searching for a break in GNP. *Journal of Business & Economic Statistics* 10, 237–250.

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

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

Harris, D., D.I. Harvey, S.J. Leybourne, & A.M.R. Taylor (2009) Testing for a unit root in the presence of a possible break in trend. *Econometric Theory* 25, 1545–1588 (this issue).

Harvey, D.I., S.J. Leybourne, & P. Newbold (2001) Innovational outlier unit root tests with an endogeneously determined break in level. *Oxford Bulletin of Economics and Statistics* 63, 559–575.

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

Hatanaka, M. & K. Yamada (1999) A unit root test in the presence of structural changes in I(1) and I(0) models. In R.F. Engle & H. White (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, pp. 256–282. Oxford University Press.

Hecq, A. & J.P. Urbain (1993) Misspecification tests, unit roots and level shifts. *Economics Letters* 43, 129–135.

Kejriwal, M. & P. Perron (2008) A Note on the Sequential Application of the Perron-Yabu Test for Shifts in Trend with an Integrated or Stationary Noise Component. Manuscript, Department of Economics, Boston University.

Kim, D. & P. Perron (2009) Unit root tests allowing for a break in the trend function under both the null and alternative hypotheses. *Journal of Econometrics* 148, 1–13.

Kim, T., S.J. Leybourne, & P. Newbold (2000) Spurious rejections by Perron tests in the presence of a break. *Oxford Bulletin of Economics and Statistics* 62, 433–444.

Lee, J. & M.C. Strazicich (2001) Break point estimation and spurious rejections with endogenous unit root tests. *Oxford Bulletin of Economics and Statistics* 63, 535–558.

Leybourne, S.J., T.C. Mills, & P. Newbold (1998) Spurious rejections by Dickey-Fuller tests in the presence of a break under the null. *Journal of Econometrics* 87, 191–203.

Leybourne, S.J. & P. Newbold (2000) Behavior of the standard and symmetric Dickey-Fuller-type tests when there is a break under the null hypothesis. *Econometrics Journal* 3, 1–15.

Montañés, A. (1997) Level shifts, unit roots and misspecification of the breaking date. *Economics Letters* 54, 7–13.

Montañés, A. & I. Olloqui (1999) Misspecification of the breaking date in segmented trend variables: Effect on the unit root tests. *Economics Letters* 65, 301–307.

Ng, S. & P. Perron (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.

Nunes, L.C., C.-M. Kuan, & P. Newbold (1995) Spurious break. *Econometric Theory* 11, 736–749.

Nunes, L.C., P. Newbold, & C.-M. Kuan (1997) Testing for unit root with breaks: Evidence on the Great Crash and the unit root hypothesis reconsidered. *Oxford Bulletin of Economics and Statistics* 59, 435–448.

Perron, P. (1989) The Great Crash, the oil price shock and the unit root hypothesis. *Econometrica* 57, 1361–1401.

Perron, P. (1990) Testing for a unit root in a time series with a changing mean. *Journal of Business & Economic Statistics* 8, 153–162.

Perron, P. (1997) Further evidence on breaking trend functions in macroeconomic variables. *Journal of Econometrics* 80, 355–385.

Perron, P. & S. Ng (1998) An autoregressive spectral density estimator at frequency zero for nonstationarity tests. *Econometric Theory* 14, 560–603.

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

Perron, P. & Z. Qu (2007) A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. *Economics Letters* 94, 12–19.

Perron, P. & G.H. Rodríguez (2003) GLS detrending, efficient unit root tests and structural change. *Journal of Econometrics* 115, 1–27.

Perron, P. & T.J. Vogelsang (1992a) Nonstationarity and level shifts with an application to purchasing power parity. *Journal of Business & Economic Statistics* 10, 301–320.

Perron, P. & T.J. Vogelsang (1992b) Testing for a unit root in a time series with a changing mean: Correction and extensions. *Journal of Business & Economic Statistics* 10, 467–470.

Perron, P. & T. Yabu (2009) Testing for shifts in trend with an integrated or stationary noise component. *Journal of Business and Economic Statistics* 27, 369–396.

Perron, P. & X. Zhu (2005) Structural breaks with deterministic and stochastic trends. *Journal of Econometrics* 129, 65–119.

Stock, J.H. (1999) A class of tests for integration and cointegration. In *Cointegration, Causality, and Forecasting: A Festschrift for Clive W.J. Granger*, pp. 135–167. Oxford University Press.

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

Vogelsang, T.J. (2001) Testing for a Shift in Trend when Serial Correlation is of Unknown Form. Manuscript, Department of Economics, Cornell University.
Vogelsang, T.J. & P. Perron (1998) Additional tests for a unit root allowing for a break in the trend function at an unknown time. *International Economic Review* 39, 1073–1100.
Zivot, E. & D.W.K. Andrews (1992) Further evidence on the Great Crash, the oil price shock and the unit root hypothesis. *Journal of Business & Economic Statistics* 10, 251–270.

APPENDIX

We start by presenting some results that will be used throughout. We have, by definition

$$DT^{*\bar{\alpha}}(T_i)=\begin{cases}0, & t\le T_i\\ 1-\bar{c}(t-T_i-1)/T, & t\ge T_i+1\end{cases}\quad\text{and}\quad DU^{\bar{\alpha}}(T_i)=\begin{cases}0, & t\le T_i\\ 1, & t=T_i+1\\ -\bar{c}/T, & t\ge T_i+2.\end{cases}$$

Also, for $T_j > T_i$,

$$DT^{*\bar{\alpha}}(T_j)-DT^{*\bar{\alpha}}(T_i)=\begin{cases}-1+\bar{c}(t-T_i-1)/T, & T_i+1\le t\le T_j\\ \bar{c}(\lambda_j-\lambda_i), & T_j+1\le t\end{cases}$$

and zero otherwise,

$$DU^{\bar{\alpha}}(T_j)-DU^{\bar{\alpha}}(T_i)=\begin{cases}-1, & t=T_i+1\\ \bar{c}/T, & T_i+2\le t\le T_j\\ 1+\bar{c}/T, & t=T_j+1\end{cases}$$

and zero otherwise. We can then show that, with $T_i \le T_j$,

$$\begin{aligned}T^{-1}DT^{*\bar{\alpha}}(T_i)'DT^{*\bar{\alpha}}(T_j)&=(1/6)(1-\lambda_j)\\ &\quad\times\left[6\bar{c}(\lambda_i-1)+\bar{c}^2\left\{\lambda_j(3\lambda_i-1)-3\lambda_i-\lambda_j^2+2\right\}+6\right]+o(1)\\ &\equiv a(\lambda_i,\lambda_j)+o(1),\end{aligned}\tag{A.1}$$

$$DU^{\bar{\alpha}}(T_i)'DU^{\bar{\alpha}}(T_j)=1(|T_i-T_j|=0)+o(1),\tag{A.2}$$

where $1(\cdot)$ denotes the indicator function. Also,

$$\begin{cases}DU^{\bar{\alpha}}(T_i)'DT^{*\bar{\alpha}}(T_j)=1(|T_i-T_j|=0)-\bar{c}(1-\lambda_j)+\bar{c}^2(1-\lambda_j)^2/2+o(1),\\ DU^{\bar{\alpha}}(T_j)'DT^{*\bar{\alpha}}(T_i)=1-\bar{c}(\lambda_j-\lambda_i)-\bar{c}(1-\lambda_j)+\bar{c}^2(1+\lambda_j-2\lambda_i)(1-\lambda_j)/2+o(1).\end{cases}\tag{A.3}$$

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

From (A.1)–(A.3), we have, with $T_i \le T_j$,

$$\begin{cases} DU^{\bar\alpha}(T_j)'[DU^{\bar\alpha}(T_j) - DU^{\bar\alpha}(T_i)] = \ 1(|T_i - T_j| \neq 0) + o(1), \\ DU^{\bar\alpha}(T_i)'[DU^{\bar\alpha}(T_j) - DU^{\bar\alpha}(T_i)] = -1(|T_i - T_j| \neq 0) + o(1), \end{cases} \tag{A.4}$$

$$\begin{cases} DU^{\bar\alpha}(T_j)'[DT^{*\bar\alpha}(T_j) - DT^{*\bar\alpha}(T_i)] = \bar c(\lambda_j - \lambda_i) - \bar c^2(\lambda_j - \lambda_i)(1-\lambda_j) + o(1), \\ DU^{\bar\alpha}(T_i)'[DT^{*\bar\alpha}(T_j) - DT^{*\bar\alpha}(T_i)] = -1(|T_i - T_j| \neq 0) + \bar c(\lambda_j - \lambda_i) \\ \qquad\qquad - \bar c^2(\lambda_j - \lambda_i)^2/2 - \bar c^2(\lambda_j - \lambda_i)(1-\lambda_j) + o(1), \end{cases} \tag{A.5}$$

$$\begin{cases} DT^{*\bar\alpha}(T_j)'[DU^{\bar\alpha}(T_j) - DU^{\bar\alpha}(T_i)] = 1(|T_i - T_j| \neq 0) + o(1), \\ DT^{*\bar\alpha}(T_i)'[DU^{\bar\alpha}(T_j) - DU^{\bar\alpha}(T_i)] = -\bar c^2(\lambda_j - \lambda_i)^2/2 + o(1), \end{cases} \tag{A.6}$$

$$\begin{cases} T^{-1}DT^{*\bar\alpha}(T_j)'[DT^{*\bar\alpha}(T_j) - DT^{*\bar\alpha}(T_i)] = \bar c(\lambda_j - \lambda_i)[(1-\lambda_j) - \bar c(1-\lambda_j)^2/2] + o(1), \\ T^{-1}DT^{*\bar\alpha}(T_i)'[DT^{*\bar\alpha}(T_j) - DT^{*\bar\alpha}(T_i)] = -(\lambda_j - \lambda_i) + \bar c(\lambda_j - \lambda_i)(1-\lambda_i) \\ \qquad\qquad - \bar c^2(\lambda_j - \lambda_i)\{3(1-\lambda_i)^2 - (\lambda_j - \lambda_i)^2\}/6 \\ \qquad\qquad + o(1). \end{cases} \tag{A.7}$$

From (A.4)–(A.7), we have (with $T_j \ge T_i$)

$$||DU^{\bar\alpha}(T_j) - DU^{\bar\alpha}(T_i)||^2 = 2 \cdot 1(|T_j - T_i| \neq 0) + o(1), \tag{A.8}$$

$$\begin{aligned} T^{-1}||DT^{*\bar\alpha}(T_j) - DT^{*\bar\alpha}(T_i)||^2 &= (\lambda_j - \lambda_i) - \bar c(\lambda_j - \lambda_i)^2 \\ &\quad + \bar c^2(\lambda_j - \lambda_i)^3/3 \\ &\quad + \bar c^2(\lambda_j - \lambda_i)^2(1-\lambda_j) + o(1), \end{aligned} \tag{A.9}$$

$$\begin{aligned} (DU^{\bar\alpha}(T_j) - DU^{\bar\alpha}(T_i))'\left(DT^{*\bar\alpha}(T_j) - DT^{*\bar\alpha}(T_i)\right) &= 1(|T_i - T_j| \neq 0) \\ &\quad + \bar c(\lambda_j - \lambda_i)^2/2 + o(1). \end{aligned} \tag{A.10}$$

Noting that $u_1^{\bar\alpha} = v_1$ and $u_t^{\bar\alpha} = v_t + T^{-1}(c - \bar c)u_{t-1}$ $(t = 2, \ldots, T)$, we deduce that

$$\begin{aligned} DU^{\bar\alpha}(T_i)'u^{\bar\alpha} &= v_{T_i+1} + T^{-1}(c - \bar c)u_{T_i} + \bar c T^{-1}\sum_{t=T_i+2}^{T}(v_t + T^{-1}(c - \bar c)u_{t-1}) \\ &= v_{T_i+1} + o_p(1), \end{aligned} \tag{A.11}$$

$$\begin{aligned} T^{-1/2}DT^{*\bar\alpha}(T_i)'u^{\bar\alpha} &= T^{-1/2}\sum_{T_i+1}^{T}(v_t + T^{-1}(c - \bar c)u_{t-1})(1 - \bar c(t - T_i - 1)/T) \\ &\Rightarrow \sigma V(\lambda_i), \end{aligned} \tag{A.12}$$

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

where

$$V(\lambda_i) = (1+\bar{c}\lambda_i)\left[W(1) - W(\lambda_i) + (c-\bar{c})\int_{\lambda_i}^{1} W_c(r)\,dr\right] - \bar{c}\int_{\lambda_i}^{1} r\,dW(r) - (c-\bar{c})\bar{c}\int_{\lambda_i}^{1} r\,W_c(r)\,dr. \tag{A.13}$$

Proof of Theorem 1. Define the quadratic form $M_T\left(c,\bar{c},\lambda^0\right) = \left(u^{\bar{\alpha}\prime}z^{\bar{\alpha}}\right)\left(z^{\bar{\alpha}\prime}z^{\bar{\alpha}}\right)^{-1} \times \left(z^{\bar{\alpha}\prime}u^{\bar{\alpha}}\right)$, where $z^{\bar{\alpha}}$ is the vector of the quasi-differences of $z_t(\lambda^0)$ defined in (3). From Elliott et al. (1996) and Perron and Rodríguez (2003), we have

$$s^2(\lambda^0)P_T^{GLS}(c,\bar{c},\lambda^0) = M_T(c,0,\lambda^0) - M_T(c,\bar{c},\lambda^0) - 2\bar{c}T^{-1}\sum_{t=2}^{T} u_{t-1}v_t + (\bar{c}^2 - 2\bar{c}c)T^{-2}\sum_{t=2}^{T} u_{t-1}^2 - \bar{c}T^{-1}u^{1\prime}u^1 + o_p(1). \tag{A.14}$$

From the invariance principle and the continuous mapping theorem (CMT), $T^{-1}\sum_{t=2}^{T} u_{t-1}v_t \Rightarrow \sigma^2[\int_0^1 W_c(r)\,dW(r) + \gamma\,]$ and $T^{-2}\sum_{t=2}^{T} u_{t-1}^2 \Rightarrow \sigma^2\int_0^1 W_c(r)^2\,dr$ where $\gamma = (\sigma^2 - \sigma_v^2)/(2\sigma^2)$ and $\sigma_v^2 = \mathrm{E}v_t^2$. By the law of large numbers, $p\lim T^{-1}u^{1\prime}u^1 = \sigma_v^2$. Note that the expression in (A.14) is similar to that in Elliott et al. (1996) and Perron and Rodríguez (2003), the only difference coming from the definition of the quadratic forms $M_T(c,0,\lambda^0)$ and $M_T(c,\bar{c},\lambda^0)$. Let $z_t' = (z_{t,1}', z_{t,2}')$ be a rearranged version of $z_t(\lambda^0)$ so that $z_{t,1}$ collects the $m+1$ regressors that correspond to the constant and the impulse dummy variables and $z_{t,2}$ collects the $m+1$ trending regressors. Also, let $z_t^{\bar{\alpha}\prime} = (z_{t,1}^{\bar{\alpha}\prime}, z_{t,2}^{\bar{\alpha}\prime})$ be the quasi-differenced z_t. Then, the scaled matrix $M_T(c,\bar{c},\lambda^0)$ can be expressed as

$$M_T(c,\bar{c},\lambda^0) = \left(u^{\bar{\alpha}\prime}z^{\bar{\alpha}}D_T\right)\left(D_T z^{\bar{\alpha}\prime}z^{\bar{\alpha}}D_T\right)^{-1}\left(D_T z^{\bar{\alpha}\prime}u^{\bar{\alpha}}\right),$$

where $D_T = \mathrm{diag}\{D_{1,T}, D_{2,T}\} = \mathrm{diag}\left(1,\ldots,1,T^{-1/2},\ldots,T^{-1/2}\right)$. From (A.11) and (A.12),

$$D_T z^{\bar{\alpha}\prime}u^{\bar{\alpha}} = \sigma\left(v_1/\sigma,\ldots,v_{T_j+1}/\sigma,\ldots,v_{T_m+1}/\sigma, V(\lambda_0^0),\ldots,V(\lambda_j^0),\ldots,V(\lambda_m^0)\right)' + o_p(1). \tag{A.15}$$

Using (A.1)–(A.3), the limit of $D_T z^{\bar{\alpha}\prime}z^{\bar{\alpha}}D_T$ is given by the following block diagonal matrix:

$$D_T z^{\bar{\alpha}\prime}z^{\bar{\alpha}}D_T \to \begin{bmatrix} I_{m+1} & 0 \\ 0 & A(\lambda^0) \end{bmatrix}, \tag{A.16}$$

where I_{m+1} is the identity matrix of order $m+1$ and $A(\lambda^0)$ is a symmetric matrix defined by (8) with $a\left(\lambda_i^0,\lambda_j^0\right)$ as given in (A.1). Therefore, we have

$$M_T\left(c,\bar{c},\lambda^0\right) = v_1^2 + \sum_{j=1}^{m} v_{T_j+1}^2 + \sigma^2 M\left(c,\bar{c},\lambda^0\right) + o_p(1),$$

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

where $M(c,\bar{c},\lambda^0) = \bar{V}(\lambda^0)'A(\lambda^0)^{-1}\bar{V}(\lambda^0)$, with $\bar{V}(\lambda^0) = \left(V(\lambda_0^0),\ldots,V(\lambda_m^0)\right)'$ and $V(\lambda_j^0)$ as defined in (A.13). This weak convergence result also holds when $\bar{c}=0$, and, thus,

$$M_T(c,0,\lambda^0) = v_1^2 + \sum_{j=1}^{m} v_{T_j+1}^2 + \sigma^2 M(c,0,\lambda^0) + o_p(1).$$

It follows that the limiting distribution is as stated in Theorem 1. ■

Proof of Theorem 2. Note that the scaled detrended variable $\tilde{y}_t$ is given by

$$T^{-1/2}\tilde{y}_t = T^{-1/2}u_t - T^{-1/2}z_t' D_T \left(D_T z^{\bar{\alpha}\prime} z^{\bar{\alpha}} D_T\right)^{-1} D_T z^{\bar{\alpha}\prime} u^{\bar{\alpha}}.$$

Using (A.15), (A.16), and the asymptotic block diagonality of $\left(D_T z^{\bar{\alpha}\prime} z^{\bar{\alpha}} D_T\right)^{-1}$, we obtain

$$T^{-1/2}\tilde{y}_t = T^{-1/2}u_t - T^{-1/2}z_{t,1}' \left(z_1^{\bar{\alpha}\prime} z_1^{\bar{\alpha}}\right)^{-1} z_1^{\bar{\alpha}\prime} u^{\bar{\alpha}}$$

$$- T^{-1/2}z_{t,2}' D_{2,T} \left(D_{2,T} z_2^{\bar{\alpha}\prime} z_2^{\bar{\alpha}} D_{2,T}\right)^{-1} D_{2,T} z_2^{\bar{\alpha}\prime} u^{\bar{\alpha}} + o_p(1).$$

Note that $T^{-1/2}z_{[\mathrm{Tr}],1}' \left(z_1^{\bar{\alpha}\prime} z_1^{\bar{\alpha}}\right)^{-1} z_1^{\bar{\alpha}\prime} u^{\bar{\alpha}} \xrightarrow{p} 0$, provided that $T^{-1/2}v_{T_{b,j}+1}/\sigma \xrightarrow{p} 0, \forall j = 0,\ldots,m$. Also, $T^{-1/2}D_{2,T}z_{[\mathrm{Tr}],2} \to z_2(r)$ uniformly in $r \in [0,1]$, where $z_2(r) = (r, (r-\lambda_1^0)1(r>\lambda_1^0),\ldots,(r-\lambda_m^0)1(r>\lambda_m^0))$. Therefore,

$$T^{-1/2}\tilde{y}_{[\mathrm{Tr}]} \Rightarrow \sigma[W_c(r) - z_2(r)A(\lambda^0)^{-1}\bar{V}(\lambda^0)] = \sigma V_{c,\bar{c}}(r,\lambda^0),$$

with $A(\lambda^0)$ and $\bar{V}(\lambda^0)$ as defined before. Using this weak convergence result, the CMT and the fact that $s(\lambda^0)^2 \xrightarrow{p} \sigma^2$, the results stated in the theorem for Models I, II, and IIb follow. ■

Proof of Proposition 1. Let λ^0 denote the true break fraction and λ a generic one. Except when indicated, $\bar{\alpha}$ is computed using λ. Note that

$$S(\bar{\alpha},\lambda) = (z^{\bar{\alpha}}(\lambda^0)\psi + u^{\bar{\alpha}})' M_\lambda^{\bar{\alpha}} (z^{\bar{\alpha}}(\lambda^0)\psi + u^{\bar{\alpha}}) = (d^{\bar{\alpha}}(\lambda)\psi + u^{\bar{\alpha}})' M_\lambda^{\bar{\alpha}} (d^{\bar{\alpha}}(\lambda)\psi + u^{\bar{\alpha}})$$

$$= \psi' d^{\bar{\alpha}\prime}(\lambda) M_\lambda^{\bar{\alpha}} d^{\bar{\alpha}}(\lambda)\psi - 2u^{\bar{\alpha}\prime} M_\lambda^{\bar{\alpha}} d^{\bar{\alpha}}(\lambda)\psi + u^{\bar{\alpha}\prime} M_\lambda^{\bar{\alpha}} u^{\bar{\alpha}} \equiv Q^{\bar{\alpha}}(\lambda) - 2G^{\bar{\alpha}}(\lambda)$$
$$+ u^{\bar{\alpha}\prime} M_\lambda^{\bar{\alpha}} u^{\bar{\alpha}}, \qquad \textbf{(A.17)}$$

where $d^{\bar{\alpha}}(\lambda) = z^{\bar{\alpha}}(\lambda) - z^{\bar{\alpha}}(\lambda^0)$, $M_\lambda^{\bar{\alpha}} = I - P_\lambda^{\bar{\alpha}}$, and $P_\lambda^{\bar{\alpha}} = z^{\bar{\alpha}}(\lambda) D_T (D_T z^{\bar{\alpha}\prime}(\lambda) z^{\bar{\alpha}}(\lambda) D_T)^{-1} D_T z^{\bar{\alpha}\prime}(\lambda)$. When the break fraction is correctly specified, $\lambda = \lambda^0$ and $S(\bar{\alpha}_0,\lambda^0) = u^{\bar{\alpha}_0\prime} M_{\lambda^0}^{\bar{\alpha}_0} u^{\bar{\alpha}_0}$. Note that the preceding expressions are functions of $u^{\bar{\alpha}_0}$ and $z^{\bar{\alpha}_0}(\lambda^0)$, which are quasi-differenced with $\bar{\alpha}_0$, i.e., the parameter $\bar{\alpha}$ computed using the true break fraction λ^0. The difference between these two sums of squared residuals is given by

$$S(\bar{\alpha},\lambda) - S(\bar{\alpha}_0,\lambda^0) = Q^{\bar{\alpha}}(\lambda) - 2G^{\bar{\alpha}}(\lambda) + u^{\bar{\alpha}\prime} M_\lambda^{\bar{\alpha}} u^{\bar{\alpha}} - u^{\bar{\alpha}_0\prime} M_{\lambda^0}^{\bar{\alpha}_0} u^{\bar{\alpha}_0}.$$

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

Consider first

$$Q^{\bar{\alpha}}(\lambda) = \psi' d^{\bar{\alpha}\prime}(\lambda) d^{\bar{\alpha}}(\lambda)\psi - \psi' d^{\bar{\alpha}\prime}(\lambda) P_{\lambda}^{\bar{\alpha}} d^{\bar{\alpha}}(\lambda)\psi.$$

For simplicity, assume that there is only one break. In Models II and IIb,

$$\psi' d^{\bar{\alpha}\prime}(\lambda) d^{\bar{\alpha}}(\lambda)\psi = \mu_1^2 ||DU^{\bar{\alpha}}(T_1) - DU^{\bar{\alpha}}(T_1^0)||^2 + \beta_1^2 ||DT^{*\bar{\alpha}}(T_1) - DT^{*\bar{\alpha}}(T_1^0)||^2$$
$$+ 2\mu_1\beta_1 [DU^{\bar{\alpha}}(T_1) - DU^{\bar{\alpha}}(T_1^0)]' [DT^{*\bar{\alpha}}(T_1) - DT^{*\bar{\alpha}}(T_1^0)].$$

From (A.8)–(A.10), we have, in Model II (if $\lambda > \lambda^0$),

$$T^{-1}\psi' d^{\bar{\alpha}\prime}(\lambda) d^{\bar{\alpha}}(\lambda)\psi = \beta_1^2 \left[(\lambda - \lambda^0) - \bar{c}(\lambda - \lambda^0)^2 + \bar{c}^2 (\lambda - \lambda^0)^3/3 \right.$$
$$\left. + \bar{c}^2 (\lambda - \lambda^0)^2 (1 - \lambda) \right] + o(1)$$

and, in Model IIb,

$$T^{-1-2\eta}\psi' d^{\bar{\alpha}\prime}(\lambda) d^{\bar{\alpha}}(\lambda)\psi = 2\kappa_1^2 \cdot 1(|T_1 - T_1^0| \neq 0) + o(1).$$

Also,

$$d^{\bar{\alpha}\prime}(\lambda) P_{\lambda}^{\bar{\alpha}} d^{\bar{\alpha}}(\lambda) = d^{\bar{\alpha}\prime}(\lambda) z^{\bar{\alpha}}(\lambda) D_T \left(D_T z^{\bar{\alpha}\prime}(\lambda) z^{\bar{\alpha}}(\lambda) D_T \right)^{-1} D_T z^{\bar{\alpha}\prime}(\lambda) d^{\bar{\alpha}}(\lambda).$$

We have shown that $\left(D_T z^{\bar{\alpha}\prime}(\lambda) z^{\bar{\alpha}}(\lambda) D_T \right)^{-1} = O(1)$ with the limit being a block diagonal matrix. It also follows from (A.2)–(A.7) that

$$D_T z^{\bar{\alpha}\prime}(\lambda) d^{\bar{\alpha}}(\lambda)\psi = [\iota^{\bar{\alpha}}, DU^{\bar{\alpha}}(T_1), T^{-1/2}\tau^{\bar{\alpha}}, T^{-1/2} DT^{*\bar{\alpha}}(T_1)]' \qquad \textbf{(A.18)}$$
$$\times \left[DU^{\bar{\alpha}}(T_1) - DU^{\bar{\alpha}}\left(T_1^0\right), DT^{*\bar{\alpha}}(T_1) - DT^{*\bar{\alpha}}\left(T_1^0\right) \right] (\mu_1, \beta_1)'$$
$$= \begin{pmatrix} o(1) & O\left(T^{-1}\left|T_1 - T_1^0\right|\right) \\ 1 + o(1) & 1(T_1 < T_1^0) + O\left(T^{-1}\left|T_1 - T_1^0\right|\right) \\ o(1) & O\left(T^{-1/2}\left|T_1 - T_1^0\right|\right) \\ o(1) & O\left(T^{-1/2}\left|T_1 - T_1^0\right|\right) \end{pmatrix} \begin{pmatrix} \mu_1 \\ \beta_1 \end{pmatrix}.$$

Hence, we have in Model II

$$T^{-1}\psi' d^{\bar{\alpha}\prime}(\lambda) P_{\lambda}^{\bar{\alpha}} d^{\bar{\alpha}}(\lambda)\psi = \beta_1^2 \begin{pmatrix} a(0,\lambda) - a(0,\lambda^0) \\ a(\lambda,\lambda) - a(\lambda,\lambda^0) \end{pmatrix}' A(\lambda)^{-1} \begin{pmatrix} a(0,\lambda) - a(0,\lambda^0) \\ a(\lambda,\lambda) - a(\lambda,\lambda^0) \end{pmatrix} + o(1)$$
$$= \left|T_1 - T_1^0\right|^2 O(T^{-2})$$

with $a(\lambda_i, \lambda_j)$ as defined in (A.1) and, in Model IIb, $T^{-1-2\eta}\psi' d^{\bar{\alpha}\prime}(\lambda) P_{\lambda}^{\bar{\alpha}} d^{\bar{\alpha}}(\lambda)\psi = o(1)$. An entirely analogous argument applies to Models 0b and I. Therefore, collecting terms,

$$Q^{\bar{\alpha}}(\lambda) = \begin{cases} O(T^{1+2\eta}) \cdot \tilde{I}, & \text{in Models 0b and IIb,} \\ \left|T_1 - T_1^0\right| O(1), & \text{in Models I and II,} \end{cases} \qquad \textbf{(A.19)}$$

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

where $\tilde{I} = 1(|T_1 - T_1^0| \neq 0)$. Consider now

$$G^{\bar{\alpha}}(\lambda) = u^{\bar{\alpha}\prime} d^{\bar{\alpha}}(\lambda)\psi - u^{\bar{\alpha}\prime} P_{\lambda}^{\bar{\alpha}} d^{\bar{\alpha}}(\lambda)\psi.$$

As before, we first analyze the order of magnitude of $u^{\bar{\alpha}\prime} d^{\bar{\alpha}}(\lambda)\psi$. Note first that

$$\begin{aligned} u^{\bar{\alpha}\prime}\left(DU^{\bar{\alpha}}(T_1) - DU^{\bar{\alpha}}\left(T_1^0\right)\right) &= -u^{\bar{\alpha}}_{\lambda^0 T+1} + \bar{c}T^{-1}\sum_{t=\lambda^0 T+2}^{\lambda T} u_t^{\bar{\alpha}} + \left(1+\bar{c}T^{-1}\right)u^{\bar{\alpha}}_{\lambda T+1} \\ &= -\left(v_{\lambda^0 T+1} + T^{-1}(c-\bar{c})u_{\lambda^0 T}\right) + \bar{c}T^{-1} \\ &\quad \times \sum_{t=\lambda^0 T+2}^{\lambda T}\left(v_t + T^{-1}(c-\bar{c})u_{t-1}\right) \\ &\quad + \left(1+\bar{c}T^{-1}\right)\left(v_{\lambda T+1} + T^{-1}(c-\bar{c})u_{\lambda T}\right) \\ &= -v_{\lambda^0 T+1} + v_{\lambda T+1} + o_p(1) = O_p(1)\cdot\tilde{I} \end{aligned}$$

and

$$\begin{aligned} &u^{\bar{\alpha}\prime}\left(DT^{*\bar{\alpha}}(T_1) - DT^{*\bar{\alpha}}\left(T_1^0\right)\right) \\ &\quad = -u^{\bar{\alpha}}_{\lambda^0 T+1} - \sum_{t=\lambda^0 T+2}^{\lambda T+1}\left(-\bar{c}T^{-1}t + 1 + \lambda^0\bar{c} + \bar{c}T^{-1}\right)u_t^{\bar{\alpha}} + u^{\bar{\alpha}}_{\lambda T+1} \\ &\qquad + \sum_{t=\lambda T+2}^{T}\left(\left(-\bar{c}T^{-1}t + 1 + \lambda\bar{c} + \bar{c}T^{-1}\right) - \left(-\bar{c}T^{-1}t + 1 + \lambda^0\bar{c} + \bar{c}T^{-1}\right)\right)u_t^{\bar{\alpha}} \\ &\quad = u^{\bar{\alpha}}_{\lambda T+1} - u^{\bar{\alpha}}_{\lambda^0 T+1} - \sum_{t=\lambda^0 T+2}^{\lambda T+1}\left(-\bar{c}T^{-1}t + 1 + \lambda^0\bar{c}\right)u_t^{\bar{\alpha}} + \left(\lambda - \lambda^0\right)\bar{c} \\ &\qquad \times \sum_{t=\lambda T+2}^{T}\left(v_t + \frac{c-\bar{c}}{T}u_{t-1}\right) + o_p(1) \\ &\quad = |T_1 - T_1^0| O_p(T^{-1/2}). \end{aligned}$$

Then,

$$u^{\bar{\alpha}\prime} d^{\bar{\alpha}}(\lambda)\psi = \begin{cases} O_p(T^{1/2+\eta})\cdot\tilde{I}, & \text{in Models 0b and IIb,} \\ \left|T_1 - T_1^0\right| O_p(T^{-1/2}), & \text{in Models I and II.} \end{cases}$$

Now,

$$\begin{aligned} u^{\bar{\alpha}\prime} P_{\lambda}^{\bar{\alpha}} d^{\bar{\alpha}}(\lambda)\psi &= u^{\bar{\alpha}\prime} z^{\bar{\alpha}}(\lambda) D_T\left(D_T z^{\bar{\alpha}\prime}(\lambda) z^{\bar{\alpha}}(\lambda) D_T\right)^{-1} D_T z^{\bar{\alpha}\prime}(\lambda) d^{\bar{\alpha}}(\lambda)\psi \\ &= O_p(1) D_T z^{\bar{\alpha}\prime}(\lambda) d^{\bar{\alpha}}(\lambda)\psi \\ &= \begin{cases} o_p(T^{1/2+\eta})\cdot\tilde{I}, & \text{in Models 0b and IIb,} \\ \left|T_1 - T_1^0\right| O_p(T^{-1/2}), & \text{in Models I and II.} \end{cases} \end{aligned}$$

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

Therefore, we have

$$G^{\bar{\alpha}}(\lambda) = \begin{cases} O_p(T^{1/2+\eta}) \cdot \tilde{I}, & \text{in Models 0b and IIb,} \\ \left| T_1 - T_1^0 \right| O_p(T^{-1/2}), & \text{in Models I and II.} \end{cases} \tag{A.20}$$

For the last term, it follows from previous results that

$$u^{\bar{\alpha}\prime} M_{\lambda}^{\bar{\alpha}} u^{\bar{\alpha}} - u^{\bar{\alpha}_0\prime} M_{\lambda^0}^{\bar{\alpha}_0} u^{\bar{\alpha}_0} \leq O_p(1) \cdot \tilde{I}.$$

Therefore, collecting terms,

$$T^{-1}\left(S(\bar{\alpha}, \lambda) - S(\bar{\alpha}_0, \lambda^0)\right) = T^{-1} Q^{\bar{\alpha}}(\lambda) + o_p(1).$$

Furthermore, note that for a given estimate of the break fraction vector, $\hat{\lambda}$, the inequality $T^{-1}S(\bar{\alpha}, \hat{\lambda}) \leq T^{-1}S(\bar{\alpha}_0, \lambda^0)$ is always satisfied. Now suppose that $\hat{\lambda} \nrightarrow_p \lambda^0$. Then, according to this inequality, we will need that, for large T, $T^{-1}Q^{\bar{\alpha}}(\hat{\lambda}) \leq 0$, but we have shown that $Q^{\bar{\alpha}}(\lambda) > 0$ when $\lambda \neq \lambda^0$. Hence a contradiction, and the only way that the inequality is satisfied is when $\hat{\lambda} \rightarrow_p \lambda^0$. Thus, the minimization of $S(\bar{\alpha}, \lambda)$ over all possible values of the break fraction vector λ results in a consistent estimate of the break fractions. Although our derivations have considered only one structural break, the arguments are also valid for the multiple break case.

To establish the rate of convergence for Models I and II, we first define the sets $V_\epsilon = \left\{T_k : \left|T_k - T_1^0\right| < \epsilon T\right\}$ for $\epsilon \in (0, 1)$ and $V_\epsilon(C) = \left\{T_k : \left|T_k - T_1^0\right| < \epsilon T, \left|T_k - T_1^0\right| > C\right\}$ for $C > 0$, so that $V_\epsilon(C) \subset V_\epsilon$. Note that $S(\bar{\alpha}, \hat{\lambda}) \leq S(\bar{\alpha}_0, \lambda^0)$ with probability one and $\Pr(\hat{T}_1 \in V_\epsilon) \rightarrow 1$, as $T \rightarrow \infty$. Given the previous results, there is a constant $C > 0$ such that

$$\Pr\left(\min_{\lambda T \in V_\epsilon(C)} \frac{S(\bar{\alpha}, \lambda) - S(\bar{\alpha}_0, \lambda^0)}{|\lambda - \lambda^0| T} \leq 0\right) < \zeta$$

for some small $\zeta > 0$ because, when C is properly chosen,

$$\Pr\left(\min_{\lambda T \in V_\epsilon(C)} \frac{S(\bar{\alpha}, \lambda) - S(\bar{\alpha}_0, \lambda^0)}{|\lambda - \lambda^0| T} \leq 0\right)$$

$$= \Pr\left(\min_{\lambda T \in V_\epsilon(C)} \frac{Q^{\bar{\alpha}}(\lambda) - 2G^{\bar{\alpha}}(\lambda) + u^{\bar{\alpha}\prime} M_{\lambda}^{\bar{\alpha}} u^{\bar{\alpha}} - u^{\bar{\alpha}_0\prime} M_{\lambda^0}^{\bar{\alpha}_0} u^{\bar{\alpha}_0}}{|\lambda - \lambda^0| T} \leq 0\right) < \zeta$$

because

$$\frac{Q^{\bar{\alpha}}(\lambda) - 2G^{\bar{\alpha}}(\lambda) + u^{\bar{\alpha}\prime} M_{\lambda}^{\bar{\alpha}} u^{\bar{\alpha}} - u^{\bar{\alpha}_0\prime} M_{\lambda^0}^{\bar{\alpha}_0} u^{\bar{\alpha}_0}}{|\lambda - \lambda^0| T} = O(1) - 2O_p(T^{-1/2}) + \frac{O_p(1) \cdot \tilde{I}}{C}$$

with the $O(1)$ and $O_p(1)$ terms positive. Hence, $[S(\bar{\alpha}, \lambda) - S(\bar{\alpha}_0, \lambda^0)]/(|\lambda - \lambda^0| T) > 0$ on $V_\epsilon(C)$ with large probability. This implies that the minimum cannot be achieved on $V_\epsilon(C)$

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

and, thus, $\Pr(T|\hat{\lambda}-\lambda^0|\geq C)\leq\zeta$, so that $(\hat{\lambda}-\lambda^0)=O_p\left(T^{-1}\right)$. For Models 0b and IIb, a similar argument can be applied to show that $(\hat{\lambda}-\lambda^0)=o_p\left(T^{-1}\right)$. ■

Proof of Proposition 2. Consider first the statistic $P_T^{GLS}(c,\bar{c},\hat{\lambda})$ in Models I and II.

$$\begin{aligned}
s(\hat{\lambda})^2 P_T^{GLS}(c,\bar{c},\hat{\lambda}) &= S(\bar{\alpha},\lambda)-\bar{\alpha}S(1,\hat{\lambda})\\
&= Q^{\bar{\alpha}}(\hat{\lambda})-2G^{\bar{\alpha}}(\hat{\lambda})+u^{\bar{\alpha}\prime}M_{\hat{\lambda}}^{\bar{\alpha}}u^{\bar{\alpha}}-(1+\bar{c}/T)\left[Q^1(\hat{\lambda})-2G^1(\hat{\lambda})+u^{1\prime}M_{\hat{\lambda}}^1u^1\right]\\
&= [Q^{\bar{\alpha}}(\hat{\lambda})-Q^1(\hat{\lambda})]-\bar{c}T^{-1}Q^1(\hat{\lambda})-2\left[G^{\bar{\alpha}}(\hat{\lambda})-(1+\bar{c}/T)G^1(\hat{\lambda})\right]\\
&\quad+\left[u^{\bar{\alpha}\prime}M_{\hat{\lambda}}^{\bar{\alpha}}u^{\bar{\alpha}}-(1+\bar{c}/T)u^{1\prime}M_{\hat{\lambda}}^1u^1\right]\\
&= [Q^{\bar{\alpha}}(\hat{\lambda})-Q^1(\hat{\lambda})]+\left[u^{\bar{\alpha}\prime}M_{\hat{\lambda}}^{\bar{\alpha}}u^{\bar{\alpha}}-(1+\bar{c}/T)u^{1\prime}M_{\hat{\lambda}}^1u^1\right]+o_p(1).
\end{aligned}$$

The last equality follows from (A.19), (A.20), and Proposition 1. From this expression, we can see that if $[Q^{\bar{\alpha}}(\hat{\lambda})-Q^1(\hat{\lambda})]$ is asymptotically negligible, the limit distribution of $s(\hat{\lambda})^2P_T^{GLS}(c,\bar{c},\hat{\lambda})$ will be the same as that of $s(\lambda^0)^2P_T^{GLS}(c,\bar{c},\lambda^0)$. Now,

$$\begin{aligned}
Q^{\bar{\alpha}}(\hat{\lambda})-Q^1(\hat{\lambda}) &= \psi'd^{\bar{\alpha}\prime}(\hat{\lambda})d^{\bar{\alpha}}(\hat{\lambda})\psi-\psi'd^{\bar{\alpha}\prime}(\hat{\lambda})P_{\hat{\lambda}}^{\bar{\alpha}}d^{\bar{\alpha}}(\hat{\lambda})\psi\\
&\quad-\psi'd^{1\prime}(\hat{\lambda})d^1(\hat{\lambda})\psi+\psi'd^{1\prime}(\hat{\lambda})P_{\hat{\lambda}}^1d^1(\hat{\lambda})\psi.
\end{aligned}$$

From (A.8)–(A.10),

$$\psi'd^{\bar{\alpha}\prime}(\hat{\lambda})d^{\bar{\alpha}}(\hat{\lambda})\psi-\psi'd^{1\prime}(\hat{\lambda})d^1(\hat{\lambda})\psi=o_p(1),$$

and we recall that

$$d^{\bar{\alpha}\prime}(\hat{\lambda})P_{\hat{\lambda}}^{\bar{\alpha}}d^{\bar{\alpha}}(\hat{\lambda})=d^{\bar{\alpha}\prime}(\hat{\lambda})z^{\bar{\alpha}}(\hat{\lambda})D_T(D_Tz^{\bar{\alpha}\prime}(\hat{\lambda})z^{\bar{\alpha}}(\hat{\lambda})D_T)^{-1}D_Tz^{\bar{\alpha}\prime}(\hat{\lambda})d^{\bar{\alpha}}(\hat{\lambda}).$$

Then, from (A.18) and Proposition 1, we have (assuming for simplicity a single break)

$$\begin{aligned}
D_Tz^{\bar{\alpha}\prime}(\hat{\lambda})d^{\bar{\alpha}}(\hat{\lambda})\psi &= [\iota^{\bar{\alpha}},DU^{\bar{\alpha}}(\hat{T}_1),T^{-1/2}\tau^{\bar{\alpha}},T^{-1/2}DT^{*\bar{\alpha}}(\hat{T}_1)]'\\
&\quad\times\left[DU^{\bar{\alpha}}(\hat{T}_1)-DU^{\bar{\alpha}}\left(T_1^0\right),DT^{*\bar{\alpha}}(\hat{T}_1)-DT^{*\bar{\alpha}}\left(T_1^0\right)\right](\mu_1,\beta_1)'\\
&= \begin{pmatrix} o_p(1) & o_p(1)\\ 1+o_p(1) & 1(T_1<T_1^0)+o_p(1)\\ o_p(1) & o_p(1)\\ o_p(1) & o_p(1)\end{pmatrix}\begin{pmatrix}\mu_1\\ \beta_1\end{pmatrix}
\end{aligned}$$

and, similarly,

$$D_Tz^{1\prime}(\hat{\lambda})d^1(\hat{\lambda})\psi=\begin{pmatrix} o_p(1) & o_p(1)\\ 1+o_p(1) & 1(T_1<T_1^0)+o_p(1)\\ o_p(1) & o_p(1)\\ o_p(1) & o_p(1)\end{pmatrix}\begin{pmatrix}\mu_1\\ \beta_1\end{pmatrix}.$$

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press

Given the asymptotic block diagonality of $D_T z^{\bar{\alpha}\prime}(\hat{\lambda}) z^{\bar{\alpha}}(\hat{\lambda}) D_T$ and $D_T z^{1\prime}(\hat{\lambda}) z^{1}(\hat{\lambda}) D_T$, we have

$$\psi' d^{\bar{\alpha}\prime}(\hat{\lambda}) P^{\bar{\alpha}}_{\hat{\lambda}} d^{\bar{\alpha}}(\hat{\lambda})\psi - \psi' d^{1\prime}(\hat{\lambda}) P^{1}_{\hat{\lambda}} d^{1}(\hat{\lambda})\psi = o_p(1).$$

We focus on $MSB^{GLS}(\lambda)$ defined by (10). Note that $\tilde{y}'_{-1}\tilde{y}_{-1}$ is of the same order of magnitude as $\tilde{y}'\tilde{y}$, and so we concentrate on $\tilde{y}'\tilde{y}$. First, note that, in Models I and II,

$$\begin{aligned}\hat{\psi} &= (z^{\bar{\alpha}}(\hat{\lambda})' z^{\bar{\alpha}}(\hat{\lambda}))^{-1} z^{\bar{\alpha}}(\hat{\lambda})' y^{\bar{\alpha}} = (z^{\bar{\alpha}}(\hat{\lambda})' z^{\bar{\alpha}}(\hat{\lambda}))^{-1} z^{\bar{\alpha}}(\hat{\lambda})' z^{\bar{\alpha}}(\lambda^0)\psi \\ &\quad + (z^{\bar{\alpha}}(\hat{\lambda})' z^{\bar{\alpha}}(\hat{\lambda}))^{-1} z^{\bar{\alpha}}(\hat{\lambda})' u^{\bar{\alpha}} \\ &= \psi + (z^{\bar{\alpha}}(\hat{\lambda})' z^{\bar{\alpha}}(\hat{\lambda}))^{-1} z^{\bar{\alpha}}(\hat{\lambda})' u^{\bar{\alpha}} + (z^{\bar{\alpha}}(\hat{\lambda})' z^{\bar{\alpha}}(\hat{\lambda}))^{-1} z^{\bar{\alpha}}(\hat{\lambda})' \left[z^{\bar{\alpha}}(\lambda^0) - z^{\bar{\alpha}}(\hat{\lambda})\right]\psi\end{aligned}$$

and

$$\begin{aligned}\tilde{y} &= y - z(\hat{\lambda})\hat{\psi} = z(\lambda^0)(\psi - \hat{\psi}) + u + [z(\lambda^0) - z(\hat{\lambda})]\hat{\psi} \\ &= u - z(\lambda^0)(z^{\bar{\alpha}}(\hat{\lambda})' z^{\bar{\alpha}}(\hat{\lambda}))^{-1} z^{\bar{\alpha}}(\hat{\lambda})' u^{\bar{\alpha}} \\ &\quad - z(\lambda^0)(z^{\bar{\alpha}}(\hat{\lambda})' z^{\bar{\alpha}}(\hat{\lambda}))^{-1} z^{\bar{\alpha}}(\hat{\lambda})' [z^{\bar{\alpha}}(\lambda^0) - z^{\bar{\alpha}}(\hat{\lambda})]\psi + [z(\lambda^0) - z(\hat{\lambda})]\hat{\psi}.\end{aligned}$$

We need to show that the last two terms are asymptotically negligible compared to the first two. Note that the first two terms are such that

$$\left\| u - z(\lambda^0)(z^{\bar{\alpha}}(\hat{\lambda})' z^{\bar{\alpha}}(\hat{\lambda}))^{-1} z^{\bar{\alpha}}(\hat{\lambda})' u^{\bar{\alpha}} \right\| = O_p(T).$$

For the third term,

$$\begin{aligned}&\left\| z(\lambda^0)(z^{\bar{\alpha}}(\hat{\lambda})' z^{\bar{\alpha}}(\hat{\lambda}))^{-1} z^{\bar{\alpha}}(\hat{\lambda})' \left[z^{\bar{\alpha}}(\lambda^0) - z^{\bar{\alpha}}(\hat{\lambda})\right]\psi \right\| \\ &\quad = \Big(\psi \left[z^{\bar{\alpha}}(\lambda^0) - z^{\bar{\alpha}}(\hat{\lambda})\right] z^{\bar{\alpha}}(\hat{\lambda}) D_T S_T D_T z(\lambda^0)' z(\lambda^0) D_T S_T D_T z^{\bar{\alpha}}(\hat{\lambda})' \left[z^{\bar{\alpha}}(\lambda^0) \right. \\ &\qquad \left. - z^{\bar{\alpha}}(\hat{\lambda})\right]\psi \Big)^{1/2} \\ &\quad = \Big(\psi \left[z^{\bar{\alpha}}(\lambda^0) - z^{\bar{\alpha}}(\hat{\lambda})\right] z^{\bar{\alpha}}(\hat{\lambda}) D_T S_T D_T z(\lambda^0)' z(\lambda^0) D_T S_T D_T z^{\bar{\alpha}}(\hat{\lambda})' \left[z^{\bar{\alpha}}(\lambda^0) \right. \\ &\qquad \left. - z^{\bar{\alpha}}(\hat{\lambda})\right]\psi \Big)^{1/2} \\ &\quad = O_p(T^{1/2}),\end{aligned}$$

where $S_T = (D_T z^{\bar{\alpha}}(\hat{\lambda})' z^{\bar{\alpha}}(\hat{\lambda}) D_T)^{-1}$. The last equality follows from the fact that $S_T = O_p(1)$, $D_T z(\lambda^0)' z(\lambda^0) D_T = O_p(T^2)$ and, from the previous proof, $D_T z^{\bar{\alpha}}(\lambda)' [z^{\bar{\alpha}}(\lambda^0) - z^{\bar{\alpha}}(\hat{\lambda})]\psi = O_p(T^{-1/2})$. For the last term,

$$\begin{aligned}\left\| [z(\lambda^0) - z(\hat{\lambda})]\hat{\psi} \right\| &= (\hat{\psi}' [z(\lambda^0) - z(\hat{\lambda})]' [z(\lambda^0) - z(\hat{\lambda})]\hat{\psi})^{1/2} \\ &= |\hat{\lambda} - \lambda^0| O_p(T^{3/2}) = O_p(T^{1/2}).\end{aligned}$$

The CMT with the consistency of $s^2(\lambda)$ completes the proof for $MSB^{GLS}(\lambda)$. An analogous argument applies to $\tilde{y}_T$, and the proofs for $MZ^{GLS}_{\alpha}(\lambda)$ and $MZ^{GLS}_{t}(\lambda)$ directly follow. ■

https://doi.org/10.1017/S0266466609990326 Published online by Cambridge University Press