Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Econometrics 129 (2005) 65–119
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

Structural breaks with deterministic and stochastic trends

Pierre Perron[a,*], Xiaokang Zhu[b]

[a]*Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA*
[b]*Cornerstone Research, 1000 El Camino Real, Menlo Park, CA 94025, USA*

Available online 13 November 2004

Abstract

This paper analyzes the consistency, rate of convergence and limiting distributions of parameter estimates in models where the trend function exhibits a slope change at some unknown date and the errors can be either stationary or have a unit root. These estimates are obtained by minimizing the sum of squared residuals in simple regressions involving a constant, a trend, a slope shift regressor and possibly a level shift regressor. Special attention is given to the effects induced by alternative specifications of the slope shift regressor and the inclusion or exclusion of a level shift regressor. Some surprising results are found for which we provide more detailed explanations. We also show via simulations that our asymptotic results provide good approximations in finite samples. We illustrate the issues analyzed applying our results to investigate dates and magnitudes of changes in the growth rates of (log) real per capita GDP series for 10 countries using a historical data set that covers the period 1870–1986.

JEL classification: C20

Keywords: Change-point; Segmented regressions; Break dates; Hypothesis testing; Model selection

*Corresponding author. Tel.: +1 617 353 3026; fax: +1 617 353 4449.
E-mail addresses: perron@bu.edu (P. Perron), kzhu@cornerstone.com (X. Zhu).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2004.09.004

1. Introduction

Both the statistics and econometrics literature contain a vast amount of work on issues related to structural change (see, e.g., the surveys by Krishnaiah and Miao, 1988; Bhattacharya, 1994 as well as the monograph by Csörgó and Horváth, 1997). The econometric literature has witnessed recently an upsurge of interest in extending procedures to various models with an unknown change point. With respect to the problem of testing for structural change, recent contributions include Andrews (1993) and Andrews and Ploberger (1994). Issues about the distributional properties of the estimates, in particular those of the break date, have also been considered by Bai (1994, 1997). These testing and inference issues have been addressed in the context of multiple structural changes by Bai and Perron (1998, 2003).

Most of the work in this literature has concentrated on the case where the regressors and the errors are stationary. Issues related to structural change are also important in the context of trending regressors and non-stationary time series following the work of Perron (1989), in particular. In that paper, it was argued that inference about unit roots are affected by changes in the intercept and slope of the trend function of the time series. In the literature that ensued, many different procedures have been suggested pertaining to the unit root problem but surprisingly little work have addressed the problem of estimating the break dates and forming confidence intervals. The aim of this paper is to fill this gap by analyzing the consistency, rate of convergence and limiting distributions of parameter estimates in models where the trend function exhibit a slope change at some unknown date and the errors can be either stationary or have a unit root.

Work related to this problem include Feder (1975) who considers estimating the joint points of polynomial type segmented regressions. Closely related to our work is the study of Bai et al. (1998) who analyze the limiting distribution of an estimated break for non-stationary type series with a slope change. The analysis is, however, quite different and more restrictive insofar as inference about a change in a linear trend is concerned (see Remark 3 below). Our results, in particular, allows different specifications concerning the role of an intercept shift. Other contributions include Chu and White (1992) who provide a test for a change in a trend function. The problem of testing for a change in the trend function of a series allowing the errors to be stationary or integrated has also been addressed by Perron (1991) and Vogelsang (1997). Of related interest are also the studies of Hansen (1992) and Hansen and Johansen (1999) who build on the work of Nyblom (1989) to study structural break in integrated variables. Lumsdaine and Papell (1997) consider unit root tests allowing for two structural changes in the trend function. Inoue (1999) presents a test to establish the cointegrating rank of a system of variables in the presence of a trend break at an unknown date. Seo (1998) derived tests for structural breaks within a cointegrated vector autoregressive system, though his study does not allow for a break in the trend function. Finally, Hansen (2003) considers multiple breaks in any parameter of a cointegrated vector autoregressive system, though he assumes the break dates to be known.

The issues to be investigated are best motivated using real data series. To that effect, we consider an historical data set of (log) real per capita GDP series from 1870 to 1986 for 10 different countries: Australia, Canada, Denmark, France, Germany, Italy, Norway, Sweden, the United Kingdom and the United States.[1] The series are presented in Fig. 1 where the dotted line is a fitted trend function obtained by regressing the series on a constant, a trend, an intercept shift and a slope shift where the break date is selected by minimizing the sum of squared residuals from the regression (see Model II in Section 2). From this figure it seems clear that most series are characterized by at least one (and in most cases only one) major shift in the slope of the trend function (with perhaps the exception of the United States). At the same time as the shift in slope occurs there is a clear tendency, for most countries, to have a level shift. This level shift is, however, of different relative importance across counties, being strongest for France and weakest for Norway and Sweden. Also, the dates of the breaks are not common across countries. They occur at the time of World War II for France, Germany and Italy; at the beginning of the 1930s for Australia and around the time of World War I for Sweden and the United Kingdom. Of interest to characterize the nature of these series is whether the noise component can be viewed as stationary or integrated (i.e., whether the series have a unit root or not). Using tests that allow for a change in intercept and slope at an unknown date, Perron (1992) concludes that the unit root is rejected at the 5% level for Australia, Canada, Denmark, France, Germany, the United Kingdom and the United States. No rejection of the unit root was possible for Italy, Norway and Sweden.

The questions that arise from these real GDP series are the following: what are the properties of the estimated break dates obtained by minimizing the sum of squared residuals from a simple regression on the deterministic components? Are they consistent, what is the rate of convergence and the limiting distribution? Do the results differ if one assumes the noise component to be stationary or to have an autoregressive unit root? Are the estimates still consistent if a unit root is present? How do these compare with those obtained in a context where both the regressors and the errors are assumed to be stationary? How should one treat the level shift? Should we discard it when the shift appears small and if so does that change the properties of the estimated break dates and other parameters? In general, what is the effect of a level shift on the precision of the estimated break date? Can a large level shift improve the rate of convergence? If so, what kind of modelling device is needed?

These and other issues will be addressed in this paper whose structure is as follows. Section 2 first describes the models considered, the assumptions made on the various

[1]This data set is the same as used by Kormendi and Meguire (1990) and Perron (1992) and was obtained through the *Journal of Money, Credit and Banking* editorial office. All series are real GDP except for the United States for which real GNP is used. For the United States, the series is real GNP from the National Income and Products Accounts for the period 1929–1986, spliced to Romer's (1989) estimates for the period 1870–1928. For the United Kingdom, the series is real GDP from Feinstein (1972) for the period 1870–1947 spliced to the International Financial Statistics (IFS) series of the IMF for the period 1948–1986. For the remaining countries, the series are indices of annual real GDP from Maddison (1982) spliced to the postwar IFS data. The population series used are from the same sources. A logarithmic transformation is applied.

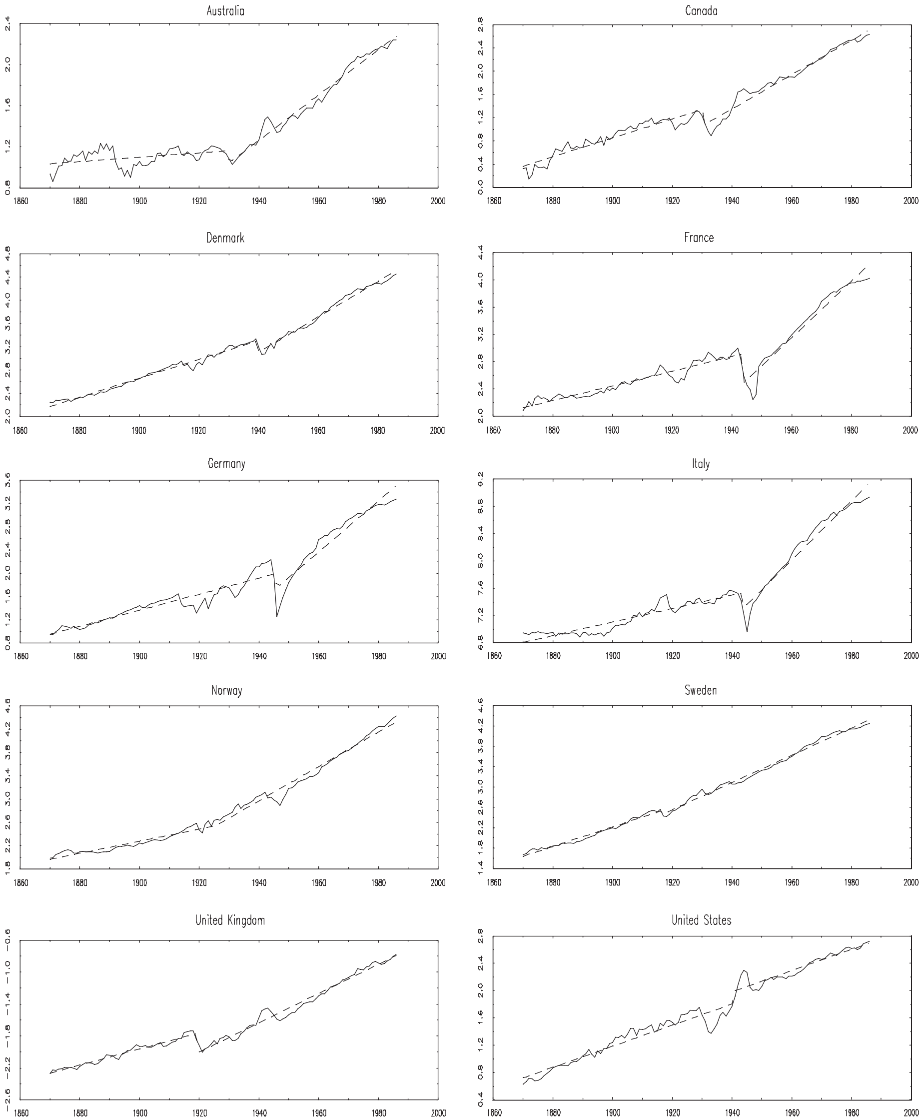

Fig. 1. Annual log per capita gross domestic product (GDP): 1870–1986. The fitted trend function is obtained by regressing the series on a constant, a trend, an intercept shift and a slope shift where the break date is selected by minimizing the sum of squared residuals from the regression (see Model II).

components and how the estimates are obtained. Three models are considered: a joint broken trend (no level shift), a local disjoint broken trend (appropriate in the case when the level shift is relatively small), and a global disjoint broken trend (appropriate when the level shift is relatively large). In all cases, we consider two assumptions on the noise component, namely that it is either stationary or is integrated of order one. Section 3 is the main body of the paper where for the six scenarios considered we derive the consistency and rate of convergence of the estimates of the break dates (or break fractions) and the limiting distributions of these estimates as well as those of the other parameters. Section 4 presents simulation results to show that our asymptotic results are good approximations to the finite sample distributions and to illustrate when one should include a level shift as a regressor. Section 5 offers intuitive explanations for some results that appear surprising at first, in particular the fact that including a level shift regressor when not needed can actually reduce the rate of convergence of the estimate of the break fraction and induce a bimodal distribution. Section 6 presents empirical results for the real per capita GDP series discussed above and Section 7 offers brief concluding remarks. All derivations are included in a mathematical appendix.

2. The models considered

2.1. Deterministic and stochastic trends

Throughout, it is assumed that some variable of interest, y_t, is the sum of some systematic part d_t and a random component, u_t, i.e.

$$y_t = d_t + u_t.$$

The models analyzed differ according to the assumptions made about both components. For u_t, we specify $\mathrm{E}(u_t) = 0$ and alternatively one of the following two assumptions:

Assumption 1. $u_t \sim I(0)$. More specifically u_t is such that $T^{-1/2}\sum_{t=1}^{[Tr]} u_t \Rightarrow \sigma W(r)$ where $\sigma^2 = \lim_{T\to\infty} T^{-1}\mathrm{E}(\sum_{t=1}^{T} u_t)^2$ exists and is strictly positive. Here "$\Rightarrow$" denotes weak convergence in distribution (under the sup metric) and $W(r)$ is the unit Wiener process.

Assumption 2. $u_t \sim I(1)$. More specifically $u_t = \sum_{j=1}^{t} \varepsilon_j$ where the sequence ε_t is assumed to be $I(0)$ as defined in Assumption 1.

Remark 1. There are many sets of sufficient conditions to guarantee that the weak convergence result stated in Assumption 1 holds. One that is fairly general is that used in Phillips and Perron (1988), namely (a) $\sup_t E|u_t|^{\gamma+\eta} < \infty$ for some $\gamma > 2$ and $\eta > 0$ and (b) $\{u_t\}_1^\infty$ is strong mixing with mixing numbers α_m that satisfy $\sum_1^\infty \alpha_m^{1-2/\gamma} < \infty$. Alternatively, we can assume that u_t is a linear process such that $u_t = \sum_{i=0}^{\infty} c_i e_{t-i}$ where $\{e_t, \mathscr{F}_{t-1}\}$ is a martingale difference sequence with $\mathscr{F}_{t-1}$ the filtration to which e_t is adapted. Also $\sum_{i=0}^{\infty} i|c_i| < \infty$ (see Phillips and Solo, 1992). Either sets of conditions include the popular stationary and invertible ARMA processes.

For the systematic component d_t, we consider three cases. The first specifies that d_t is a first-order linear trend with a one time change in slope such that the trend function is joined at the time of break. The second specifies that d_t is a first-order linear trend with a one time change in intercept and slope such that without an intercept change, the trend function is joined at the time of break. The third specification is similar except that the trend function is not restricted to be joined at the time of break (in the absence of a change in intercept). The time of break is denoted T_1 and we define the break fraction as $\lambda = T_1/T$. Hence, we have six different models labelled as follows: I.a-joint broken trend with $I(1)$ errors; I.b-joint broken trend with $I(0)$ errors; II.a-local disjoint broken trend with $I(1)$ errors; II.b-local disjoint broken trend with $I(0)$ errors; III.a-global disjoint broken trend with $I(1)$ errors; and III.b-global disjoint broken trend with $I(0)$ errors. We start by specifying more precisely the data generating process (DGP) for each model.

2.1.1. *Models I.a and I.b: joint broken trend with $I(1)$ or $I(0)$ errors*

For the first two models, d_t is specified by

$$d_t = \mu_1 + \beta_1 t + \beta_b B_t, \tag{1}$$

where B_t is a dummy variable for the slope change defined by

$$B_t = \begin{cases} 0 & \text{if } t \leqslant T_1, \\ t - T_1 & \text{if } t > T_1. \end{cases}$$

Here, the slope coefficient (or the rate of growth when a logarithmic transform is applied) changes from β_1 to $\beta_1 + \beta_b$ at the time point T_1. However, the trend function is continuous at T_1. For this reason, we refer to this specification as a "joint broken trend".

2.1.2. *Models II.a and II.b: local disjoint broken trend with $I(1)$ or $I(0)$ errors*

For these two models, d_t is specified by

$$d_t = \mu_1 + \beta_1 t + \mu_b C_t + \beta_b B_t, \tag{2}$$

where C_t is a dummy variable for the level shift defined by

$$C_t = \begin{cases} 0 & \text{if } t \leqslant T_1, \\ 1 & \text{if } t > T_1. \end{cases}$$

Note that μ_b and β_b capture the change in the intercept and slope coefficients. At the break point T_1, the slope changes by β_b and the level shifts by μ_b, which is negligible compared to the level of the series $\mu_1 + \beta_1 T_1$, hence the label "local disjoint segmented trend".

Remark 2. As pointed out by Hatanaka and Yamada (1999) and others, when the error term is an $I(1)$ process, the shift in the intercept cannot be identified. Intuitively, it is impossible to distinguish a permanent shift in the intercept from a permanent shock to the process (the $I(1)$ error). Later, we provide a rigorous discussion about this identification issue. More surprisingly, we will show that even when the error term is an $I(0)$ process, the intercept shift cannot be identified, either. An intuitive explanation will be provided later.

2.1.3. Models III.a and III.b: global disjoint broken trend with $I(1)$ or $I(0)$ errors

If one wants to model a permanent shift in the level of the series such that the trend function is discontinuous at the break date even asymptotically, we can specify the DGP as

$$d_t = \mu_1 + \beta_1 t + \mu_b C_t + \beta_b B_t^{dj}, \tag{3}$$

where

$$B_t^{dj} = \begin{cases} 0 & \text{if } t \leqslant T_1, \\ t & \text{if } t > T_1. \end{cases}$$

We label this model as a "global disjoint segmented trend" since, in contrast to the previous "local disjoint segmented trend", the implied relative (to the overall level of the trend function) level shift at the break date converges to $\beta_b/\beta_1 \neq 0$ as $T \to \infty$, since $d_{T_1+1} - d_{T_1} = \beta_1 + \mu_b + \beta_b T_1$.

Note that, in practice, using Model II or III yields exactly the same results for the estimates of the parameters T_1, μ_1, β_1 and β_b. Nevertheless, as we shall see, the two specifications yield drastically different asymptotic results, in particular pertaining to the rate of convergence and the asymptotic distribution of the estimated break date. As our simulations will highlight, limiting results obtained from model II (local disjoint trend) will provide good approximations to the finite sample distributions when the shift in level is small while those from model III will do so when the shift in level is large. Hence, both asymptotic frameworks are complementary. These issues are discussed in details in Section 4.

Remark 3. A special case of the general model considered by Bai et al. (1998) is that of Model III with I(0) errors, with or without restricting $\mu_b = 0$. They also consider an asymptotic framework, whereby the coefficient β_b shrinks to zero at some suitable rate. In the following, all results are obtained for fixed coefficients.

Remark 4. The DGP specified by Models II and III can be generalized to encompass both as special cases by extending Model II allowing μ_b to be a function of the sample size, i.e. $\mu_b = \kappa T^{\alpha}$ for some $\kappa > 0$ and $\alpha \geqslant 0$. Model II obtains when $\alpha = 0$ while we recover Model III with $\alpha = 1$. It turns out that the results we shall derive for Model II.a holds more generally when $\alpha < 1/2$ while those derived for Model III.a hold when $\alpha > 1/2$. When the errors are $I(0)$, the results for Model III.b applies for all values of $\alpha > 0$. Hence, we shall continue with the classification described above and discuss in various remarks how the results extend to the more general case. Note, however, that Model III is not quite a special case since the regressors are not the same. Nevertheless, the general conclusions will be the same, in particular for the estimate of the break date.

2.2. The estimation method and a key inequality

All specifications discussed above can be expressed in matrix notation as

$$Y = X_{T_1}\gamma + U,$$

where $Y' = [y_1, \ldots, y_T]$, $U' = [u_1, \ldots, u_T]$, $X'_{T_1} = [x(T_1)_1, \ldots, x(T_1)_T]$, $\gamma' = (\mu_1, \beta_1, \mu_b, \beta_b)$ and where, for Models I, $x(T_1)'_t = [1, t, B_t]$, for Models II, $x(T_1)'_t = [1, t, C_t, B_t]$, and for Models III, $x(T_1)'_t = [1, t, C_t, B_t^{dj}]$. Note that the matrix X_{T_1} depends on the postulated value of the break date T_1. The parameters are assumed to be obtained using a global least-squares criterion. In particular, we have the following estimate for the break date:

$$\hat{T}_1 = \arg\min_{T_1} \; Y'(1 - P_{T_1})Y,$$

where P_{T_1} is the projection matrix constructed using X_{T_1}, i.e., $P_{T_1} = X_{T_1}(X'_{T_1}X_{T_1})^{-1}X'_{T_1}$. Denoting by $X_{\hat{T}_1}$ the matrix X constructed using the least-squares estimate of the break date $\hat{T}_1$, the least-squares estimate of the coefficients γ is

$$\hat{\gamma} = (X'_{\hat{T}_1}X_{\hat{T}_1})^{-1}X'_{\hat{T}_1}Y$$

and the resulting sum of squared residuals is, for an estimated break fraction $\hat{\lambda} = \hat{T}_1/T$,

$$\mathrm{SSR}(\hat{\lambda}) = \sum_{t=1}^{T} \hat{u}_t^2 = \sum_{t=1}^{T} (y_t - x(\hat{T}_1)'_t\hat{\gamma})^2 = Y'(I - P_{\hat{T}_1})Y$$

where $P_{\hat{T}_1}$ is the projection matrix associated with $X_{\hat{T}_1}$, i.e. $P_{\hat{T}_1} = X_{\hat{T}_1}(X'_{\hat{T}_1}X_{\hat{T}_1})^{-1}X'_{\hat{T}_1}$.

The true values of the unknown coefficients will be denoted with a 0 superscript, i.e. $\gamma^0 = (\mu_1^0, \beta_1^0, \mu_b^0, \beta_b^0)'$, T_1^0, $\lambda^0 = T_1^0/T$; $X_{T_1^0}$ is the matrix of regressors constructed using the true value T_1^0 for the break date, and $P_{T_1^0}$ is the associated projection matrix, i.e. $P_{T_1^0} = X_{T_1^0}(X'_{T_1^0}X_{T_1^0})^{-1}X'_{T_1^0}$. So the true data generating process is assumed to be

$$Y = X_{T_1^0}\gamma^0 + U.$$

Our aim is to derive the limit distribution of $(\hat{\lambda} - \lambda^0)$ and $(\hat{\gamma} - \gamma^0)$ for the six cases described above assuming that there exists at least a shift in slope. To that effect, we make the following assumption on the true coefficients.

Assumption 3. $\beta_b^0 \neq 0$ and $\lambda^0 \in (0, 1)$.

This assumption basically ensures that we have a one time break in the systematic part and that the pre and post break samples are not asymptotically negligible which is a standard assumption needed to derive any useful asymptotic result. Note that in this asymptotic framework, the break date T_1^0 increases as T increases. Hence, one cannot properly consider the issue of whether the estimate $\hat{T}_1$ is consistent for T_1. Accordingly, we derive all consistency and distributional results in terms of the estimate of the break fraction $\hat{\lambda} = \hat{T}_1/T$ for λ^0 which is a fixed quantity as T increases. We then use the resulting limit distributions to provide approximations to the finite sample distributions of $(\hat{T}_1 - T_1^0)$.

2.3. A key inequality

We now discuss a key inequality that will be used extensively to derive the various limiting results. From the properties of projections, we have for all T,

$$\mathrm{SSR}(\hat{\lambda}) \leqslant \mathrm{SSR}(\lambda^0),$$

or

$$Y'(I - P_{\hat{T}_1})Y \leqslant Y'(I - P_{T_1^0})Y.$$

Since $Y = X_{T_1^0}\gamma^0 + U$, the above inequality implies that

$$\gamma^{0\prime} X'_{T_1^0}(P_{T_1^0} - P_{\hat{T}_1})X_{T_1^0}\gamma^0 + 2\gamma^{0\prime} X'_{T_1^0}(P_{T_1^0} - P_{\hat{T}_1})U + U'(P_{T_1^0} - P_{\hat{T}_1})U \leqslant 0$$

for all T. Note that $P_{T_1^0} X_{T_1^0} = X_{T_1^0}$ and $X'_{\hat{T}_1}(I - P_{\hat{T}_1}) = 0$, hence we can rewrite the above inequality as follows:

$$\begin{aligned} &\gamma^{0\prime}(X_{T_1^0} - X_{\hat{T}_1})'(I - P_{\hat{T}_1})(X_{T_1^0} - X_{\hat{T}_1})\gamma^0 \\ &\quad + 2\gamma^{0\prime}(X_{T_1^0} - X_{\hat{T}_1})'(I - P_{\hat{T}_1})U + U'(P_{T_1^0} - P_{\hat{T}_1})U \leqslant 0. \end{aligned} \tag{4}$$

This is the key inequality that we shall use repeatedly when deriving the asymptotic properties for the different models. Finally, it is worth noting that

$$\begin{aligned} \arg\min_{T_1}[\mathrm{SSR}(T_1)] &= \arg\min_{T_1}[\mathrm{SSR}(T_1) - \mathrm{SSR}(T_1^0)] \\ &= \arg\min_{T_1}[\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'(I - P_{T_1})(X_{T_1^0} - X_{T_1})\gamma^0 \\ &\qquad + 2\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'(I - P_{T_1})U + U'(P_{T_1^0} - P_{T_1})U]. \end{aligned}$$

This will be employed to derive the asymptotic distribution of the least-squares estimate of the break fraction, $\hat{\lambda} = \hat{T}_1/T$.

3. Asymptotic properties

3.1. Consistency

We first consider the issue of the consistency of the estimated break fraction. We show that $\hat{\lambda}$ is consistent for λ^0 in all models. To prove this, we show that, in inequality (4), the first (non-negative) term would asymptotically dominate the second and third terms (i.e., the first term grows at a faster rate than the other two) if $\hat{\lambda}$ does not converge to λ^0. Hence, the inequality cannot hold asymptotically if $\hat{\lambda}$ does not converge to λ^0. To this end, we first prove the following lemma. Note that here and throughout the text, we use the label $\mathrm{O}(T^a)$ and $\mathrm{O_p}(T^a)$ in its strict sense, i.e. meaning that the variables are not $\mathrm{o}(T^a)$ and $\mathrm{o_p}(T^a)$.

Lemma 1. *Define*

$$(XX) \equiv \gamma'(X_{T_1^0} - X_{T_1})'(I - P_{T_1})(X_{T_1^0} - X_{T_1})\gamma,$$

$$(XU) \equiv \gamma'(X_{T_1^0} - X_{T_1})'(I - P_{T_1})U,$$

$$(UU) \equiv U'(P_{T_1^0} - P_{T_1})U.$$

Under Assumptions 1–3, *we have that uniformly over all generic* $T_1 \in [\pi T, (1-\pi)T]$ *for some arbitrarily small* π *such that* $\lambda^0 \in [\pi, 1-\pi]$:[2]

1. *In Model I.a* (*joint broken trend with* $I(1)$ *errors*),
$$(XX) = |T_1 - T_1^0|^2 \mathrm{O}(T), \quad (XU) = |T_1 - T_1^0| \mathrm{O_p}(T^{3/2}) \text{ and}$$
$$(UU) = |T_1 - T_1^0| \mathrm{O_p}(T).$$

2. *In Model I.b* (*joint broken trend with* $I(0)$ *errors*),
$$(XX) = |T_1 - T_1^0|^2 \mathrm{O}(T), \quad (XU) = |T_1 - T_1^0| \mathrm{O_p}(T^{1/2}) \text{ and}$$
$$(UU) = |T_1 - T_1^0| \mathrm{O_p}(T^{-1}).$$

3. *In Model II.a* (*local disjoint broken trend with* $I(1)$ *errors*),
$$(XX) = |T_1 - T_1^0|^3 \mathrm{O}(1), \quad (XU) = |T_1 - T_1^0|^2 \mathrm{O_p}(T^{1/2}) \text{ and}$$
$$(UU) = |T_1 - T_1^0| \mathrm{O_p}(T).$$

4. *In Model II.b* (*local disjoint broken trend with* $I(0)$ *errors*),
$$(XX) = |T_1 - T_1^0|^3 \mathrm{O}(1), \quad (XU) = |T_1 - T_1^0|^{3/2} \mathrm{O_p}(1) \text{ and}$$
$$(UU) = |T_1 - T_1^0|^{1/2} \mathrm{O_p}(T^{-1/2}).$$

5. *In Model III.a* (*global disjoint broken trend with* $I(1)$ *errors*),
$$(XX) = |T_1 - T_1^0| \mathrm{O}(T^2), \quad (XU) \leqslant |T_1 - T_1^0| \mathrm{O_p}(T^{3/2}) \text{ and}$$
$$(UU) \leqslant |T_1 - T_1^0| \mathrm{O_p}(T).$$

6. *In Model III.b* (*global disjoint broken trend with* $I(0)$ *errors*),
$$(XX) = |T_1 - T_1^0| \mathrm{O}(T^2), \quad (XU) \leqslant |T_1 - T_1^0| \mathrm{O_p}(T) \text{ and}$$
$$(UU) \leqslant |T_1 - T_1^0| \mathrm{O_p}\left(T^{-1/2}\right).$$

[2] This trimming is assumed to ensure the invertibility of $X'_{T_1} X_{T_1}$ in the projection matrix. Alternatively, one could simply drop the regressors C_t and B_t to calculate SSR when $T_1 = 0$ or T. This trimming is just a technical device used for simplicity. In practice, we need not use a trimming since $\hat{T}_1$ will equal to 0 or T with zero probability given that we assume a break exists.

Note that the term (XX) is always non-negative since it is quadratic and $(I - P_{\hat{T}_1})$ is positive semi-definite. Given the above results, it is easy to enquire about the consistency of $\hat{\lambda}$. For example, in Model I, if $\hat{\lambda} \nrightarrow \lambda^0$, then $(XX) = \mathrm{O}(T^3)$, $(XU) = \mathrm{O_p}(T^{5/2})$ and $(UU) = \mathrm{O_p}(T)$. Therefore for large enough T, with some positive probability, the positive term (XX) dominates the other two terms (XU) and (UU) such that inequality (4) will not hold with probability 1. Since we know that the inequality (4) is true for all T, we have a contradiction and, hence, this implies that $\hat{\lambda} \rightarrow_{\mathrm{p}} \lambda^0$. The consistency of $\hat{\lambda}$ to λ^0 is summarized in the following theorem.

Theorem 2. *Under Assumptions* 1–3, *in Models* I–III, $\hat{\lambda}$ *converges to* λ^0 *in probability.*

Remark 5. For the generalized Model II with $\mu_b^0 = \kappa T^{\alpha}$, consistency holds as well.

3.2. Rate of convergence

Having investigated the issue of convergence of the estimate of the break fraction, we can then derive the convergence rates which are summarized in the following Theorem.

Theorem 3. *Under Assumptions* 1–3, *the rates of convergence of* $\hat{\lambda}$ *are* (*with* $I(1)$ *errors for Models*-*a and* $I(0)$ *errors for Models*-*b*):

1. *In Models I.a and II.a*, $\hat{\lambda} - \lambda^0 = \mathrm{O_p}(T^{-1/2})$;
2. *In Model I.b*, $\hat{\lambda} - \lambda^0 = \mathrm{O_p}(T^{-3/2})$;
3. *In Model II.b*, $\hat{\lambda} - \lambda^0 = \mathrm{O_p}(T^{-1})$;
4. *In Models III.a and III.b*, $|\hat{\lambda} - \lambda^0| = \mathrm{o_p}(T^{-3})$.

Remark 6. In Model II.b, $(\hat{\lambda} - \lambda^0)$ is $\mathrm{O_p}(T^{-1})$ while it is $\mathrm{O_p}(T^{-3/2})$ in Model I.b, hence the break fraction converges at a faster rate when the intercept shift regressor is absent. This is an important and rather surprising result. Later, we will show in details how allowing for an intercept shift may contaminate how precisely the break date is estimated.

Remark 7. Consider the generalized Model II.a with $\mu_b^0 = \kappa T^{\alpha}$. When $\alpha < 1/2$, the result is the same as for Model II.a; when $\alpha > 1/2$, we have $(\hat{\lambda} - \lambda^0) = \mathrm{o_p}(T^{-2\alpha-1})$ and the qualitative conclusions are the same as for Model III.a. When the errors are $I(0)$, things are different. We have $(\hat{\lambda} - \lambda^0) = \mathrm{o_p}(T^{-2\alpha-1})$ for all $\alpha > 0$.

3.3. Limiting distribution of the estimate of the break date

Having considered the issue of convergence of the estimated break fraction to its true value and the rate at which it does converge, we can now consider the problem of deriving its limiting distribution. Note, however, that this can only be done for those cases for which it was possible to obtain a given rate of convergence. Accordingly, this section omits Model III from the analysis. Of particular interest, our results show that the estimated break date has a Normal asymptotic distribution in Model I while for Models II the limiting distribution is a complicated function of two-sided random processes which can, nevertheless, be simulated to obtain appropriate confidence intervals. The following Theorem summarizes the main results obtained whose proofs can be found in the appendix.

Theorem 4. *Under Assumptions* 1–3, *we have* (*with* $I(1)$ *errors for Model.a and* $I(0)$ *errors for Model.b*):

1. *In Model I.a*, $\sqrt{T}(\hat{\lambda}-\lambda)\rightarrow^{d} \mathrm{N}(0, 2\sigma^2/(15(\beta_b^0)^2))$;
2. *In Model I.b*, $T^{3/2}(\hat{\lambda}-\lambda)\rightarrow^{d} \mathrm{N}(0, 4\sigma^2/[\lambda^0(1-\lambda^0)(\beta_b^0)^2])$;
3. *For Model II.a, define* $\xi_1 \equiv [\int_0^1 W(r)\,\mathrm{d}r, \int_0^1 rW(r)\,\mathrm{d}r, \int_{\lambda^0}^1 W(r)\,\mathrm{d}r, \int_{\lambda^0}^1 (r-\lambda^0)W(r)\,\mathrm{d}r]'$, $\xi_2 = [0, 0, W(\lambda^0), \int_{\lambda^0}^1 W(r)\,\mathrm{d}r]'$, $\xi_3 \equiv \int_0^{\lambda^0} [(3r^2 - 2r\lambda^0)/(\lambda^0)^2]\,\mathrm{d}W(r)$, $\xi_4 \equiv \int_{\lambda^0}^1 [(r-1)(3r-2\lambda^0-1)/(1-\lambda^0)^2]\,\mathrm{d}W(r)$,

$$\Omega_1 \equiv \begin{bmatrix} \frac{4}{\lambda^0} & -\frac{6}{(\lambda^0)^2} & \frac{2}{\lambda^0} & \frac{6}{(\lambda^0)^2} \\ -\frac{6}{(\lambda^0)^2} & \frac{12}{(\lambda^0)^3} & -\frac{6}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3} \\ \frac{2}{\lambda^0} & -\frac{6}{(\lambda^0)^2} & \frac{4}{\lambda^0(1-\lambda^0)} & 6\,\frac{1-2\lambda^0}{(\lambda^0)^2(1-\lambda^0)^2} \\ \frac{6}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3} & 6\,\frac{1-2\lambda^0}{(\lambda^0)^2(1-\lambda^0)^2} & 12\,\frac{3(\lambda^0)^2-3\lambda^0+1}{(\lambda^0)^3(1-\lambda^0)^3} \end{bmatrix},$$

$$\Omega_2 \equiv \begin{bmatrix} -\frac{4}{(\lambda^0)^2} & \frac{12}{(\lambda^0)^3} & -\frac{2}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3} \\ \frac{12}{(\lambda^0)^3} & -\frac{36}{(\lambda^0)^4} & \frac{12}{(\lambda^0)^3} & \frac{36}{(\lambda^0)^4} \\ -\frac{2}{(\lambda^0)^2} & \frac{12}{(\lambda^0)^3} & 4\,\frac{2\lambda^0-1}{(\lambda^0)^2(1-\lambda^0)^2} & \frac{12}{(\lambda^0)^3}\,\frac{3(\lambda^0)^2-3\lambda^0+1}{(\lambda^0-1)^3} \\ -\frac{12}{(\lambda^0)^3} & \frac{36}{(\lambda^0)^4} & \frac{12}{(\lambda^0)^3}\,\frac{3(\lambda^0)^2-3\lambda^0+1}{(\lambda^0-1)^3} & \frac{36}{(\lambda^0)^4}\,\frac{4(\lambda^0)^3-6(\lambda^0)^2+4\lambda^0-1}{(\lambda^0-1)^4} \end{bmatrix}.$$

Also define $Z^*(m)$ *as follows*: $Z^*(0)=0$, $Z^*(m)=Z_1(m)$ *for* $m<0$ *and* $Z^*(m)=Z_2(m)$ *for* $m>0$, *with*

$$Z_1(m) = (\beta_b^0)^2|m|^3/3 + m^2\sigma\beta_b^0\xi_4 + m\sigma^2[2\xi_2'\Omega_1\xi_1 - \xi_1'\Omega_2\xi_1], \quad m<0,$$

$$Z_2(m) = (\beta_b^0)^2|m|^3/3 + m^2\sigma\beta_b^0\xi_3 + m\sigma^2[2\xi_2'\Omega_1\xi_1 - \xi_1'\Omega_2\xi_1], \quad m>0.$$

Then, $\sqrt{T}(\hat{\lambda}-\lambda)\rightarrow^{d} m_{\mathrm{III}}^{\infty} \equiv \arg\min_m Z^*(m)$.

4. *For Model II.b, define a stochastic process* $S^*(m)$ *on the set of integers as follows*: $S^*(0)=0$, $S^*(m)=S_1(m)$ *for* $m<0$ *and* $S^*(m)=S_2(m)$ *for* $m>0$, *with*

$$S_1(m) = \sum_{k=m+1}^{0} (\mu_b^0 + \beta_b^0 k)^2 - 2\sum_{k=m+1}^{0} (\mu_b^0 + \beta_b^0 k)u_k, \quad m=-1,-2,\ldots,$$

$$S_2(m) = \sum_{k=1}^{m} (\mu_b^0 + \beta_b^0 k)^2 + 2\sum_{k=1}^{m} (\mu_b^0 + \beta_b^0 k)u_k, \quad m=1,2,\ldots\ .$$

If $\{u_t\}$ is strictly stationary with a continuous distribution, S^* is a two-sided random walk with drift, and $T(\hat{\lambda}-\lambda)\to^{d} m_{\mathrm{IV}}^{\infty} \equiv \arg\min_m S^*(m)$.

Remark 8. Consider the generalized Model II.a with $\mu_b^0 = \kappa T^{\alpha}$. When $\alpha<1/2$, the limit distribution is the same as for Model II.a; when $\alpha = 1/2$, it is slightly different due to the presence of an additional terms.

These results show interesting qualitative differences across models. First, note that for Models I.a and I.b, the limiting distributions of the break date do no depend on the structure of the errors (apart from the variance term σ^2 needed to properly scale the distribution). In particular, the results remain the same irrespective of the nature of the serial correlation. This is in stark contrast to results obtained in a stationary context in which case the limiting distribution of the estimated break date, in this fixed shift case, not only depends on the properties of the residuals but in particular on their exact distribution in finite samples (see, e.g., Bai, 1997). This feature in the stationary case, has led to the development of asymptotic distributions obtained under the so-called "shrinking shifts asymptotic experiment" to get rid of the dependence of the limiting distribution on the exact distribution of the errors (and other regressors). Our results show that in the non-stationary trending case, there is no need to resort to such alternative asymptotic approximations.

Models I.a and I.b show further interesting differences. First, when the errors are $I(1)$, the limiting distribution is invariant to the location of the break. In contrast, when the errors are $I(0)$, the limiting distribution depends on the location of the break in such a way that (given other parameter values) the variance is smaller the closer the break is to the middle of the sample. In both cases, as expected, the variance decreases as the shift in slope increases. The simple limiting distributions obtained make it fairly easy to construct confidence intervals.

For Models II.a and II.b which incorporate a level shift, the limiting distributions obtained are strikingly different from those for Models I.a and I.b and are similar in structure to those obtained in the stationary case. Here, the limiting distributions are expressed as particular functions of a two sided random process involving many nuisance parameters. However, when the errors are $I(1)$, it does not depend on the exact distribution of these errors. Hence, confidence intervals can be computed using consistent estimates of the various nuisance parameters and simulations of the various functionals of the Weiner process to approximate the distribution. When the errors are $I(0)$, it is still possible to use simulations to compute the confidence intervals but the limiting distribution depends on the exact distribution of the errors making the result less attractive in practice. In this case, it is possible to get rid of the dependence of the limiting distribution on the exact distribution of the errors by adopting an asymptotic framework whereby the shift is shrinking as the sample size increases (see Bai et al., 1998).

Comparing the results for Models I and II (with either $I(1)$ or $I(0)$ errors), we find that the level shift plays an important role in the limiting distribution of the estimated break date. Suppose that the data generating process specifies no level shift, i.e. $\mu_b^0 = 0$. In Model I, no level shift is allowed in the regression while in Model II it is allowed via the regressor C_t. Our results show that introducing such an irrelevant regressor changes the rate of convergence of the estimated break date and its asymptotic distribution. We return to this important issue in a subsequent section with a detailed explanation.

Note that the level shift coefficient μ_b^0 does not enter into the limiting distribution of the estimated break date in Model II.a. The intuition for this feature is that as the sample size increases, the magnitude of the level shift, relative to the level of the trend function, becomes negligible and in the limit its effect is masked by the variation in the errors which are $I(1)$.[3] Hence, we can expect limiting results obtained from Model II.a to be adequate approximations in finite samples when the level shift μ_b^0 is relatively small and that the quality of the approximation would deteriorate as the level shift increases.[4] For breaking trend function with large level shifts and $I(1)$ errors, it would then seem more appropriate to use limiting results from Model III.a where the importance of the level shift does not vanish as the sample size increases. A problem, however, is that in this case it was not possible to establish a rate of convergence of the break fraction $\hat{\lambda}$. Accordingly, no limiting distribution is available to provide a confidence interval. Hence, in the case of Model III.a, to obtain an asymptotic approximation that is influenced by the level shift μ_b^0, we consider using an asymptotic expansion of the distribution obtained for Model II.a that retains higher order terms affected by the level shift μ_b^0 and the sample size T. This expansion, derived in the appendix is described in the following Theorem.

Theorem 5. *For Model II.a* ($I(1)$ *errors*), *using terms defined in Theorem* 4, *define a stochastic process* $V^*(n; T, \lambda^0, \mu_b^0, \beta_b^0, \sigma)$ *on the set of integers as follows*: $V^*(0) = 0, V^*(n) = V_1(n)$ *for* $n<0$ *and* $V^*(n) = V_2(n)$ *for* $n>0$, *with*

$$V_1(n; T, \lambda^0, \mu_b^0, \beta_b^0, \sigma) = \sum_{k=n+1}^{0} [\mu_b^0 + \beta_b^0 k]^2 + 2\sigma\xi_4 T^{1/2} \sum_{k=n+1}^{0} [\mu_b^0 + \beta_b^0 k] + n\sigma^2 T[2\xi_2'\Omega_1\xi_1 - \xi_1'\Omega_2\xi_1],$$

$$V_2(n; T, \lambda^0, \mu_b^0, \beta_b^0, \sigma) = \sum_{k=1}^{n} [\mu_b^0 + \beta_b^0 k]^2 + 2\sigma\xi_3 T^{1/2} \sum_{k=1}^{n} [\mu_b^0 + \beta_b^0 k] + n\sigma^2 T[2\xi_2'\Omega_1\xi_1 - \xi_1'\Omega_2\xi_1].$$

Then under Assumptions 2–3, *we have the following approximation to the finite sample distribution of* $\hat{T}_1$: $\hat{T}_1 - T_1^0 \sim \arg\min_n V^*(n; T, \lambda^0, \mu_b^0, \beta_b^0, \sigma)$.

Note that this approximation still does not depend on the finite sample distribution of the errors but only on various nuisance parameters which can be estimated. Hence, it is possible to simulate the confidence interval. It is important to note that now not only the slope change β_b^0 but also the level shift μ_b^0 affect the stated distribution. Simulations to be presented subsequently show that this expansion provides an excellent approximation to the finite sample distribution.

[3]In Model II.b, the level shift also becomes small relative to the overall level of the trend function but it still has an effect on the asymptotic distribution since the errors are $I(0)$ and do not mask its effect.

[4]These assertions are corroborated by simulation experiments reported in a later section of the paper.

3.4. The limit distribution of the other parameters

We now turn to the limiting distributions of the other parameter estimates involved in each model, namely, $(\hat{\mu}_1, \hat{\beta}_1, \hat{\beta}_b)$ and $\hat{\mu}_b$ for Models II, and III. A standard result in the stationary case (see, e.g., Bai, 1997 or Bai and Perron, 1998) is that the limiting distribution of the parameters of the model (other than the break date) is the same whether one uses the estimated break date or its true value. To investigate whether or not such an equivalence holds with trend and/or $I(1)$ errors, we derive the limiting distributions assuming first that the break date is estimated (by minimizing the sum of squared residuals) and then assuming it is fixed at some known true value, in which case the estimates are denoted by $(\bar{\mu}_1, \bar{\beta}_1, \bar{\beta}_b)$ and $\bar{\mu}_b$. The results, derived in the appendix, are summarized in the following theorem.

Theorem 6. *Under Assumptions* 1–3, *we have that*:

1(a). *In Model I.a, the limiting distribution of* $\hat{\gamma}$ (*using the estimated* $\hat{T}_1$) *is*

$$\begin{bmatrix} T^{-1/2}(\hat{\mu}_1 - \mu_1^0) \\ T^{1/2}(\hat{\beta}_1 - \beta_1^0) \\ T^{1/2}(\hat{\beta}_b - \beta_b^0) \end{bmatrix} \xrightarrow{d} \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \frac{2}{15}\lambda^0 & -\frac{1}{10} & \frac{1}{10} \\ -\frac{1}{10} & \frac{6}{5\lambda^0} & -\frac{6}{5\lambda^0} \\ \frac{1}{10} & -\frac{6}{5\lambda^0} & \frac{6}{5\lambda^0(1-\lambda^0)} \end{bmatrix} \right).$$

1(b). *In Model I.a, the limiting distribution of* $\bar{\gamma}$ (*assuming a known break date* T_1^0) *is*

$$\begin{bmatrix} T^{-1/2}(\bar{\mu}_1 - \mu_1^0) \\ T^{1/2}(\bar{\beta}_1 - \beta_1^0) \\ T^{1/2}(\bar{\beta}_b - \beta_b^0) \end{bmatrix}$$
$$\xrightarrow{d} \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \frac{1}{10}\lambda^0 + \frac{1}{30} & -\frac{1}{10}\frac{(\lambda^0)^2-\lambda^0+1}{\lambda^0} & \frac{1}{20}\frac{3\lambda^0-2}{\lambda^0(\lambda^0-1)} \\ -\frac{1}{10}\frac{(\lambda^0)^2-\lambda^0+1}{\lambda^0} & \frac{3}{10}\frac{2(\lambda^0)^2+\lambda^0+1}{(\lambda^0)^2} & -\frac{3}{20}\frac{2(\lambda^0)^2-\lambda^0-2}{(\lambda^0)^2(\lambda^0-1)} \\ \frac{1}{20}\frac{3\lambda^0-2}{\lambda^0(\lambda^0-1)} & -\frac{3}{20}\frac{2(\lambda^0)^2-\lambda^0-2}{(\lambda^0)^2(\lambda^0-1)} & \frac{3}{10(\lambda^0)^2(\lambda^0-1)^2} \end{bmatrix} \right).$$

2(a). *In Model I.b, the limiting distribution of* $\hat{\gamma}$ (*using the estimated* $\hat{T}_1$) *is*

$$\begin{bmatrix} T^{1/2}(\hat{\mu}_1 - \mu_1^0) \\ T^{3/2}(\hat{\beta}_1 - \beta_1^0) \\ T^{3/2}(\hat{\beta}_b - \beta_b^0) \end{bmatrix} \xrightarrow{d} \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \frac{4}{\lambda^0} & -\frac{6}{(\lambda^0)^2} & \frac{6}{(\lambda^0)^2} \\ -\frac{6}{(\lambda^0)^2} & \frac{12}{(\lambda^0)^3} & -\frac{12}{(\lambda^0)^3} \\ \frac{6}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3} & 12\,\frac{3(\lambda^0)^2-3\lambda^0+1}{(\lambda^0)^3(1-\lambda^0)^3} \end{bmatrix} \right).$$

2(b). *In Model I.b, the limiting distribution of* $\bar{\gamma}$ (*assuming a known break date* T_1^0) *is*

$$\begin{bmatrix} T^{1/2}(\bar{\mu}_1-\mu_1^0) \\ T^{3/2}(\bar{\beta}_1-\beta_1^0) \\ T^{3/2}(\bar{\beta}_b-\beta_b^0) \end{bmatrix} \xrightarrow{\mathrm{d}} \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \frac{\lambda^0+3}{\lambda^0} & -3\,\frac{1+\lambda^0}{(\lambda^0)^2} & \frac{3}{(\lambda^0)^2(1-\lambda^0)} \\ -3\,\frac{1+\lambda^0}{(\lambda^0)^2} & 3\,\frac{3\lambda^0+1}{(\lambda^0)^3} & 3\,\frac{2\lambda^0+1}{(\lambda^0)^3(\lambda^0-1)} \\ \frac{3}{(\lambda^0)^2(1-\lambda^0)} & 3\,\frac{2\lambda^0+1}{(\lambda^0)^3(\lambda^0-1)} & \frac{3}{(\lambda^0)^3(1-\lambda^0)^3} \end{bmatrix} \right).$$

3(a). *In Model II.a, the limiting distribution of* $\hat{\gamma}$ (*using the estimated* $\hat{T}_1$) *is*

$$\begin{bmatrix} T^{-1/2}(\hat{\mu}_1-\mu_1^0) \\ T^{1/2}(\hat{\beta}_1-\beta_1^0) \\ T^{-1/2}(\hat{\mu}_b-\mu_b^0)-\beta_b^0(\hat{T}_1-T_1^0)/\sqrt{T} \\ T^{1/2}(\hat{\beta}_b-\beta_b^0) \end{bmatrix} \Rightarrow \sigma\Omega_1 \begin{bmatrix} \int_0^1 W(r)\,\mathrm{d}r \\ \int_0^1 rW(r)\,\mathrm{d}r \\ \int_{\lambda^0}^1 W(r)\,\mathrm{d}r \\ \int_{\lambda^0}^1 (r-\lambda^0)W(r)\,\mathrm{d}r \end{bmatrix}.$$

This implies that the marginal limit distribution of $\hat{\mu}_1$, $\hat{\beta}_1$ *and* $\hat{\beta}_b$ *is the same as in Model I.a and that* $\hat{\mu}_b$ *is asymptotically unidentified since*

$$T^{-1/2}[(\hat{\mu}_b-\mu_b^0)-\beta_b^0(\hat{T}_1-T_1^0)] \Rightarrow \xi_3+\xi_4,$$

where ξ_3 *and* ξ_4 *are random variables defined in Theorem* 4.3.

3(b). *In Model II.a, the limiting distribution of* $\bar{\gamma}$ (*assuming a known break date* T_1^0) *is*

$$\begin{bmatrix} T^{-1/2}(\bar{\mu}_1-\mu_1^0) \\ T^{1/2}(\bar{\beta}_1-\beta_1^0) \\ T^{-1/2}(\bar{\mu}_b-\mu_b^0) \\ T^{1/2}(\bar{\beta}_b-\beta_b^0) \end{bmatrix} \xrightarrow{\mathrm{d}} \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \frac{2}{15}\lambda^0 & -\frac{1}{10} & -\frac{1}{30}\lambda^0 & \frac{1}{10} \\ -\frac{1}{10} & \frac{6}{5\lambda^0} & -\frac{1}{10} & -\frac{6}{5\lambda^0} \\ -\frac{1}{30}\lambda^0 & -\frac{1}{10} & \frac{2}{15} & 0 \\ \frac{1}{10} & -\frac{6}{5\lambda^0} & 0 & \frac{6}{5(1-\lambda^0)\lambda^0} \end{bmatrix} \right).$$

4(a). *For Model II.b, the marginal limiting distribution of* $\hat{\mu}_1$, $\hat{\beta}_1$ *and* $\hat{\beta}_b$ *is the same as in Model I.b. Also,* $\hat{\mu}_b$ *is asymptotically unidentified and* $\hat{\mu}_b-\mu_b^0 \Rightarrow \beta_b^0 m_{\mathrm{IV}}^{\infty}$, *with* m_{IV}^{∞}, *defined in Theorem* 4.4

4(b). *In Model II.b, the limiting distribution of* $\bar{\gamma}$ (*assuming a known break date* T_1^0) *is*

$$\begin{bmatrix} T^{1/2}(\bar{\mu}_1-\mu_1^0) \\ T^{3/2}(\bar{\beta}_1-\beta_1^0) \\ T^{1/2}(\bar{\mu}_b-\mu_b^0) \\ T^{3/2}(\bar{\beta}_b-\beta_b^0) \end{bmatrix} \xrightarrow{\mathrm{d}}$$

$$
N\left(\begin{bmatrix}0\\0\\0\\0\end{bmatrix}, \sigma^2\begin{bmatrix} \frac{4}{\lambda^0} & -\frac{6}{(\lambda^0)^2} & \frac{2}{\lambda^0} & \frac{6}{(\lambda^0)^2} \\ -\frac{6}{(\lambda^0)^2} & \frac{12}{(\lambda^0)^3} & -\frac{6}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3} \\ \frac{2}{\lambda^0} & -\frac{6}{(\lambda^0)^2} & -\frac{4}{(\lambda^0-1)\lambda^0} & -6\,\frac{-1+2\lambda^0}{(\lambda^0)^2(\lambda^0-1)^2} \\ \frac{6}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3} & -6\,\frac{-1+2\lambda^0}{(\lambda^0)^2(\lambda^0-1)^2} & 12\,\frac{3(\lambda^0)^2-3\lambda^0+1}{(\lambda^0)^3(1-\lambda^0)^3} \end{bmatrix}\right).
$$

5. *In Model III.a, the limiting distributions of* $\hat{\gamma}$ (*using the estimated* $\hat{T}_1$) *and* $\bar{\gamma}$ (*assuming a known break date* T_1^0) *are the same and given by*

$$
\begin{bmatrix} T^{-1/2}(\hat{\mu}_1-\mu_1^0) \\ T^{1/2}(\hat{\beta}_1-\beta_1^0) \\ T^{-1/2}(\hat{\mu}_b-\mu_b^0) \\ T^{1/2}(\hat{\beta}_b-\beta_b^0) \end{bmatrix} \to N\left(\begin{bmatrix}0\\0\\0\\0\end{bmatrix}, \sigma^2\begin{bmatrix} \frac{2}{15}\lambda^0 & -\frac{1}{10} & -\frac{2}{15}\lambda^0 & \frac{1}{10} \\ -\frac{1}{10} & \frac{6}{5\lambda^0} & \frac{11}{10} & -\frac{6}{5\lambda^0} \\ -\frac{2}{15}\lambda^0 & \frac{11}{10} & \frac{2}{15}\frac{8\lambda^0+1}{1-\lambda^0} & \frac{6}{5(\lambda^0-1)} \\ \frac{1}{10} & -\frac{6}{5\lambda^0} & \frac{6}{5(\lambda^0-1)} & \frac{6}{5\lambda^0(1-\lambda^0)} \end{bmatrix}\right).
$$

6. *In Model III.b, the limiting distributions of* $\hat{\gamma}$ (*using the estimated* $\hat{T}_1$) *and* $\bar{\gamma}$ (*assuming a known break date* T_1^0) *are the same and given by*

$$
\begin{bmatrix} T^{1/2}(\hat{\mu}_1-\mu_1^0) \\ T^{3/2}(\hat{\beta}_1-\beta_1^0) \\ T^{1/2}(\hat{\mu}_b-\mu_b^0) \\ T^{3/2}(\hat{\beta}_b-\beta_b^0) \end{bmatrix}
$$

$$
\xrightarrow{d} N\left(\begin{bmatrix}0\\0\\0\\0\end{bmatrix}, \sigma^2\begin{bmatrix} \frac{4}{\lambda^0} & -\frac{6}{(\lambda^0)^2} & -\frac{4}{\lambda^0} & \frac{6}{(\lambda^0)^2} \\ -\frac{6}{(\lambda^0)^2} & \frac{12}{(\lambda^0)^3} & \frac{6}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3} \\ -\frac{4}{\lambda^0} & \frac{6}{(\lambda^0)^2} & 4\,\frac{4(\lambda^0)^2-2\lambda^0+1}{\lambda^0(1-\lambda^0)^3} & -6\,\frac{4(\lambda^0)^2-3\lambda^0+1}{(\lambda^0)^2(1-\lambda^0)^3} \\ \frac{6}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3} & -6\,\frac{4(\lambda^0)^2-3\lambda^0+1}{(\lambda^0)^2(1-\lambda^0)^3} & 12\,\frac{3(\lambda^0)^2-3\lambda^0+1}{(\lambda^0)^3(1-\lambda^0)^3} \end{bmatrix}\right).
$$

Remark 9. Consider the generalized Model II.a ($I(1)$ errors) with $\mu_b^0 = \kappa T^{\alpha}$. When $\alpha < 1/2$, the result is the same as for Model II.a; when $\alpha > 1/2$, they are the same as for Model III.a. When the errors are $I(0)$, things are different and the results for Model III.b applies for all $\alpha > 0$.

Interestingly, except for the unidentified intercept shift μ_b, the other parameters, $\hat{\mu}_1$, $\hat{\beta}_1$ and $\hat{\beta}_b$, share the same limiting distribution in Models I.a, II.a and III.a. A similar feature holds across Models I.b, II.b and III.b. The important implication of this result is that if one is mainly interested in making asymptotic inference on these three parameters, the exact model specification does not matter (of course, except for the fact that we need the condition that $\mu_b = 0$ for Model I.a and I.b). However, the model specification does matter for asymptotic inference on the break date and the intercept shift.

As we discussed earlier, allowing for a level shift can possibly affect how good an estimate $\hat{T}_1$ is for the break date in Models II.a and II.b. On the other hand, we see the ''feedback'' effect of contaminated identification on the limiting distribution for $\hat{\mu}_b$. For example, in Model II.b, $\hat{\mu}_b$ is asymptotically linearly correlated with $\hat{T}_1$ in the sense that the limiting distribution of $\hat{\mu}_b$ is proportional to the limiting distribution of $\hat{T}_1$, the proportionality factor being the change in slope β_b^0. We provide some intuition for this result in the next section.

It is of some interest to note that it is possible to estimate $\hat{\mu}_1$ and $\hat{\beta}_1$ more precisely by estimating the break date rather than imposing a known fixed break date even if the latter correspond to the true break date. An example can be obtained by looking at the results for model I.a and comparing parts 1(a) and 1(b) of the Theorem. We have $var(\bar{\mu}_1) = (1/10)\lambda^0 + (1/30) > (2/15)\lambda^0 = var(\hat{\mu}_1)$ and $var(\bar{\beta}_1) - var(\hat{\beta}_1) = 3(1-\lambda^0)(2\lambda^0 - 1)/10(\lambda^0)^2)$, the sign of which depends on λ^0.

4. Simulation experiments

The aim of the simulation experiments we present is to provide answers to the following questions: (1) What are the important features of the distributions of the estimates?; (2) How adequate are the asymptotic distributions obtained as approximations to the finite sample distributions?; (3) For a time series with a given magnitude for the slope change and level shift, which model should be used to carry inference?

4.1. Finite sample versus asymptotics

In the first group of simulations (Figs. 2–6), we show that the limiting distributions derived are indeed very good approximations to the finite sample distributions. To that effect, we focus on the three estimates, $\hat{T}_1$ (break date), $\hat{\beta}_b$ (slope change) and $\hat{\mu}_b$ (level shift). All simulation experiments involve 2000 replications and we present results for the sample sizes $T = 200$ and 800.

To assess the quality of the approximation, we compare the asymptotic and finite sample probability density functions (pdf). To obtain the finite sample pdf, we use the 2000 simulated statistics and construct an empirical pdf using a non-parametric kernel density smoothing method.[5]The limiting pdf can be obtained directly in cases

[5]That is, for a given set of statistics, say $\{X_i\}_{i=1,\dots,N}$, the pdf at value x is estimated by $\tilde{f}(x) = (N \cdot h_x)^{-1}\sum_{i=1}^{N} K((x - X_i)/h_x)$ where $K(\cdot)$ is the kernel function and h_x is the bandwidth. In our case, $N = 2000$ and we use the standard normal distribution as the kernel function. Since the estimates of the break date are discrete integers, the cross-validation method for choosing the optimal bandwidth does not work well in this case. As a rule of thumb, we simply let $h_x = 0.3\hat{\sigma}_x$ where $\hat{\sigma}_x$ is the estimated standard deviation of a given sample of statistics $\{X_i\}_{i=1,\dots,N}$.

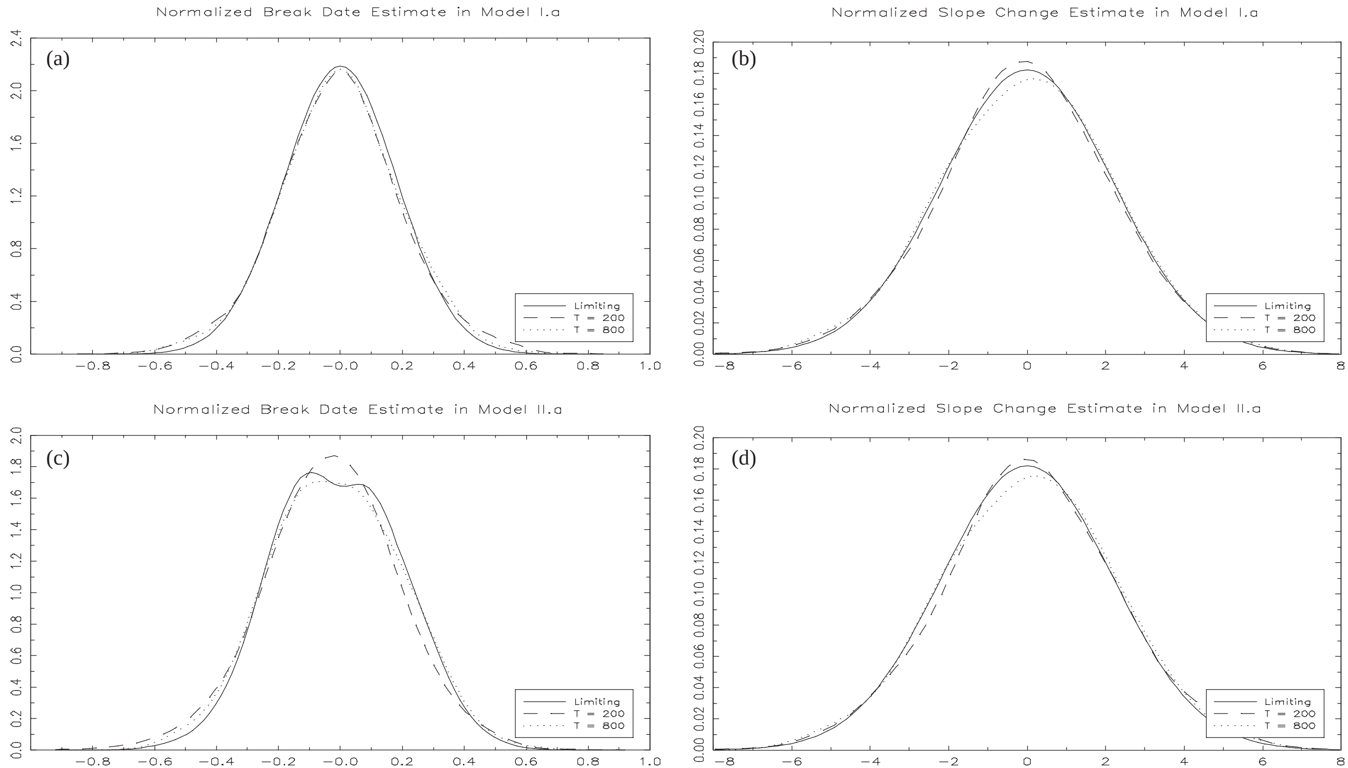

Fig. 2. Finite sample versus asymptotic distributions in models I.a and II.a: $\mu_b^0 = 0$. The graphs plot the empirical finite sample probability density functions (pdf) ($T = 200$ and 800) against the limiting pdf for the estimates of the break date, $\hat{T}_1$, and the slope change, $\hat{\beta}_b$, in Models I.a and II.a. The statistics are normalized as follows: $(\hat{T}_1 - T_1^0)/\sqrt{T}$ for the break date and $\sqrt{T}(\hat{\beta}_b - \beta_b^0)/\sigma$ for the slope break. The finite sample distributions are obtained using $u_t = \sum_{j=1}^{t} \varepsilon_j$ and $\varepsilon_j \sim \mathrm{N}(0, \sigma^2)$ with 2000 replications. The parameters of the models are set to $\lambda^0 = 0.5$, $\mu_1^0 = 1.72$, $\mu_b^0 = 0$, $\beta_1^0 = 0.03$, $\beta_b^0 = -0.02$, $\sigma = 0.01$. For the asymptotic distributions, Theorems 4.1 and 6.1(a) are used for Model I.a and Theorems 4.3 and 6.3(a) are used for Model II.a. When the limiting distributions are non-Normal, we use 5000 simulated values to construct the pdf.

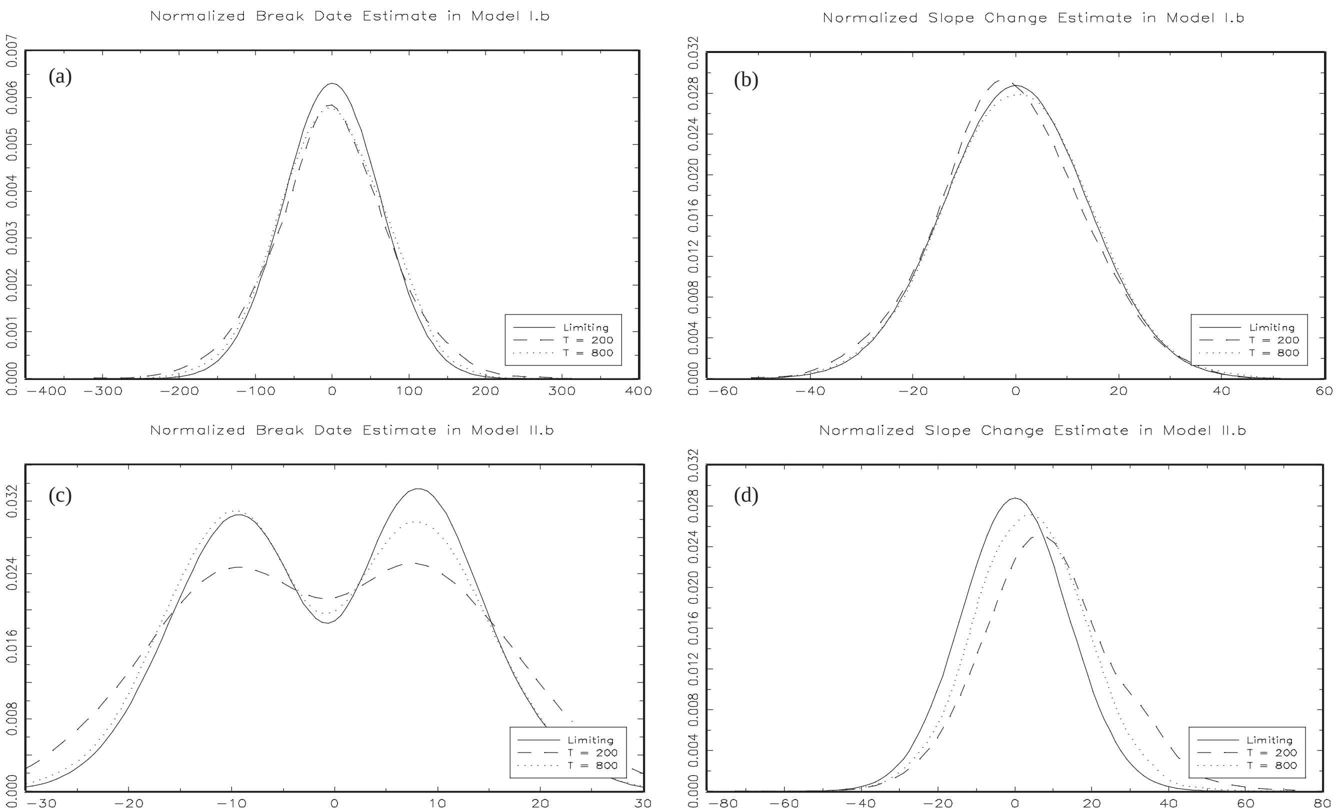

Fig. 3. Finite sample versus asymptotic distributions in Models I.b and II.b: $\mu_b^0 = 0$. The graphs plot the empirical finite sample probability density functions (pdf) ($T = 200$ and 800) against the limiting pdf for the estimates of the break date, $\hat{T}_1$, and the slope change, $\hat{\beta}_b$, in Models I.b and II.b. The statistics are normalized as follows: $(\hat{T}_1 - T_1^0)\sqrt{T}$ for the break date in Model I.b but $\hat{T}_1 - T_1^0$ in Model II.b; and $T^{3/2}(\hat{\beta}_b - \beta_b^0)/\sigma$ for the slope break in both models. The finite sample distributions are obtained using $u_t \sim \mathrm{N}(0, \sigma^2)$ with 2000 replications. The parameters of the models are set to $\lambda^0 = 0.5$, $\mu_1^0 = 1.72$, $\mu_b^0 = 0$, $\beta_1^0 = 0.03$, $\beta_b^0 = -0.02$, $\sigma^2 = 0.1$. For the asymptotic distributions, Theorems 4.2 and 6.2(a) are used for Model I.b and Theorems 4.4 and 6.4(a) are used for Model II.b. When the limiting distributions are non-normal, we use 5000 simulated values to construct the pdf.

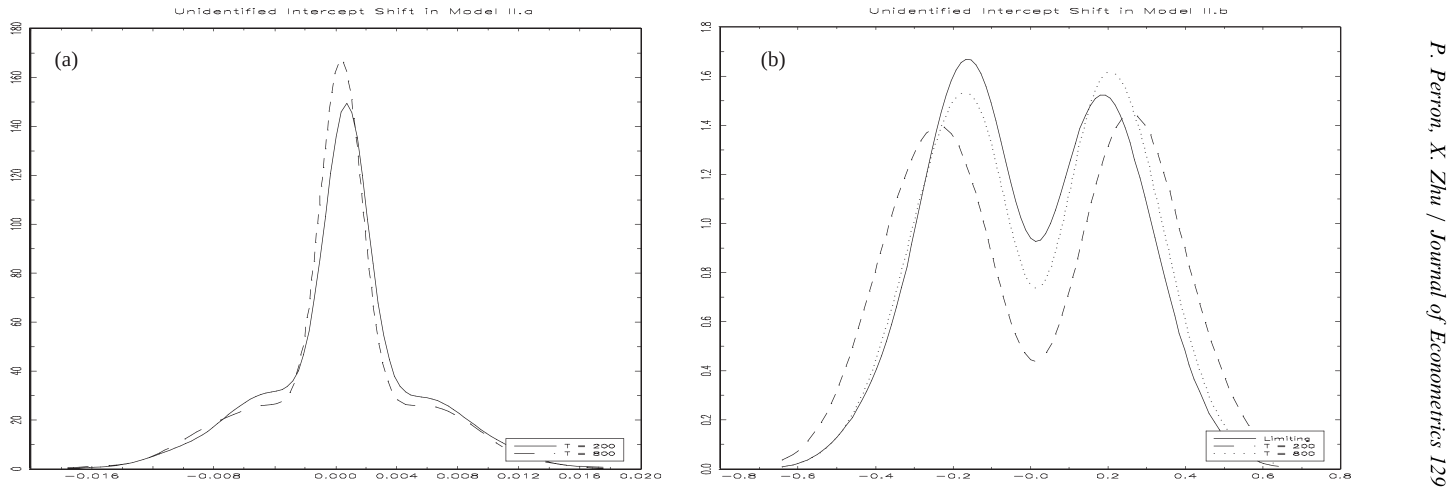

Fig. 4. Unidentified intercept shift in Model II.a and II.b: $\mu_b^0 = 0$. The graphs plot the empirical finite sample probability density functions (pdf) ($T = 200$ and 800) against the limiting pdf for the estimates of the unidentified intercept shift, $\hat{\mu}_b$, in Models II.a and II.b. The finite sample distributions are obtained using $u_t = \sum_{j=1}^{t} \varepsilon_j$ and $\varepsilon_j \sim \mathrm{N}(0, 0.01^2)$ with 2000 replications in Model II.a while they are obtained using $u_t \sim \mathrm{N}(0, 0.1)$ in Model II.b. The parameters of the models are set to $\lambda^0 = 0.5$, $\mu_1^0 = 1.72$, $\mu_b^0 = 0$, $\beta_1^0 = 0.03$, $\beta_b^0 = -0.02$. Panel (a) plots $(\hat{\mu}_b - \mu_b^0)/\sqrt{T}$ in Model II.a for $T = 200$ and 800; Panel (b) plots $\hat{\mu}_b - \mu_b^0$ in Model II.b for $T = 200$ and 800 as well as the limiting distribution $\hat{\mu}_b - \mu_b^0 \xrightarrow{d} \beta_b^0 m_{\mathrm{IV}}^{\infty}$. We use 5000 simulated values to construct the pdf of m_{IV}^{∞} according to Theorem 4.4.

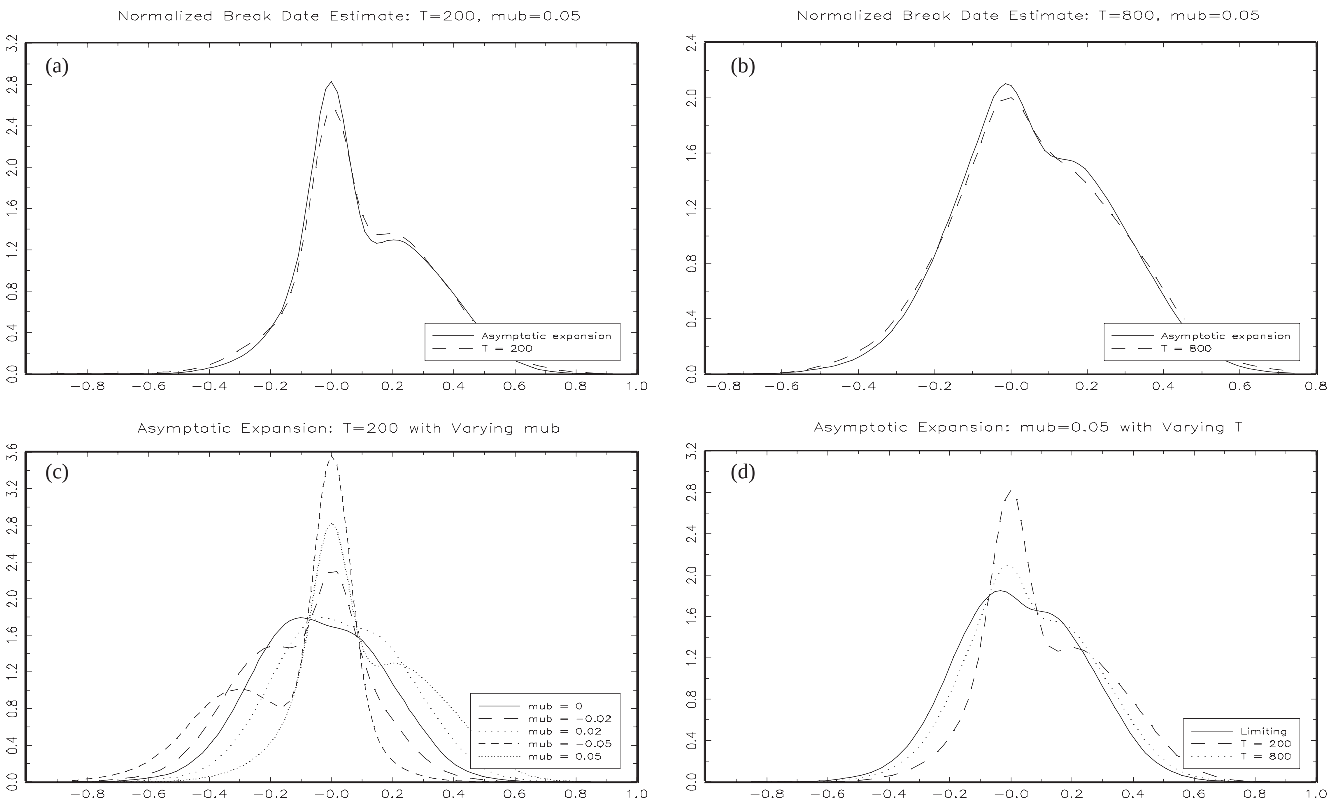

Fig. 5. Finite sample versus expanded asymptotic distributions in Model II.a: $\mu_b^0 \neq 0$. In these graphs, we consider the effect of a non-zero intercept shift on the empirical finite sample distributions of the estimated break date in Model II.a. And we want to examine how well the expanded asymptotic distributions approximate the finite sample distributions in this case. The finite sample distributions are obtained using $u_t = \sum_{j=1}^{t} \varepsilon_j$ and $\varepsilon_j \sim \mathrm{N}(0, \sigma^2)$ with 2000 replications. The parameters of the models are set to $\lambda^0 = 0.5$, $\mu_1^0 = 1.72$, $\beta_1^0 = 0.03$, $\beta_b^0 = -0.02$, $\sigma = 0.01$. We use 5000 simulated values to construct the pdf of the expanded asymptotic distribution of $\hat{T}_1$ according to Theorem 5. In Panel (a), we compare the empirical finite sample pdf for $(\hat{T}_1 - T_1^0)/\sqrt{T}$ against the pdf of the asymptotic expansion in the case where $T = 200$ and $\mu_b^0 = 0.05$; in panel (b), we consider the case where $T = 800$ and $\mu_b^0 = 0.05$; in panel (c), we only examine the pdfs of the asymptotic expansion for fixed sample size $T = 200$ but varying μ_b^0; while in panel (d), we fix $\mu_b^0 = 0.05$ but vary the sample size.

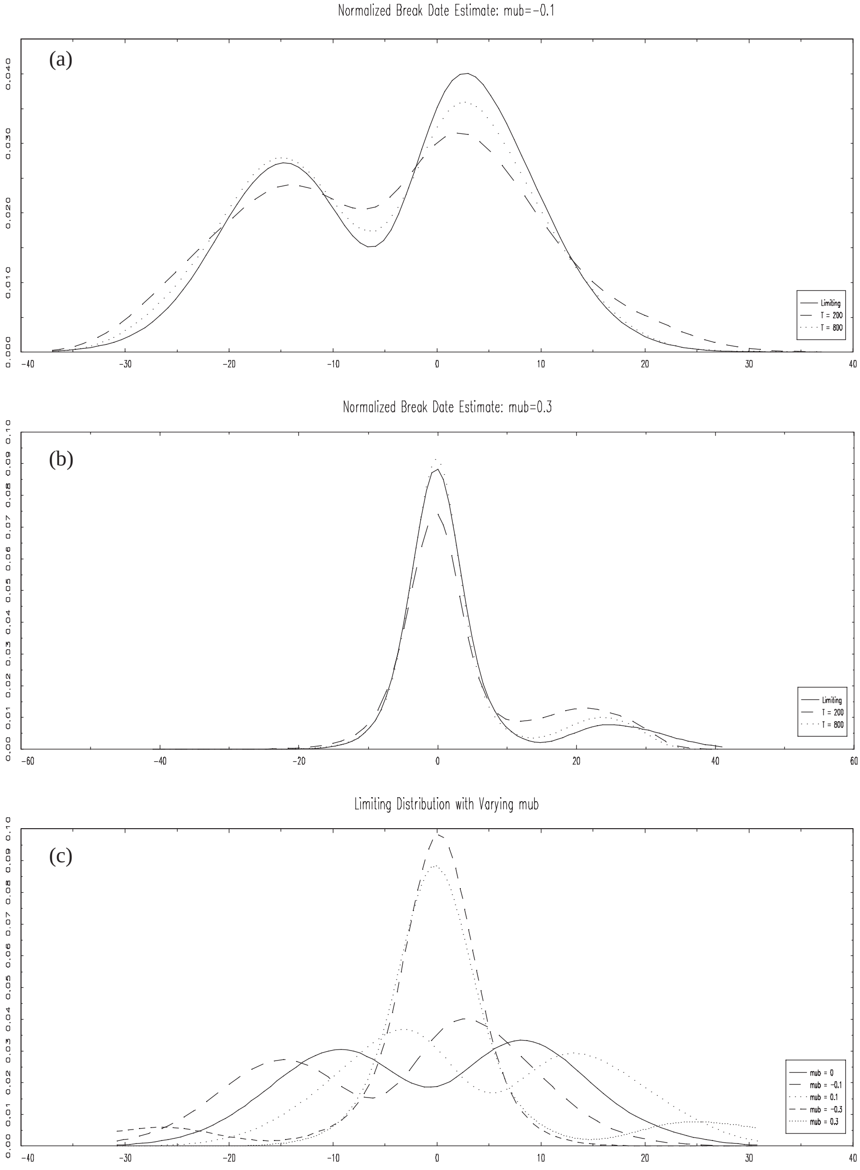

Fig. 6. Finite sample versus asymptotic distributions in Model II.b: $\mu_b^0 \neq 0$. In these graphs, we consider the effect of a non-zero intercept shift on the empirical finite sample distributions of the estimated break date in Model II.b. And we want to examine how well the asymptotic distributions approximate the finite sample distributions in this case. The statistics for the break date is normalized as $\hat{T}_1 - T_1^0$. The finite sample distributions are obtained using $u_t \sim \mathrm{N}(0, \sigma^2)$ with 2000 replications. The parameters of the models are set to $\lambda^0 = 0.5$, $\mu_1^0 = 1.72$, $\beta_1^0 = 0.03$, $\beta_b^0 = -0.02$, $\sigma^2 = 0.1$. We use 5000 simulated values to construct the pdf of the asymptotic distribution of $\hat{T}_1$ according to Theorem 4.4. In Panel (a), where $\mu_b^0 = -0.1$, we compare the finite sample distributions for $T = 200$ and 800 against the limiting distribution; in Panel (b), we let $\mu_b^0 = 0.3$; in Panel (c), we compare the limiting distributions for $\hat{T}_1 - T_1^0$ with varying μ_b^0.

where the limiting distribution is Normal. For the other cases, it can be obtained by simulation and a similar kernel smoothing method ($\hat{T}_1$ in Models II.a and II.b, for examples).

We first consider the case where the DGP is given by

$$y_t = \mu_1^0 + \beta_1^0 t + \beta_b^0 B_t + u_t, \tag{5}$$

where $u_t = \sum_{j=1}^t e_j$ with $e_j \sim$ i.i.d. $\mathrm{N}(0, \sigma^2)$, i.e. the sequence u_t is a random walk. We recall that $B_t = 1(t > T_1^0)(t - T_1^0)$ with $T_1^0 = \lambda^0 T$, hence, the series y_t is characterized by a joined segmented trend with no level shift at the time of the break. We set the various parameters at the following values: $\lambda^0 = 0.5$, $\mu_1^0 = 1.72$, $\beta_1^0 = 0.03$, $\beta_b^0 = -0.02$ and $\sigma = 0.01$.

Fig. 2 presents the finite sample and asymptotic pdf of the normalized estimate of the break date $(\hat{T}_1 - T_1^0)/\sqrt{T}$ and the normalized estimate of the slope change $\sqrt{T}(\hat{\beta}_b - \beta_b^0)/\sigma$ when these are obtained from the regression corresponding to model I.a, i.e.

$$y_t = \hat{\mu}_1 + \hat{\beta}_1 t + \hat{\beta}_b B_t + \hat{u}_t \tag{6}$$

and from the regression corresponding to Model II.a, i.e.

$$y_t = \hat{\mu}_1 + \hat{\mu}_b C_t + \hat{\beta}_1 t + \hat{\beta}_b B_t + \hat{u}_t, \tag{7}$$

where we recall that $C_t = 1(t > T_1)$. The results show that the asymptotic distribution provides a good approximation when the regression from Model I.a is used (which is well specified). When the regression from Model II.a is used (which incorporates an unnecessary level shift regressor), the approximation for the estimate of the slope change is still very good. However, for the estimate of the break date, the limiting distribution exhibits a slight bimodal pattern which is not present in the finite sample distribution. Nevertheless, for $T = 800$, the approximation is quite satisfactory.

Fig. 3 presents a similar set of results but now the errors are $I(0)$, i.e., the DGP is still (5) but with $u_t \sim$ i.i.d. $\mathrm{N}(0, \sigma^2)$ and $\sigma^2 = 0.1$. The asymptotic distributions used are those corresponding to Models I.b and II.b. When the estimates are constructed from (6), see the top panel, the finite sample distributions of $T^{1/2}(\hat{T}_1 - T_1^0)$ and $T^{3/2}(\hat{\beta}_b - \beta_b^0)/\sigma$ are close to that of a Normal and indeed well approximated by the asymptotic distributions for both estimates. The results, in the bottom panel are for estimates constructed from the regression (7) using now $(\hat{T}_1 - T_1^0)$ with no scaling. Here, the results are strikingly different. For the estimate of the break date, the distribution is clearly bimodal. The asymptotic approximation is good when $T = 800$ but less so when $T = 200$ (though the same qualitative shape emerges). For the slope change, the finite sample distribution is skewed to the right, but again when $T = 800$, it is close to the asymptotic distribution.

Fig. 4 presents results pertaining to the distribution of the estimate of the level shift $\hat{\mu}_b$. The DGP is given by (5) where $u_t \sim I(1)$ in the left panel, i.e. $u_t = \sum_{j=1}^t e_j$ with $e_j \sim$ i.i.d. $\mathrm{N}(0, \sigma^2)$ (corresponding to Model II.a) and $u_t \sim I(0)$ in the right panel, i.e. $u_t \sim$ i.i.d. $\mathrm{N}(0, \sigma^2)$ (corresponding to Model II.b). The slope change is set to $\beta_b^0 = -0.02$, and $\sigma = 0.01$ when the errors are $I(1)$ and $\sigma = 0.1$ when the errors are $I(0)$ (the value of the other parameters are as stated above). The regression used is

(7). When the errors are $I(1)$ (left panel) we only plot the finite sample distributions which show little changes between $T = 200$ and 800. In the right panel, with $I(0)$ errors, the distributions are clearly bimodal and the asymptotic distribution is a good approximation. This bimodal feature parallels that found for the estimate of the break, a feature we explain in more detail below.

Figs. 5 and 6 consider the case where the DGP specifies a non-zero level shift in, i.e.

$$y_t = \mu_1^0 + \mu_b^0 C_t + \beta_1^0 t + \beta_b^0 B_t + u_t \tag{8}$$

and the regression used is (7). In Fig. 5, the errors are $I(1)$, i.e. $u_t = \sum_{j=1}^{t} e_j$ with $e_j \sim$ i.i.d. $\mathrm{N}(0, \sigma^2)$ (corresponding to Model II.a) with the slope change set to $\beta_b^0 = -0.02$ and $\sigma = 0.01$. The aim is to assess the extent to which the asymptotic expansion given by Theorem 5 provides a better (and adequate) approximation to the finite sample distribution of the estimated break date compared to the standard limiting distribution obtained in Theorem 4 (part 3). The base case uses the value $\mu_b^0 = 0.05$. Panels (a) and (b) show that this expansion provides a very good approximation to the finite sample distribution for both $T = 200$ and 800, even though the distributions change substantially when changing the sample size. Panel (d) compares the asymptotic expansion with the standard asymptotic distribution (from Theorem 4.3). The results show that the latter provides a poor approximation even when $T = 800$. The results in Panels (a), (b) and (d) show how the expansion of Theorem 5 provides a much better approximation than the standard limiting distribution. Finally, panel (c) shows how, for a given sample size set at $T = 200$, the distribution provided by the expansion is substantially affected by changes in the level shift coefficient μ_b^0. Since the standard asymptotic distribution is not affected by this parameter, this is further evidence of the usefulness of the asymptotic expansion.

Fig. 6 presents similar results for the case where the errors are $I(0)$, i.e. $u_t \sim$ i.i.d. $\mathrm{N}(0, \sigma^2)$ (corresponding to Model II.b). Here, we compare the finite sample distribution of the estimated break date with its asymptotic counterpart stated in Theorem 4 (part 4). Panel (a) considers the value $\mu_b^0 = -0.1$, in which case the results show strong bimodality (the right mode being more important) and that the asymptotic distribution is a good approximation. When $\mu_b^0 = 0.3$, the asymptotic approximation is even better and the left mode now clearly dominates. Panel (c) shows the behavior of the asymptotic distribution for different values of the level shift μ_b^0. What transpires from the results is that when $|\mu_b^0|$ is small, the distribution is bimodal with two modes that are important (and more symmetric as μ_b^0 gets closer to 0). When μ_b^0 increases (positive values), the left mode becomes more important and is more centered around 0 as μ_b^0 increases. The effect is opposite when μ_b^0 decreases (negative values), the right mode becoming more important. We provide some explanations for this feature below. Note that when the level shift is very big (in absolute value), the distribution is more centered around the true value (the second mode becoming negligible). Hence, large level shifts help to better identify the true break date.

4.2. Model selection

A feature of importance that transpires from the theory and the simulations reported above is that the inclusion of a level shift regressor has a substantial effect on the properties of the estimate of the break date as well as other parameters. In the next set of simulations, our aim is to see in which cases it is advisable to use a regression with or without a level shift.

The data generating process is again given by (8) with the parameter values $\lambda^0 = 0.5$, $\mu_1^0 = 1.72$, $\beta_1^0 = 0.03$, $\beta_b^0 = -0.02$ and $\sigma = 0.01$. Table 1 considers the case with $I(1)$ errors while Table 2 the case with $I(0)$ errors (in which case we use $\sigma^2 = 0.1$). We use 10 different values of μ_b^0 ranging from -0.3 to 0.5. We report the mean squared

Table 1
Simulation analysis of mean squared errors using Model I.a and II.a

Sample size	Intercept shift	Model I.a				Model II.a				
		$\hat{T}_1$	$\hat{\mu}_1$	$\hat{\beta}_1$	$\hat{\beta}_b$	$\hat{T}_1$	$\hat{\mu}_1$	$\hat{\beta}_1$	$\hat{\mu}_b$	$\hat{\beta}_b$
$T = 200$	$\mu_b^0 = -0.3$	136.8	12.2	13.1	32.3	0.0	13.8	11.5	25.5	22.7
	$\mu_b^0 = -0.1$	27.6	13.2	12.2	23.0	8.2	13.3	12.1	69.6	23.3
	$\mu_b^0 = -0.05$	13.0	13.4	12.0	22.7	10.8	13.3	12.2	69.9	23.3
	$\mu_b^0 = -0.02$	8.5	13.5	11.9	22.7	10.0	13.3	12.1	54.3	23.4
	$\mu_b^0 = 0$	7.7	13.6	11.8	22.7	9.1	13.4	12.0	44.3	23.3
	$\mu_b^0 = 0.02$	8.6	13.8	11.6	22.7	9.6	13.5	11.9	46.5	23.3
	$\mu_b^0 = 0.05$	13.4	14.0	11.4	22.7	10.8	13.7	11.7	62.7	23.4
	$\mu_b^0 = 0.1$	28.8	14.5	11.1	22.8	10.4	13.8	11.6	81.6	23.3
	$\mu_b^0 = 0.3$	145.5	24.0	16.9	30.3	0.0	13.8	11.5	25.5	22.7
	$\mu_b^0 = 0.5$	289.2	65.9	47.8	62.1	0.0	13.8	11.5	25.5	22.7
$T = 800$	$\mu_b^0 = -0.3$	222.9	48.3	3.1	5.9	20.5	49.6	3.0	198.0	5.8
	$\mu_b^0 = -0.1$	51.8	49.4	3.0	5.8	43.2	49.1	3.1	267.2	5.9
	$\mu_b^0 = -0.05$	34.1	49.7	3.0	5.8	38.5	49.3	3.1	210.7	5.9
	$\mu_b^0 = -0.02$	29.2	50.0	3.0	5.8	36.0	49.5	3.0	180.2	5.9
	$\mu_b^0 = 0$	28.2	50.1	3.0	5.8	35.0	49.6	3.0	169.6	5.9
	$\mu_b^0 = 0.02$	29.1	50.3	3.0	5.8	35.0	49.7	3.0	172.4	5.9
	$\mu_b^0 = 0.05$	34.1	50.4	3.0	5.8	37.0	49.9	3.0	191.8	5.9
	$\mu_b^0 = 0.1$	51.8	50.8	3.0	5.8	43.2	50.2	3.0	253.6	5.8
	$\mu_b^0 = 0.3$	224.3	52.7	2.9	5.8	21.4	50.4	3.0	233.1	5.8
	$\mu_b^0 = 0.5$	526.7	58.5	3.1	6.1	2.0	50.4	3.0	113.6	5.8

In this simulation study, we examine the model selection problem by considering the mean squared errors (MSE) of different estimates in Model I.a and Model II.a for different intercept shifts. The underlying true parameters used in simulation are: $\lambda^0 = 0.5$, $\mu_1^0 = 1.72$, $\beta_1^0 = 0.03$, $\beta_b^0 = -0.02$, $\sigma = 0.01$ and replication number 2000. Two sample sizes are considered, $T = 200$ and 800. The data is generated by $y_t = \mu_1^0 + \beta_1^0 t + \mu_b^0 C_t^0 + \beta_b^0 B_t^0 + u_t$ where $u_t = \sum_{s=1}^{t} e_s$ and $e_s \sim N(0, \sigma^2)$. For Model I.a, the estimates are obtained from the least squares regression $y_t = \hat{\mu}_1 + \hat{\beta}_1 t + \hat{\beta}_b \hat{B}_t + \hat{u}_t$, and for Model II.a they are obtained from the regression $y_t = \hat{\mu}_1 + \hat{\beta}_1 t + \hat{\mu}_b \hat{C}_t + \hat{\beta}_b \hat{B}_t + \hat{u}_t$. We normalize the MSE of $\hat{\beta}_1$ and $\hat{\beta}_b$ by multiplying 10^7, and the MSE of μ_1^0 and μ_b^0 by multiplying 10^4.

Table 2
Simulation analysis of mean squared errors using Model I.b and II.b

Sample size	Intercept shift	Model I.b				Model II.b				
		$\hat{T}_1$	$\hat{\mu}_1$	$\hat{\beta}_1$	$\hat{\beta}_b$	$\hat{T}_1$	$\hat{\mu}_1$	$\hat{\beta}_1$	$\hat{\mu}_b$	$\hat{\beta}_b$
$T=200$	$\mu_b^0=-0.3$	156.6	45.9	180.5	370.7	100.4	47.4	198.4	733.4	301.8
	$\mu_b^0=-0.1$	44.7	43.1	149.0	243.5	171.9	50.5	225.7	896.2	374.5
	$\mu_b^0=-0.05$	31.7	42.7	144.0	235.8	174.6	50.1	220.2	847.5	403.3
	$\mu_b^0=-0.02$	27.4	42.3	138.2	233.8	171.3	49.3	209.5	819.8	409.3
	$\mu_b^0=0$	26.8	42.1	135.3	233.7	173.6	48.8	203.5	816.0	415.0
	$\mu_b^0=0.02$	28.7	41.7	132.5	234.1	172.9	48.1	195.7	807.7	414.6
	$\mu_b^0=0.05$	33.3	41.0	125.4	235.8	170.2	47.6	187.5	807.8	410.2
	$\mu_b^0=0.1$	48.3	39.8	114.0	241.9	170.9	46.4	172.6	855.3	389.4
	$\mu_b^0=0.3$	165.7	45.7	156.9	351.4	115.8	43.7	145.0	798.4	321.0
	$\mu_b^0=0.5$	319.6	84.6	459.3	737.6	19.7	42.3	132.4	203.5	301.0
$T=800$	$\mu_b^0=-0.3$	198.7	10.5	2.1	4.2	68.7	10.3	2.0	334.0	3.8
	$\mu_b^0=-0.1$	28.9	10.2	1.9	3.6	152.4	10.6	2.2	658.6	4.0
	$\mu_b^0=-0.05$	11.4	10.2	1.9	3.6	142.0	10.6	2.2	602.1	4.1
	$\mu_b^0=-0.02$	6.4	10.2	1.9	3.6	136.5	10.6	2.1	575.4	4.1
	$\mu_b^0=0$	5.5	10.2	1.9	3.6	134.9	10.5	2.1	570.8	4.1
	$\mu_b^0=0.02$	6.4	10.2	1.9	3.6	132.7	10.5	2.1	560.9	4.1
	$\mu_b^0=0.05$	11.5	10.1	1.8	3.6	137.0	10.4	2.1	578.8	4.1
	$\mu_b^0=0.1$	29.4	10.0	1.8	3.6	145.2	10.4	2.1	617.3	4.1
	$\mu_b^0=0.3$	201.9	10.4	2.1	4.1	84.8	10.2	1.9	397.2	3.8
	$\mu_b^0=0.5$	504.2	15.6	5.0	7.1	7.8	10.1	1.9	53.1	3.8

In this simulation study, we examine the model selection problem by considering the mean squared errors (MSE) of different estimates in Model I.b and Model II.b for different intercept shifts. The underlying true parameters used in simulation are: $\lambda^0=0.5$, $\mu_1^0=1.72$, $\beta_1^0=0.03$, $\beta_b^0=-0.02$, $\sigma^2=0.1$ and replication number 2000. Two sample sizes are considered, $T=200$ and 800. The data is generated by $y_t=\mu_1^0+\beta_1^0 t+\mu_b^0 C_t^0+\beta_b^0 B_t^0+u_t$ where $u_t\sim \mathrm{N}(0,\sigma^2)$. For Model I.b, the estimates are obtained from the least squares regression $y_t=\hat{\mu}_1+\hat{\beta}_1 t+\hat{\beta}_b\hat{B}_t+\hat{u}_t$, and for Model II.b they are obtained from the regression $y_t=\hat{\mu}_1+\hat{\beta}_1 t+\hat{\mu}_b\hat{C}_t+\hat{\beta}_b\hat{B}_t+\hat{u}_t$. We normalize the MSE of $\hat{\beta}_1$ and $\hat{\beta}_b$ by multiplying 10^8, and the MSE of μ_1^0 and μ_b^0 by multiplying 10^4.

errors (MSE) of the parameter estimates obtained from regression (6) corresponding to Model I and regression (7) corresponding to Model II. What transpires from the result is that when the true level shift is small, it is better to exclude the level shift regressor. The MSE of the parameter estimates are then reduced. The reduction is especially pronounced when the errors are $I(0)$ to the extent that one is better off including the level shift regressor only for large values of μ_b^0, say greater than 0.3 in absolute value. The practical implementation of the choice of an appropriate model is an open question. The difficulty lies in the fact that with an estimated break date, the estimate $\hat{\mu}_b$ is not asymptotically identified. Hence, one cannot construct a valid pre-test procedure or use the confidence interval of $\hat{\mu}_b$ to map a confidence interval for the estimated break date. Visual inspection should be a useful starting point.

5. Contamination and "feedback" effects between level shifts and estimated break dates

Our theoretical and simulation results showed some surprising features related to the effect of the inclusion of a level shift regressor (whether the true DGP specifies such a level shift or not) on the estimate of the break date. In particular, including such a level shift regressor reduces the rate of convergence of the estimated break fraction when the errors are $I(0)$ and induces bimodality in the distribution of this estimate (both in finite samples and asymptotically). In this section, we provide intuitive explanations for these features.

Consider the simplified example depicted in Fig. 7. Suppose the break occurs at T_1^0 and there is no level shift at the break date. In the estimation, we allow the possibility of a level shift by incorporating the regressor C_t. Suppose that, by random chance, the realization of $y_{T_1^0-1}$ is the data point b instead of the data point a while all other data points are on the trend lines. Then the estimated break date (obtained by minimizing the sum of squared residuals) will be T_1^0-1 instead of T_1^0 and the associated estimate of the intercept shift at the estimated break date is the distance between a and b. Note that the same argument holds on the other side, i.e. with the realization at the data point d instead of e and all other points on the trend lines, one would estimate the break date at T_1^0+1. Hence, including a level shift regressor induces a kind of "contamination effect" with respect to how precisely we estimate the break date. In a sense, the point of intersection of the segmented trend becomes random since the level shift regressor can accommodate random departures from the

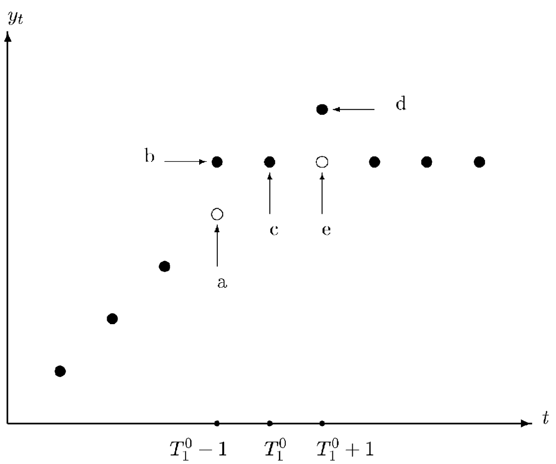

Fig. 7. Identification: contamination and feedback effects.

trend lines around the true break date. This effect remains even in the presence of small level shifts, i.e. small enough relative to the magnitude of the random deviations around the true trend lines. It also remains if the sample size is large and, hence, induces a reduction in the rate of convergence of the estimated break fraction. Moreover, using our simplified example, it is easy to see the following relation between the estimate of the level shift, the slope change and the estimated break date, namely that $\hat{\mu}_b - \mu_b^0 = \beta_b^0(\hat{T}_1 - T_1^0)$. This is similar to the limiting result derived in Theorem 6 (part 4.a). Hence the "contaminated" estimate $\hat{T}_1$ also influences the estimate $\hat{\mu}_b$. We call this a "feedback" effect. This "feedback" effect makes μ_b^0 a parameter that cannot be identified.

Following the same line of argument, it is easy to understand why the distribution of the estimate of the break date is bimodal when the true level shift is small. Indeed, with purely random deviations around the true trend function, we have less chances of estimating a break at the true value and more at either sides since the level shift regressor can categorize random deviations as level shifts with an estimated break date before or after the true one. By not incorporating a level shift regressor such a "contamination effect" disappears and it becomes easier to estimate the break date more precisely, hence the faster rate of convergence. If the true level shift is positive but not too large, the horizontal trend line in Fig. 7 shifts up so that there are less chances that a level shift regressor can accommodate a random departure from the trend line on the left side of the true break date. Hence, the left mode becomes less important (and vice versa with a negative level shift). As the level shift becomes very large, random deviations around the trend lines become negligible compared to the magnitude of the level shift. Hence, the level shift regressor will actually then reflect the true shift that occurs. The "contamination effect" disappears and the break date becomes easier to identify. The situation then corresponds to Model III where the rate of convergence is now faster.

6. Empirical applications

We now return to the (log) real per capita GDP series discussed in the introduction. We present estimates and confidence intervals for the break dates (T_1) and the slope change (β_b) for both cases where a level shift regressors is included or not. To assess whether to use the results corresponding to $I(0)$ or $I(1)$ errors we categorize the error structures for different countries according to the unit root tests reported in Perron (1992). Hence, the noise component is treated as $I(0)$ for Australia, Canada, Denmark, France, Germany, the United Kingdom and the United States. It is treated as $I(1)$ for Italy, Norway and Sweden.

The estimation results are reported in Table 3. The results under the heading "Model I" are obtained using the regression $y_t = \hat{\mu}_1 + \hat{\beta}_1 t + \hat{\beta}_b B_t + \hat{u}_t$ and the results under the heading "Model II" are obtained from the regression $y_t = \hat{\mu}_1 + \hat{\mu}_b C_t + \hat{\beta}_1 t + \hat{\beta}_b B_t + \hat{u}_t$. The confidence intervals for the estimates from Model I are simply computed from the normal asymptotic distribution stated in Theorem 4 (1 and 2). For Model II, we use 5000 replications (using the expansion of Theorem 5 for

Table 3
Empirical results for the log real per capita GDP series (1870–1986)

Break date, $\hat{T}_1$

Country	Error structure	Model I			Model II		
		Break date	90% Confidence interval		Break date	90% Confidence interval	
Australia	$I(0)$	1934	1930	1938	1929	1926	1943
Canada		1936	1924	1948	1930	1928	1951
Denmark		1948	1942	1954	1939	1938	1949
France		1950	1945	1955	1943	1943	1948
Germany		1948	1943	1953	1945	1941	1964
UK		1934	1928	1940	1919	1919	1919
USA		1934	1893	1975	1940	1927	1952
Italy	$I(1)$	1947	1934	1960	1943	1932	1975
Norway		1925	1913	1937	1925	1914	1944
Sweden		1920	1896	1944	1916	1902	1959

Slope change, $\hat{\beta}_b$

Country	Error structure	Model I			Model II		
		Slope change	95% Confidence interval		Slope change	95% Confidence interval	
Australia	$I(0)$	0.0203	0.0174	0.0233	0.0199	0.0170	0.0229
Canada		0.0137	0.0076	0.0199	0.0133	0.0084	0.0181
Denmark		0.0146	0.0108	0.0184	0.0142	0.0118	0.0167
France		0.0290	0.0217	0.0362	0.0307	0.0264	0.0350
Germany		0.0271	0.0205	0.0337	0.0300	0.0230	0.0370
UK		0.0114	0.0089	0.0140	0.0082	0.0063	0.0102
USA		0.0026	−0.0013	0.0065	0.0001	− 0.0048	0.0050
Italy	$I(1)$	0.0323	0.0059	0.0587	0.0336	0.0127	0.0546
Norway		0.0192	0.0046	0.0337	0.0192	0.0046	0.0337
Sweden		0.0082	−0.0037	0.0201	0.0076	− 0.0071	0.0223

For Model I, the estimates are obtained from the regression $y_t = \hat{\mu}_1 + \hat{\beta}_1 t + \hat{\beta}_b \hat{B}_t + \hat{u}_t$ while for Model II they are obtained from the regression $y_t = \hat{\mu}_1 + \hat{\beta}_1 t + \hat{\mu}_b \hat{C}_t + \hat{\beta}_b \hat{B}_t + \hat{u}_t$. We categorize the error structures as $I(0)$ or $I(1)$ according to the unit root tests reported in Perron (1992). Five thousand simulated values are used to estimate the 90% confidence intervals for $\hat{T}_1$ in Models II.a and II.b.

$I(1)$ errors and, using Theorem 4, part 4 assuming Normal errors for the case with $I(0)$ errors). To estimate the parameter σ^2, we first calculated the fitted residual $\hat{u}_t$ using the OLS estimates of the break date and the other parameters. If the error term is assumed to be $I(1)$, then $\hat{\sigma}^2 = T^{-1}\sum_{t=2}^{T}(\Delta\hat{u}_t)^2 + 2T^{-1}\sum_{j=2}^{T-1} w(j,m)\sum_{t=j+1}^{T}\Delta\hat{u}_t\Delta\hat{u}_{t-j}$ where $w(j,m)$ is the quadratic spectral kernel and the bandwidth m is selected using Andrews (1991) method assuming an AR(1) approximation for $\Delta\hat{u}_t$. If the error term is assumed to be $I(0)$, the same method is used to construct $\hat{\sigma}^2$ with $\hat{u}_t$ instead of $\Delta\hat{u}_t$. In the case of Model II.b, the limit distribution also depends on the exact distribution of the errors u_t, which accordingly needs to be specified. To that effect, we assume that u_t is a linear process that can be approximated by an autoregressive model. We estimate an AR(p) with p selected using the BIC (Bayesian information criterion). We assume Normality for the errors, and given the estimates for the autoregressive parameters and the variance of the residuals, we simulate a realization for the errors.

The results clearly show interesting patterns in accordance with our theoretical results. Consider the case of Australia and Germany for which the estimated level shift is small. Our theoretical results indicated that more precise estimates could be obtained by not including the level shift regressor (i.e., using Model I). The confidence intervals obtained for the break dates and slope changes clearly indicates this. Consider the case of Australia. Using Model I, the break date is estimated at 1934 and the confidence interval is (1930, 1938) which is rather small. Using Model II with a level shift the estimated break date is 1929 and the confidence interval is now (1926, 1943), indeed much larger. The same holds for Germany.

The results also show that including the level shift regressor can lead to better estimates when the true level shift is large. This is illustrated by looking at the results for France (see Fig. 1 where the large level shift is evident). Using Model I, the break date is estimated at 1950 and the confidence interval is (1945, 1955); using Model II with a level shift the estimated break date is 1943 and the confidence interval is now (1943, 1948) indeed smaller. The same feature holds for the United Kingdom in which case the 90% confidence interval from Model II.b includes only the year 1919.

The effect of having $I(1)$ errors is also clear from the results. Indeed, whether using Model I or II, the confidence interval for the break dates are quite large for Italy, Norway and Sweden in accordance with our theoretical results. Finally, the results for the United States show very wide confidence intervals. This should not be surprising in view of the fact that Fig. 1 suggests that a slope change is likely not to have occurred.

7. Conclusion

We considered asymptotic distributions in the context of a breaking trend function with $I(1)$ or $I(0)$ errors. Our results show interesting qualitative differences from those that are obtained in a stationary context. First, the rate of convergence and the ensuing asymptotic distribution of the estimated break date can be quite different. Second, we have uncovered how the inclusion or exclusion of a level shift regressor can change the results in important ways. The presence of a level shift in the data generating process also has important qualitative effects. We have shown that our limiting results, in particular the limiting distributions derived, are good approximations in finite samples and can accordingly be useful tools for inference. When the standard asymptotic distribution was found to be a poor approximation, we provided an asymptotic expansion that delivers very accurate results. The usual caveat applies in that these conclusions are drawn from a limited set of simulation experiments, and it is always possible that they do generalize to cases not considered.

Our analysis has implications for unit root tests that allow for a change in the trend function at some unknown date (see Perron, 1997; Zivot and Andrews, 1992; Vogelsang and Perron, 1998, among others). The standard practice is to choose the break date by minimizing the unit root t-statistic. It is natural to consider choosing the break date by minimizing the sum of squared residuals in the types of regressions that we analyzed. It is possible then to use our asymptotic results to derive the limiting distribution of the ensuing unit root tests. This is the object of ongoing research.

Acknowledgements

Pierre Perron acknowledges financial support from the National Science Foundation under Grant SES-0078492. We are especially grateful to Zhongjun Qu for detailed comments on a previous draft as well as three referees.

Appendix A

In this appendix, we prove results model by model. Since we use almost identical strategies to derive all the asymptotic properties, we only give detailed proofs for Model I.a and we outline the main differences and derive the relevant results with less details for the other models. In most cases, the argument of symmetry will apply. So we simply assume, without loss of generality, that $\hat{T}_1 \geqslant T_1^0$ (or $T_1 \geqslant T_1^0$ for any generic potential break date T_1). In cases where the symmetry argument does not hold, we treat both cases ($\hat{T}_1 \geqslant T_1^0$ and $\hat{T}_1 < T_1^0$) separately. Throughout, we use "$\to$" to denote the uniform convergence of a sequence of non-random elements, "$\to_p$" convergence in probability, "$\to^d$" convergence in distribution, and "$\Rightarrow$" weak convergence in the space $D[0,1]$ under the Skorohod metric. Throughout, we let $\lambda = T_1/T$ for some generic T_1.

A.1. Model I.a—joint broken trend with $I(1)$ errors

Here, $X = [\iota, \mathbf{t}, B]'$ with $\iota = (1,\ldots,1)'$ and $\mathbf{t} = (1,\ldots,T)'$. Define $\tilde{\iota}_b = (\tilde{\iota}_b(1),\ldots,\tilde{\iota}_b(T))$ and $\iota_b = (\iota_b(1),\ldots,\iota_b(T))$ where,

$$\text{if } T_1 > T_1^0, \quad \tilde{\iota}_b(t) \equiv \begin{cases} 0 & \text{if } 1 \leqslant t \leqslant T_1^0, \\ (t - T_1^0)/(T_1 - T_1^0) & \text{if } T_1^0 < t < T_1, \\ 1 & \text{if } T_1 \leqslant t \leqslant T, \end{cases}$$

$$\text{if } T_1 = T_1^0, \quad \tilde{\iota}_b(t) = \iota_b(t) \equiv \begin{cases} 0 & \text{if } 1 \leqslant t \leqslant T_1^0, \\ 1 & \text{if } T_1^0 < t \leqslant T. \end{cases}$$

It follows that $(X_{T_1^0} - X_{T_1})\gamma^0 = \beta_b^0(T_1 - T_1^0)\tilde{\iota}_b$. Note that $\tilde{\iota}_b([Tr])$ converges to a continuous function $f_{\tilde{\iota}_b}(r)$ over [0, 1] such that

$$\text{if } \lambda > \lambda^0, \quad f_{\tilde{\iota}_b}(r) = \begin{cases} 0 & \text{if } 0 \leqslant r \leqslant \lambda^0, \\ (r-\lambda^0)/(\lambda-\lambda^0) & \text{if } \lambda^0 < r < \lambda, \\ 1 & \text{if } \lambda \leqslant r \leqslant 1, \end{cases}$$

$$\text{if } \lambda = \lambda^0, \quad f_{\tilde{\iota}_b}(r) = f_{\iota_b} \equiv \begin{cases} 0 & \text{if } 0 \leqslant r < \lambda^0, \\ 1 & \text{if } \lambda^0 \leqslant r \leqslant 1. \end{cases}$$

A.1.1. Consistency

Proof of Lemma 1.a. Note that

$$(XX) \equiv \gamma^{0\prime}(X_{T_1^0} - X_{T_1})'(I - P_{T_1})(X_{T_1^0} - X_{T_1})\gamma^0 = (T_1 - T_1^0)^2(\beta_b^0)^2\tilde{\iota}_b'(I - P_{T_1})\tilde{\iota}_b.$$

It suffices to show that $\tilde{\iota}_b'(1 - P_{T_1})\tilde{\iota}_b$ is $\mathrm{O}(T)$ uniformly over all generic $T_1 \in [\pi T, (1-\pi)T]$. Note that $\tilde{\iota}_b'(1 - P_{T_1})\tilde{\iota}_b$ is the sum of squared residuals from a regression of $\tilde{\iota}_b$ on $[\iota, \mathbf{t}, B]$. Denote it as SSR_T. Now consider the continuous time least-squares projection of the function $f_{\tilde{\iota}_b}(r)$ on $[1, r, f_B(r)]$, where $f_B(r) = 1(r \geqslant \lambda)(r - \lambda)$. Denote the resulting sum of squared residuals by SSR_∞ and the estimated coefficients as $\hat{\delta} = (\hat{\alpha}, \hat{\beta}, \hat{\psi})$. From the definition of a Riemann integral, $T^{-1}\mathrm{SSR}_T \to \mathrm{SSR}_\infty$. Now

$$\mathrm{SSR}_\infty = \int_0^1 (f_{\tau_b}(r) - \hat{\alpha} - \hat{\beta}r - \hat{\psi}f_B(r))^2\,\mathrm{d}r.$$

If $\hat{\alpha} = \hat{\beta} = 0$, we obviously have $\mathrm{SSR}_\infty > 0$. Otherwise, we have

$$\mathrm{SSR}_\infty \geqslant \int_0^{\min(\lambda,\lambda^0)} (f_{\tau_b}(r) - \hat{\alpha} - \hat{\beta}r - \hat{\psi}f_B(r))^2\,\mathrm{d}r = \int_0^{\min(\lambda,\lambda^0)} (\hat{\alpha} + \hat{\beta}r)^2\,\mathrm{d}r > 0.$$

The last inequality follows since both λ and λ^0 are bounded away from zero. Hence, $\mathrm{SSR}_\infty > 0$ and $\mathrm{SSR}_T = \mathrm{O}(T)$. Also, $\mathrm{SSR}_\infty < \infty$. Now consider the term (XU). We have

$$(XU) = \gamma^{0\prime}(X_{T_1^0} - X_{T_1})'(I - P_{T_1})U = (T_1 - T_1^0)\beta_b^0\tilde{\iota}_b'(I - P_{T_1})U,$$

hence it suffices to show that $\tilde{\iota}_b'(I - P_{T_1})U = \mathrm{O_p}(T^{3/2})$ over all $T_1 \in [\pi T, (1-\pi)T]$. Define $f^*_{\tilde{\iota}_b}(r)$ as the projection residual of a least-squares regression of $f_{\tilde{\iota}_b}(r)$ on a constant, r and $f_B(r)$. According to the properties of projections and the results for the term (XX) above, $\int_0^1 f^*_{\tilde{\iota}_b}(r)\,\mathrm{d}r = 0$ and $\int_0^1 (f^*_{\tilde{\iota}_b}(r))^2\,\mathrm{d}r = \mathrm{O}(1)$ uniformly over all λ. By the continuous mapping theorem, we have that uniformly over all λ,

$$T^{-3/2}\tilde{\iota}_b'(I - P_{T_1})U \Rightarrow \sigma\int_0^1 f^*_{\tilde{\iota}_b}(r)W(r)\,\mathrm{d}r$$

where $W(r)$ is a standard Weiner process. Define $F^*_{\tilde{\imath}_b}(r) = \int_0^r f^*_{\tilde{\imath}_b}(r)\,\mathrm{d}s$, then

$$\int_0^1 f^*_{\tilde{\imath}_b}(r)W(r)\,\mathrm{d}r = \int_0^1 W(r)\,\mathrm{d}F^*_{\tilde{\imath}_b}(r)$$
$$= [W(r)F^*_{\tilde{\imath}_b}(r)]_0^1 - \int_0^1 F^*_{\tilde{\imath}_b}(r)\,\mathrm{d}W(r) = -\int_0^1 F^*_{\tilde{\imath}_b}(r)\,\mathrm{d}W(r).$$

It is easy to show that $\mathrm{E}(\int_0^1 F^*_{\tilde{\imath}_b}(r)\,\mathrm{d}W(r)) = 0$ and $Var(\int_0^1 F^*_{\tilde{\imath}_b}(r)\,\mathrm{d}W(r)) = \int_0^1 (F^*_{\tilde{\imath}_b}(r))^2\,\mathrm{d}r = \mathrm{O}(1) > 0$ uniformly over λ. Hence, $\int_0^1 f^*_{\tilde{\imath}_b}(r)W(r)\,\mathrm{d}r$ is $\mathrm{O_p}(1)$ and $\tilde{\imath}_b'(I - P_{T_1})U = \mathrm{O_p}(T^{3/2})$ uniformly over $\lambda \in [\pi, 1-\pi]$.

Consider now the term (UU). Let $D_T = diag(T, T^3, T^3)$. We have the decomposition,

$$\begin{aligned}
&U'(P_{T_1^0} - P_{T_1})U \\
&= U'(X_{T_1^0} - X_{T_1})D_T^{-1/2}\left(D_T^{-1/2}X'_{T_1^0}X_{T_1^0}D_T^{-1/2}\right)^{-1} D_T^{-1/2}X'_{T_1^0}U \\
&\quad + U'X_{T_1}D_T^{-1/2}(D_T^{-1/2}X'_{T_1}X_{T_1}D_T^{-1/2})^{-1}D_T^{-1/2}(X'_{T_1}X_{T_1} - X'_{T_1^0}X_{T_1^0})D_T^{-1/2} \\
&\quad \times(D_T^{-1/2}X'_{T_1^0}X_{T_1^0}D_T^{-1/2})^{-1}D_T^{-1/2}X'_{T_1^0}U \\
&\quad + U'X_{T_1}D_T^{-1/2}(D_T^{-1/2}X'_{T_1}X_{T_1}D_T^{-1/2})^{-1}D_T^{-1/2}(X_{T_1^0} - X_{T_1})'U.
\end{aligned}$$

We shall make use of the following results which are standard $T^{-3/2}\sum_{t=1}^T u_t \Rightarrow \sigma\int_0^1 W(r)\,\mathrm{d}r$, $T^{-5/2}\sum_{t=1}^T tu_t \Rightarrow \sigma\int_0^1 rW(r)\,\mathrm{d}r$, $T^{-5/2}\sum_{t=T_1+1}^T (t-T_1)u_t \Rightarrow \sigma\int_\lambda^1 (r-\lambda)W(r)\,\mathrm{d}r$, $T^{-3}\sum_{t=T_1+1}^T (t-T_1)^2 \to \int_\lambda^1 (r-\lambda)^2\,\mathrm{d}r$, $T^{-3}\sum_{t=T_1+1}^T (t-T_1)t \to \int_\lambda^1 (r-\lambda)r\,\mathrm{d}r$, and $T^{-2}\sum_{t=T_1+1}^T (t-T_1) \to \int_\lambda^1 (r-\lambda)\,\mathrm{d}r$. Letting B_{T_1} be a vector of dimension T with tth entry given by $1(t > T_1)(t - T_1)$, it is easy to show that:

1. $D_T^{-1/2}X'_{T_1}X_{T_1}D_T^{-1/2}$ and $D_T^{-1/2}X'_{T_1^0}X_{T_1^0}D_T^{-1/2}$ are $\mathrm{O}(1)$ uniformly in λ.
2. $D_T^{-1/2}X'_{T_1}U$ and $U'X_{T_1^0}D_T^{-1/2}$ are $\mathrm{O_p}(T)$ uniformly in λ.
3. For the term $U'(X_{T_1^0} - X_{T_1})D_T^{-1/2}$, note that the first two columns in $X_{T_1^0} - X_{T_1}$ are zero and the third column is $B_{T_1^0} - B_{T_1}$ and, uniformly in λ,
$$T^{-3/2}U'(B_{T_1^0} - B_{T_1}) = T^{-3/2}\sum_{t=T_1^0}^{T_1}(t - T_1^0)u_t + T^{-3/2}(T_1 - T_1^0)\sum_{t=T_1+1}^T u_t$$
$$= (T_1 - T_1^0)\mathrm{O_p}(1).$$
4. For the term $D_T^{-1/2}(X'_{T_1}X_{T_1} - X'_{T_1^0}X_{T_1^0})D_T^{-1/2}$, all the elements of $(X'_{T_1}X_{T_1} - X'_{T_1^0}X_{T_1^0})$ are zero except those associated with B_{T_1} or $B_{T_1^0}$. For these

nonzero terms,

$$B'_{T_1^0}B_{T_1^0} - B'_{T_1}B_{T_1} = \sum_{t=T_1^0}^{T_1}(t-T_1^0)^2 + \sum_{t=T_1+1}^{T}(T_1-T_1^0)(2t-T_1-T_1^0)$$
$$= |T_1 - T_1^0|\mathrm{O}(T^2),$$

$$B'_{T_1^0}\mathbf{t} - B'_{T_1}\mathbf{t} = \sum_{t=T_1^0}^{T_1}(t-T_1^0)t + (T_1-T_1^0)\sum_{t=T_1+1}^{T} t = |T_1-T_1^0|\mathrm{O}(T^2),$$

$$B'_{T_1^0}\iota - B'_{T_1}\iota = \sum_{t=T_1^0}^{T_1}(t-T_1^0) + (T_1-T_1^0)(T-T_1) = |T_1-T_1^0|\mathrm{O}(T).$$

Hence, $D_T^{-1/2}(X'_{T_1}X_{T_1} - X'_{T_1^0}X_{T_1^0})D_T^{-1/2} = |T_1 - T_1^0|\mathrm{O}(T^{-1})$.

The conclusion of the lemma follows from the above four results. We can now prove Theorem 2 (Model 1.a) using a contradiction argument. Define

$$(\hat{X}\hat{X}) \equiv \gamma^{0\prime}(X_{T_1^0} - X_{\hat{T}_1})'(I - P_{\hat{T}_1})(X_{T_1^0} - X_{\hat{T}_1})\gamma^0,$$

$$(\hat{X}\hat{U}) \equiv \gamma^{0\prime}(X_{T_1^0} - X_{\hat{T}_1})'(I - P_{\hat{T}_1})U,$$

$$(\hat{U}\hat{U}) \equiv U'(P_{T_1^0} - P_{\hat{T}_1})U.$$

Suppose that $\hat{\lambda} \nrightarrow_{\mathrm{p}} \lambda^0$, then from Lemma 1, $(\hat{X}\hat{X}) = \mathrm{O}(T^3)$, $(\hat{X}\hat{U}) = \mathrm{O_p}(T^{5/2})$ and $(\hat{U}\hat{U}) = \mathrm{O_p}(T)$. Therefore for large enough T, with some positive probability, the positive term $(\hat{X}\hat{X})$ dominates the others such that the inequality $(\hat{X}\hat{X}) + 2(\hat{X}\hat{U}) + (\hat{U}\hat{U}) < 0$ cannot hold with probability 1. Since we know that the above inequality is true for all T, we have a contradiction which implies that $\hat{\lambda} \rightarrow_{\mathrm{p}} \lambda^0$.

A.1.2. Rate of convergence

Proof of Theorem 3.1. Consider the set $V(\varepsilon) = \{|T_1 - T_1^0| < \varepsilon T\}$. From the result of Theorem 2, $\Pr(\hat{T}_1 \in V(\varepsilon)) \rightarrow 1$ as $T \rightarrow \infty$. Hence, we need only examine the behavior of the sum of squared residuals $\mathrm{SSR}(T_1)$ for those break dates T_1 that satisfy $|T_1 - T_1^0| < \varepsilon T$. Now consider the set

$$V_C(\varepsilon) = \{T_1 : |T_1 - T_1^0| < \varepsilon T \text{ and } |T_1 - T_1^0| > CT^{1/2}\}.$$

Note that $V_C(\varepsilon) \subset V(\varepsilon)$. Since $\mathrm{SSR}(\hat{T}_1) \leqslant \mathrm{SSR}(T_1^0)$ with probability 1, it is enough to show that for each η, there exists a number $C > 0$ such that

$$\Pr\left(\min_{T_1 \in V_C(\varepsilon)} \{S(T_1) - S(T_1^0)\} < 0\right) < \eta. \tag{A.1}$$

Establishing (A.1) implies that a minimum cannot be achieved over $V_C(\varepsilon)$ and that $|T_1 - T_1^0| \leqslant CT^{1/2}$ must hold with an arbitrarily large probability. Now (A.1) is equivalent to

$$\Pr\left(\min_{T_1 \in V_C(\varepsilon)} \{[(XX) - 2(XU) + (UU)]\} < 0\right) < \eta.$$

We can normalize these three terms by dividing them by $|T_1 - T_1^0|T^{3/2}$. Using Lemma 1 and the fact that on the set $V_C(\varepsilon)$ we have $|T_1 - T_1^0| < \varepsilon T$ and $|T_1 - T_1^0| > CT^{1/2}$, we have

$$\frac{(XX)}{|T_1 - T_1^0|T^{3/2}} > aC + \mathrm{o_p}(1), \quad \frac{(XU)}{|T_1 - T_1^0|T^{3/2}} = \mathrm{O_p}(1), \quad \frac{(UU)}{|T_1 - T_1^0|T^{3/2}} = \mathrm{o_p}(1),$$

where a is a positive constant. Hence, given any small ε, we can choose a C large enough so that (A.1) is satisfied.

A.1.3. Limiting distribution of the estimated break

Proof of Theorem 4.1. Define the set $D(C) = \{T_1 : |T_1 - T_1^0| < \sqrt{T}C\}$ for positive number C, and $m_T = |T_1 - T_1^0|/\sqrt{T}$. To derive the limiting distribution, we analyze

$$\arg\min_{T_1 \in D(C)} [\mathrm{SSR}(T_1) - \mathrm{SSR}(T_0)]/T^2.$$

For $T_1 \in D(C)$, we have $|T_1 - T_1^0| = \mathrm{O}(T^{1/2})$. Hence, $(XX) = |T_1 - T_1^0|^2 \mathrm{O}(T) = \mathrm{O}(T^2)$, $(XU) = |T_1 - T_1^0| \mathrm{O_p}(T^{3/2}) = \mathrm{O_p}(T^2)$ and $(UU) = |T_1 - T_1^0| \mathrm{O_p}(T) = \mathrm{O_p}(T^{3/2})$. Then,

$$\begin{aligned} &\arg\min_{T_1}[\mathrm{SSR}(T_1) - \mathrm{SSR}(T_0)]/T^2 \\ &\quad = \arg\min_{T_1}[(XX) + (XU) + (UU)]/T^2 \\ &\quad = \arg\min_{T_1}[(XX)/T^2 + (XU)/T^2 + \mathrm{o_p}(1)] \end{aligned}$$

and we only need to consider the first two terms. Note that on the set $D(C)$, $|\lambda - \lambda^0| = \mathrm{O}(T^{-1/2})$. Using this fact, we can derive the following results that will be used subsequently:

$$D_T^{-1/2} X'_{T_1} X_{T_1} D_T^{-1/2} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{(1-\lambda^0)^2}{2} \\ \frac{1}{2} & \frac{1}{3} & \frac{(1-\lambda^0)^2(\lambda^0+2)}{6} \\ \frac{(1-\lambda^0)^2}{2} & \frac{(1-\lambda^0)^2(\lambda^0+2)}{6} & \frac{(1-\lambda^0)^3}{3} \end{bmatrix} + \mathrm{o}(1)$$

and the inverse is $(D_T^{-1/2} X'_{T_1} X_{T_1} D_T^{-1/2})^{-1} = \Sigma_a^{-1} + \mathrm{o}(1)$ with

$$\Sigma_a^{-1} = \begin{bmatrix} \frac{(\lambda^0+3)}{\lambda^0} & -\frac{3(\lambda^0+1)}{(\lambda^0)^2} & \frac{3}{(\lambda^0)^2(1-\lambda^0)} \\ -\frac{3(\lambda^0+1)}{(\lambda^0)^2} & \frac{3(3\lambda^0+1)}{(\lambda^0)^3} & -\frac{3(2\lambda^0+1)}{(\lambda^0)^3(1-\lambda^0)} \\ \frac{3}{(\lambda^0)^2(1-\lambda^0)} & -\frac{3(2\lambda^0+1)}{(\lambda^0)^3(1-\lambda^0)} & \frac{3}{(\lambda^0)^3(1-\lambda^0)^3} \end{bmatrix}.$$

Also,

$$T^{-1}(B_{T_1^0} - B_{T_1})' X_{T_1} D_T^{-1/2} = \begin{bmatrix} 1-\lambda^0 & \frac{1-(\lambda^0)^2}{2} & \frac{(1-\lambda^0)^2}{2} \end{bmatrix} m_T + \mathrm{o}(1).$$

Using the results above,

$$\begin{aligned} & T^{-1}\left(B_{T_1^0} - B_{T_1}\right)' X_{T_1} D_T^{-1/2}\left(D_T^{-1/2} X'_{T_1} X_{T_1} D_T^{-1/2}\right)^{-1} \\ & \quad = \begin{bmatrix} -\frac{1-\lambda^0}{2} & \frac{3(1-\lambda^0)}{2\lambda^0} & \frac{3(2\lambda^0-1)}{2\lambda^0(1-\lambda^0)} \end{bmatrix} m_T + \mathrm{o}(1). \end{aligned} \tag{A.2}$$

Hence,

$$T^{-2}(B_{T_1^0} - B_{T_1})' X_{T_1} (X'_{T_1} X_{T_1})^{-1} X'_{T_1} (B_{T_1^0} - B_{T_1}) = \left[\frac{(1-\lambda^0)(4-\lambda^0)}{4}\right] m_T + \mathrm{o}(1) \tag{A.3}$$

and

$$T^{-2}(B_{T_1^0} - B_{T_1})'(B_{T_1^0} - B_{T_1}) = (1-\lambda^0) m_T + \mathrm{o}(1). \tag{A.4}$$

Using (A.3) and (A.4), we obtain

$$T^{-2}(B_{T_1^0} - B_{T_1})'(I - P_{T_1})(B_{T_1^0} - B_{T_1}) = \left[\frac{(1-\lambda^0)\lambda^0}{4}\right] m_T + \mathrm{o}(1).$$

Now

$$\begin{aligned} T^{-2}(B_{T_1^0} - B_{T_1})' U &= T^{-2} \sum_{t=T_1^0+1}^{T_1} (t - T_1^0) u_t + T^{-2} \sum_{t=T_1+1}^{T} (T_1 - T_1^0) u_t \\ &= T^{-2} \sum_{t=T_1^0+1}^{T} (T_1 - T_1^0) u_t + \mathrm{o_p}(1) \\ &= m_T T^{-3/2} \sum_{t=T_1^0+1}^{T} u_t + \mathrm{o_p}(1) \end{aligned} \tag{A.5}$$

and

$$\begin{aligned} & T^{-1} D_T^{-1/2} X'_{T_1} U \\ & \quad = \begin{bmatrix} T^{-3/2} \sum_{t=1}^{T} u_t & T^{-5/2} \sum_{t=1}^{T} t u_t & T^{-5/2} \sum_{t=T_1+1}^{T} (t - T_1) u_t \end{bmatrix}' \\ & \quad = \begin{bmatrix} T^{-3/2} \sum_{t=1}^{T} u_t & T^{-5/2} \sum_{t=1}^{T} t u_t & T^{-5/2} \sum_{t=T_1^0+1}^{T} (t - T_1^0) u_t + \mathrm{o_p}(1) \end{bmatrix}'. \end{aligned} \tag{A.6}$$

From (A.2), (A.5) and (A.6)

$$
\begin{aligned}
&T^{-2}\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'(I - P_{T_1})U \\
&= \left\{ T^{-3/2} \sum_{t=T_1^0+1}^{T} u_t + \frac{1-\lambda^0}{2}\frac{\sum_{t=1}^T u_t}{T^{3/2}} - \frac{3(1-\lambda^0)}{2\lambda^0}\frac{\sum_{t=1}^T tu_t}{T^{5/2}} \right. \\
&\quad \left. - \frac{3(2\lambda^0-1)}{2\lambda^0(1-\lambda^0)}\frac{\sum_{t=T_1^0+1}^T (t-T_1^0)u_t}{T^{5/2}} \right\} \beta_b^0 m_T + \mathrm{o_p}(1) \\
&= \sigma\left\{ \int_{\lambda^0}^1 W(r)\,\mathrm{d}r + \frac{1-\lambda^0}{2}\int_0^1 W(r)\,\mathrm{d}r - \frac{3(1-\lambda^0)}{2\lambda^0}\int_0^1 rW(r)\,\mathrm{d}r \right. \\
&\quad \left. - \frac{3(2\lambda^0-1)}{2\lambda^0(1-\lambda^0)}\int_{\lambda^0}^1 (r-\lambda^0)W(r)\,\mathrm{d}r \right\} \beta_b^0 m_T + \mathrm{o_p}(1) \\
&\equiv \sigma\beta_b^0 m_T \int_{\lambda^0}^1 W^*(r)\,\mathrm{d}r + \mathrm{o_p}(1),
\end{aligned}
$$

where $W^*(r)$ is the residuals function from a continuous time least-squares regression of $W(r)$ on $\{1, r, 1(r > \lambda^0)(r-\lambda^0)\}$. Given the above results

$$
\begin{aligned}
m_T^* &= \arg\min_{m_T \in D(C)} \left[(XX)/T^2 + (XU)/T^2 + \mathrm{o_p}(1)\right] \\
&= \arg\min_{m_T \in D(C)} \left[m_T^2(\beta_b^0)^2 \frac{\lambda^0(1-\lambda^0)}{4} + 2\sigma m_T(\beta_b^0)\int_{\lambda^0}^1 W^*(r)\,\mathrm{d}r \right] + \mathrm{o_p}(1)
\end{aligned}
$$

by the continuous mapping theorem. Note that the objective function does not change if $T_1 - T_1^0 < 0$. Hence,

$$
m_T^* = \frac{\hat{T}_1 - T_1^0}{\sqrt{T}} \Rightarrow -\frac{4\sigma\int_{\lambda^0}^1 W^*(r)\,\mathrm{d}r}{\lambda^0(1-\lambda^0)\beta_b^0}.
$$

Now since $m_T \in D(C)$ implies that λ is in the set $\{|\lambda - \lambda_0| < CT^{-1/2}\}$ and $\lambda_0 \in [\pi, 1-\pi]$, if we consider the minimization over $[\pi, 1-\pi]$, the result remains valid. Now, $\int_{\lambda^0}^1 W^*(r)\,\mathrm{d}r$ is Normally distributed with mean 0 and tedious algebra shows that its variance is $(\lambda^0)^2(1-\lambda^0)^2/120$. Hence, we have the equivalent result $\sqrt{T}(\hat{\lambda} - \lambda_0) \to^{\mathrm{d}} \mathrm{N}(0, 2\sigma^2/[15(\beta_b^0)^2])$.

A.1.4. The other parameters

Proof of Theorem 6.1. We have

$$
\begin{aligned}
\hat{\gamma} &= (X'_{\hat{T}_1}X_{\hat{T}_1})^{-1}X'_{\hat{T}_1}Y = (X'_{\hat{T}_1}X_{\hat{T}_1})^{-1}X'_{\hat{T}_1}X_{T_1^0}\gamma^0 + (X'_{\hat{T}_1}X_{\hat{T}_1})^{-1}X'_{\hat{T}_1}U \\
&= \gamma^0 + D_T^{-1/2}(D_T^{-1/2}X'_{\hat{T}_1}X_{\hat{T}_1}D_T^{-1/2})^{-1}D_T^{-1/2}X'_{\hat{T}_1}(X_{T_1^0} - X_{\hat{T}_1})\gamma^0 \\
&\quad + D_T^{-1/2}(D_T^{-1/2}X'_{\hat{T}_1}X_{\hat{T}_1}D_T^{-1/2})^{-1}D_T^{-1/2}X'_{\hat{T}_1}U.
\end{aligned}
$$

Hence,

$$T^{-1}D_T^{1/2}(\hat{\gamma}-\gamma^0)=\left(D_T^{-1/2}X'_{\hat{T}_1}X_{\hat{T}_1}D_T^{-1/2}\right)^{-1}[T^{-1}D_T^{-1/2}X'_{\hat{T}_1}(X_{T_1^0}-X_{\hat{T}_1})\gamma^0 + T^{-1}D_T^{-1/2}X'_{\hat{T}_1}U].$$

Note that

$$\begin{aligned}
&T^{-1}D_T^{-1/2}X'_{\hat{T}_1}(X_{T_1^0}-X_{\hat{T}_1})\gamma^0+T^{-1}D_T^{-1/2}X'_{\hat{T}_1}U\\
&=\beta_b^0\,\frac{\hat{T}_1-T_1^0}{\sqrt{T}}\,D_T^{-1/2}X'_{\hat{T}_1}\hat{\imath}_bT^{-1/2}+T^{-1}D_T^{-1/2}X'_{\hat{T}_1}U\\
&\Rightarrow -\sigma\int_{\lambda^0}^1 W^*(r)\,\mathrm{d}r\begin{bmatrix}\frac{4}{\lambda^0}\\ \frac{2(1+\lambda^0)}{\lambda^0}\\ \frac{2(1-\lambda^0)}{\lambda^0}\end{bmatrix}+\sigma\begin{bmatrix}\int_0^1 W(r)\,\mathrm{d}r\\ \int_0^1 rW(r)\,\mathrm{d}r\\ \int_{\lambda^0}^1(r-\lambda^0)W(r)\,\mathrm{d}r\end{bmatrix}.
\end{aligned}$$

Tedious calculation shows that this limiting functional is equal to

$$-\sigma\begin{bmatrix}\int_0^{\lambda^0}\frac{1}{(\lambda^0)^2}[3(1-\lambda^0)r+\lambda^0](r-\lambda^0)\,\mathrm{d}W(r)+\int_{\lambda^0}^1\frac{3}{1-\lambda^0}(r-1)(r-\lambda^0)\,\mathrm{d}W(r)\\ \int_0^{\lambda^0}\frac{1}{2(\lambda^0)^2}[(3-2(\lambda^0)^2)r+\lambda^0](r-\lambda^0)\,\mathrm{d}W(r)+\int_{\lambda^0}^1\frac{\lambda^0+2}{1-\lambda^0}(r-1)(r-\lambda^0)\,\mathrm{d}W(r)\\ \int_0^{\lambda^0}\frac{(1-\lambda^0)^2}{2(\lambda^0)^2}[3r+\lambda^0](r-\lambda^0)\,\mathrm{d}W(r)+\int_{\lambda^0}^1 2(r-1)(r-\lambda^0)\,\mathrm{d}W(r)\end{bmatrix}.$$

We can show that it has a multivariate Normal distribution $\mathrm{N}(0,\Sigma_b)$ with

$$\Sigma_b=\sigma^2\begin{bmatrix}\frac{-6(\lambda^0)^2+7\lambda^0+9}{30} & \frac{-10(\lambda^0)^3+3(\lambda^0)^2+8\lambda^0+24}{120} & \frac{(1-\lambda^0)^2(-9(\lambda^0)^2+20\lambda^0+24)}{120}\\ \frac{-10(\lambda^0)^3+3(\lambda^0)^2+8\lambda^0+24}{120} & \frac{-2(\lambda^0)^4-(\lambda^0)^3+2(\lambda^0)^2+\lambda^0+8}{60} & \frac{(1-\lambda^0)^2(-3(\lambda^0)^3+10\lambda^0+16)}{120}\\ \frac{(1-\lambda^0)^2(-9(\lambda^0)^2+20\lambda^0+24)}{120} & \frac{(1-\lambda^0)^2(-3(\lambda^0)^3+10\lambda^0+16)}{120} & \frac{(1-\lambda^0)^4(8+9\lambda^0)}{60}\end{bmatrix}.$$

Since $(D_T^{-1/2}X'_{\hat{T}_1}X_{\hat{T}_1}D_T^{-1/2})^{-1}\to_{\mathrm{p}}\Sigma_a^{-1}$, we have $T^{-1}D_T^{1/2}(\hat{\gamma}-\gamma^0)\to^{\mathrm{d}}\mathrm{N}(0,\Sigma_a^{-1}\Sigma_b\Sigma_a^{-1})$ with $\Sigma_a^{-1}\Sigma_b\Sigma_a^{-1}$ as defined in Theorem 6.1.a. Suppose now that the true break date T_1^0 is known (in which case the least-squares estimate is denoted $\bar{\gamma}$). In this case, we have

$$\bar{\gamma}=\gamma^0+D_T^{-1/2}\left(D_T^{-1/2}X'_{T_1^0}X_{T_1^0}D_T^{-1/2}\right)^{-1}D_T^{-1/2}X'_{T_1^0}U.$$

Since

$$T^{-1}D_T^{-1/2}X'_{T_1^0}U\Rightarrow\begin{bmatrix}\int_0^1 W(r)\,\mathrm{d}r\\ \int_0^1 rW(r)\,\mathrm{d}r\\ \int_{\lambda^0}^1(r-\lambda^0)W(r)\,\mathrm{d}r\end{bmatrix}$$

which has a Normal distribution with variance-covariance matrix Σ_c defined by

$$\begin{bmatrix} \frac{1}{3} & \frac{5}{24} & \frac{(5+2\lambda^0-(\lambda^0)^2)(1-\lambda^0)^2}{24} \\ \frac{5}{24} & \frac{2}{15} & \frac{1}{12}(\lambda^0)^3 - \frac{1}{120}(\lambda^0)^5 - \frac{5}{24}\lambda^0 + \frac{2}{15} \\ \frac{(5+2\lambda^0-(\lambda^0)^2)(1-\lambda^0)^2}{24} & \frac{1}{12}(\lambda^0)^3 - \frac{1}{120}(\lambda^0)^5 - \frac{5}{24}\lambda^0 + \frac{2}{15} & (\frac{7}{60}\lambda^0 + \frac{2}{15})(1-\lambda^0)^4 \end{bmatrix}.$$

The limiting distribution of $T^{-1}D_T^{1/2}(\bar{\gamma}-\gamma^0)$ is then $\mathrm{N}(0, \Sigma_a^{-1}\Sigma_c\Sigma_a^{-1})$ with $\Sigma_a^{-1}\Sigma_c\Sigma_a^{-1}$ as defined in Theorem 6.1.b.

A.2. Model I.b—joint broken trend with $I(0)$ errors

A.2.1. Asymptotics for break date

The proof for consistency and rate of convergence is similar to Model I.a, so we omit a detailed proof and focus on the limiting distribution. Since the (XX) term is same as Model I.a, we only need to consider the terms (XU) and (UU). We have $T^{-1/2}\tilde{\imath}_b' U \Rightarrow \int_{\lambda^0}^1 \mathrm{d}W(r)$,

$$T^{-1/2}\tilde{\imath}_b' X_{T_1} D_T^{-1/2}(D_T^{-1/2}X_{T_1}'X_{T_1}D_T^{-1/2})^{-1} \to_{\mathrm{p}} \begin{bmatrix} -\frac{1-\lambda^0}{2} & \frac{3(1-\lambda^0)}{2\lambda^0} & \frac{3(2\lambda^0-1)}{2\lambda^0(1-\lambda^0)} \end{bmatrix}$$

$$U'X_{T_1}D_T^{-1/2} \Rightarrow \begin{bmatrix} \int_0^1 \mathrm{d}W(r) & \int_0^1 r\,\mathrm{d}W(r) & \int_{\lambda^0}^1 (r-\lambda^0)\,\mathrm{d}W(r) \end{bmatrix}.$$

Now define $m_T = \sqrt{T}(T_1 - T_1^0)$, then

$$\begin{aligned} (XU) &= \beta_b^0 (B_{T_1^0} - B_{T_1})'(I - P_{T_1})U = T^{-1/2}\beta_b^0 m_T \tilde{\imath}_b'(I-P_{T_1})U \\ &= \beta_b^0 m_T \sigma \left[\int_{\lambda^0}^1 \mathrm{d}W(r) - \begin{bmatrix} \frac{\lambda^0-1}{2} & \frac{3(1-\lambda^0)}{2\lambda^0} & \frac{3(2\lambda^0-1)}{2\lambda^0(1-\lambda^0)} \end{bmatrix} \begin{bmatrix} \int_0^1 \mathrm{d}W(r) \\ \int_0^1 r\,\mathrm{d}W(r) \\ \int_{\lambda^0}^1 (r-\lambda^0)\,\mathrm{d}W(r) \end{bmatrix} \right] \\ &\quad + \mathrm{o_p}(1) \\ &= \beta_b^0 m_T \sigma \left[\int_0^{\lambda^0} \frac{\lambda^0 - (\lambda^0)^2 - 3r + 3r\lambda^0}{2\lambda^0}\,\mathrm{d}W(r) + \int_{\lambda^0}^1 \lambda^0 \frac{2+\lambda^0-3r}{2(1-\lambda^0)}\,\mathrm{d}W(r) \right] \\ &\quad + \mathrm{o_p}(1). \end{aligned}$$

Hence, $(XU) \equiv \beta_b^0 m_T \sigma \zeta + \mathrm{o_p}(1)$ where $\zeta \sim \mathrm{N}(0, \lambda^0(1-\lambda^0)/4)$. Consider now the term (UU),

$$\begin{aligned} (UU) &= U'\left(X_{T_1^0} - X_{T_1}\right) D_T^{-1/2} \left(D_T^{-1/2} X_{T_1^0}' X_{T_1^0} D_T^{-1/2} \right)^{-1} D_T^{-1/2} X_{T_1^0}' U \\ &\quad + U'X_{T_1}D_T^{-1/2}\left(D_T^{-1/2}X_{T_1}'X_{T_1}D_T^{-1/2}\right)^{-1} D_T^{-1/2}\left(X_{T_1}'X_{T_1} - X_{T_1^0}'X_{T_1^0}\right)D_T^{-1/2} \end{aligned}$$

$$\times\left(D_T^{-1/2}X'_{T_1^0}X_{T_1^0}D_T^{-1/2}\right)^{-1}D_T^{-1/2}X'_{T_1^0}U$$
$$+\,U'X_{T_1}D_T^{-1/2}\left(D_T^{-1/2}X'_{T_1}X_{T_1}D_T^{-1/2}\right)^{-1}D_T^{-1/2}(X_{T_1^0}-X_{T_1})'U.$$

Since U is $I(0)$, $U'(X_{T_1^0}-X_{T_1})D_T^{-1/2}=|T_1-T_1^0|\mathrm{O_p}(T^{-1})$, $D_T^{-1/2}X'_{T_1^0}U=\mathrm{O_p}(1)$, $U'X_{T_1}D_T^{-1/2}=\mathrm{O_p}(1)$, and as in Model I.a, $D_T^{-1/2}X'_{T_1^0}X_{T_1^0}D_T^{-1/2}=\mathrm{O}(1)$ and $D_T^{-1/2}(X'_{T_1}X_{T_1}-X'_{T_1^0}X_{T_1^0})D_T^{-1/2}=|T_1-T_1^0|\mathrm{O}(T^{-1})$. Given these results, we have $(UU)=|T_1-T_1^0|\mathrm{O_p}(T^{-1})$ which is dominated by (XU) asymptotically. Using the same arguments as for Model I.a, we then obtain

$$T^{3/2}(\hat{\lambda}-\lambda^0)\to^{\mathrm{d}}\frac{-4\sigma\zeta}{\beta_b^0\lambda^0(1-\lambda^0)}=^{\mathrm{d}}\mathrm{N}\left(0,\frac{4\sigma^2}{\lambda^0(1-\lambda^0)(\beta_b^0)^2}\right).$$

A.2.2. The other parameters

As in Model I.a, we have

$$D_T^{1/2}(\hat{\gamma}-\gamma^0)=(D_T^{-1/2}X_{\hat{T}_1}X'_{\hat{T}_1}\hat{X}D_T^{-1/2})^{-1}[D_T^{-1/2}X'_{\hat{T}_1}(X_{T_1^0}-X_{\hat{T}_1})\gamma^0+D_T^{-1/2}X'_{\hat{T}_1}U].$$

Note that

$$\begin{aligned}
&D_T^{-1/2}X'_{\hat{T}_1}(X_{T_1^0}-X_{\hat{T}_1})\gamma^0+D_T^{-1/2}X'_{\hat{T}_1}U\\
&\quad=\beta_b^0\sqrt{T}(\hat{T}_1-T_1^0)D_T^{-1/2}X'_{\hat{T}_1}\tilde{\imath}_bT^{-1/2}+D_T^{-1/2}X'_{\hat{T}_1}U\\
&\quad\Rightarrow\frac{-4\sigma\zeta}{\lambda^0(1-\lambda^0)}\begin{bmatrix}1-\lambda^0\\ \frac{1-(\lambda^0)^2}{2}\\ \frac{(1-\lambda^0)^2}{2}\end{bmatrix}+\sigma\begin{bmatrix}\int_0^1\mathrm{d}W(r)\\ \int_0^1 r\,\mathrm{d}W(r)\\ \int_{\lambda^0}^1(r-\lambda^0)\,\mathrm{d}W(r)\end{bmatrix}\\
&\quad=\sigma\int_0^{\lambda^0}\begin{bmatrix}\frac{-2\lambda^0+3(\lambda^0)^2+6r-6r\lambda^0}{(\lambda^0)^2}\\ \frac{(\lambda^0)^3-2r(\lambda^0)^2-\lambda^0+3r}{(\lambda^0)^2}\\ (\lambda^0)(\lambda^0-1)+3r(1-\lambda^0)\,\frac{1-\lambda^0}{(\lambda^0)^2}\end{bmatrix}\mathrm{d}W(r)\\
&\qquad+\sigma\int_{\lambda^0}^1\begin{bmatrix}3\,\frac{1+\lambda^0-2r}{-1+\lambda^0}\\ \frac{(\lambda^0)^2+3\lambda^0-2r\lambda^0-4r+2}{-1+\lambda^0}\\ -2-2\lambda^0+4r\end{bmatrix}\mathrm{d}W(r).
\end{aligned}$$

Tedious calculations show that this limiting functional has a $\mathrm{N}(0,\sigma^2\Sigma_d)$ distribution with

$$\Sigma_d = \begin{bmatrix} \frac{4-3\lambda^0}{\lambda^0} & \frac{4-4(\lambda^0)^2+\lambda^0}{2\lambda^0} & \frac{4-7\lambda^0+2(\lambda^0)^2+(\lambda^0)^3}{2\lambda^0} \\ \frac{4-4(\lambda^0)^2+\lambda^0}{2\lambda^0} & \frac{3-3(\lambda^0)^2-3(\lambda^0)^3+4\lambda^0}{3\lambda^0} & \frac{-9(\lambda^0)^2+6(\lambda^0)^3+(\lambda^0)^4+6-4\lambda^0}{6\lambda^0} \\ \frac{4-7\lambda^0+2(\lambda^0)^2+(\lambda^0)^3}{2\lambda^0} & \frac{-9(\lambda^0)^2+6(\lambda^0)^3+(\lambda^0)^4+6-4\lambda^0}{6\lambda^0} & \frac{3-8\lambda^0+6(\lambda^0)^2-(\lambda^0)^4}{3\lambda^0} \end{bmatrix}.$$

Hence, the limiting distribution of $D_T^{1/2}(\hat{\gamma}-\gamma^0)$ is $\mathrm{N}\left(0,\sigma^2\Sigma_a^{-1}\Sigma_d\Sigma_a^{-1}\right)$ with $\Sigma_a^{-1}\Sigma_d\Sigma_a^{-1}$ as stated in Theorem 6.2.a. When the break date is assumed known, we have

$$D_T^{1/2}(\bar{\gamma}-\gamma^0) = (D_T^{-1/2}X'_{T_1^0}X_{T_1^0}D_T^{-1/2})^{-1}D_T^{-1/2}X'_{T_1^0}U,$$

where

$$U'X'_{T_1^0}D_T^{-1/2} \Rightarrow \sigma\left[\int_0^1 \mathrm{d}W(r) \quad \int_0^1 r\,\mathrm{d}W(r) \quad \int_{\lambda^0}^1 \left(r-\lambda^0\right)\mathrm{d}W(r)\right]$$

which is Normally distributed with mean zero and variance $\sigma^2\Sigma_e$ with

$$\Sigma_e = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2}-\lambda^0+\frac{1}{2}(\lambda^0)^2 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{3}-\frac{1}{2}\lambda^0+\frac{1}{6}(\lambda^0)^3 \\ \frac{1}{2}-\lambda^0+\frac{1}{2}(\lambda^0)^2 & \frac{1}{3}-\frac{1}{2}\lambda^0+\frac{1}{6}(\lambda^0)^3 & \frac{1}{3}-\lambda^0+(\lambda^0)^2-\frac{1}{3}(\lambda^0)^3 \end{bmatrix}.$$

Hence $D_T^{1/2}(\bar{\gamma}-\gamma^0)\to^{\mathrm{d}}\mathrm{N}\left(0,\sigma^2\Sigma_a^{-1}\Sigma_e\Sigma_a^{-1}\right)$ as defined in Theorem 6.2.b.

A.3. Model II.a: local disjoint broken trend with $I(1)$ errors

A.3.1. Asymptotics for break date

The proofs for consistency and the rate of convergence are similar to those of Model I.a. Hence, we again concentrate on the limiting distribution. In the following, we therefore work on the set $D(C) = \{T_1 : |T_1 - T_1^0| < \sqrt{T}C\}$ for positive number C, which implies that $\lambda = T_1/T$ is such that $|\lambda-\lambda^0| = \mathrm{O}(T^{-1/2})$. Note that $X = [\iota, \mathbf{t}, C, B]$ and $(X_{T_1^0}-X_{T_1})\gamma^0 \neq (B_{T_1^0}-B_{T_1})\beta_b^0$ in this type of model, hence we need to apply a new transformation on (XX) and (XU). Since $(T_1-T_1^0)(I-P_{T_1})C_{T_1}=0$, we have

$$\begin{aligned}&(I-P_{T_1})(X_{T_1^0}-X_{T_1})\gamma^0 \\ &\quad = (I-P_{T_1})[(C_{T_1^0}-C_{T_1})\mu_b^0 + (B_{T_1^0}-B_{T_1}-(T_1-T_1^0)C_{T_1})\beta_b^0].\end{aligned}$$

When $T_1 > T_1^0$,

$$C_{T_1^0} - C_{T_1} = 1 \quad \text{if } T_1^0 + 1 \leqslant t \leqslant T_1 \quad \text{and 0 otherwise,}$$
$$B_{T_1^0} - B_{T_1} - \left(T_1 - T_1^0\right)C_{T_1} = t - T_1^0 \quad \text{if } T_1^0 + 1 \leqslant t \leqslant T_1 \quad \text{and 0 otherwise,}$$

while when $T_1 < T_1^0$,

$$C_{T_1^0} - C_{T_1} = -1 \quad \text{if } T_1 + 1 \leqslant t \leqslant T_1^0 \quad \text{and 0 otherwise,}$$
$$B_{T_1^0} - B_{T_1} - (T_1 - T_1^0)C_{T_1} = -(t - T_1^0) \quad \text{if } T_1 + 1 \leqslant t \leqslant T_1^0 \quad \text{and 0 otherwise.}$$

We shall use the following notation. For $T_1^0 > T_1$,

$$g_1(T_1 - T_1^0) \equiv \sum_{t=T_1+1}^{T_1^0} \left[\mu_b^0 + \beta_b^0(t - T_1^0)\right],$$
$$h_1(T_1 - T_1^0) \equiv \sum_{t=T_1+1}^{T_1^0} \left[\mu_b^0 + \beta_b^0(t - T_1^0)\right]^2$$

and for $T_1^0 < T_1$,

$$g_2(T_1 - T_1^0) \equiv \sum_{t=T_1^0+1}^{T_1} \left[\mu_b^0 + \beta_b^0\left(t - T_1^0\right)\right],$$
$$h_2(T_1 - T_1^0) \equiv \sum_{t=T_1^0+1}^{T_1} \left[\mu_b^0 + \beta_b^0\left(t - T_1^0\right)\right]^2.$$

Let $n = T_1 - T_1^0$ and $k = t - T_1^0$, then

$$\text{For } n<0, \quad g_1(n) \equiv \sum_{k=n+1}^{0} [\mu_b^0 + \beta_b^0 k], \quad h_1(n) \equiv \sum_{k=n+1}^{0} [\mu_b^0 + \beta_b^0 k]^2.$$

$$\text{For } n>0, \quad g_2(n) \equiv \sum_{k=1}^{n} [\mu_b^0 + \beta_b^0 k], \quad h_2(n) \equiv \sum_{k=1}^{n} [\mu_b^0 + \beta_b^0 k]^2.$$

Henceforth, we suppress the argument $T_1 - T_1^0$ and simply use the short-hand notation g_1, g_2, h_1 and h_2. Let $D_T = diag\left(T, T^3, T, T^3\right)$, we can show that if $T_1 > T_1^0$,

$$\begin{aligned}
&\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'(I - P_{T_1})(X_{T_1^0} - X_{T_1})\gamma^0 \\
&\quad = [(C_{T_1^0} - C_{T_1})\mu_b^0 + (B_{T_1^0} - B_{T_1} - \left(T_1 - T_1^0\right)C_{T_1})\beta_b^0]' \\
&\qquad \times (I - P_{T_1})[(C_{T_1^0} - C_{T_1})\mu_b^0 \\
&\qquad + (B_{T_1^0} - B_{T_1} - (T_1 - T_1^0)C_{T_1})\beta_b^0]
\end{aligned}$$

$$= \sum_{t=T_1^0+1}^{T_1} [\mu_b^0 + \beta_b^0(t - T_1^0)]^2$$

$$- \sum_{t=T_1^0+1}^{T_1} [\mu_b^0 + \beta_b^0(t - T_1^0)] x(T_1)_t' D_T^{-1/2} (D_T^{-1/2} X_{T_1}' X_{T_1} D_T^{-1/2})^{-1}$$

$$\times D_T^{-1/2} \sum_{t=T_1^0+1}^{T_1} x(T_1)_t [\mu_b^0 + \beta_b^0(t - T_1^0)].$$

Note that $\sum_{t=T_1^0+1}^{T_1} [\mu_b^0 + \beta_b^0(t - T_1^0)]^2 = h_2$ and

$$\sum_{t=T_1^0+1}^{T_1} [\mu_b^0 + \beta_b^0(t - T_1^0)] x(T_1)_t' D_T^{-1/2}$$

$$= T^{-1/2} \sum_{t=T_1^0+1}^{T_1} [\mu_b^0 + \beta_b^0(t - T_1^0)][1 \quad t/T \quad 0 \quad 0]$$

$$= T^{-1/2} g_2 \left[1 \quad T_1^0/T \quad 0 \quad 0\right] + T^{-1/2} \sum_{k=1}^{T_1 - T_1^0} \left[\mu_b^0 + \beta_b^0 k\right] \left[0 \quad k/T \quad 0 \quad 0\right]$$

$$\leqslant T^{-1/2} |g_2| \left[1 \quad T_1^0/T \quad 0 \quad 0\right] + |g_2| T^{-1/2} |T_1 - T_1^0|/T = \mathrm{O_p}(|g_2| T^{-1/2}).$$

The last step is due to the fact that $|T_1 - T_1^0|/T \to_{\mathrm{p}} 0$. Based on this result and the fact that $\left(D_T^{-1/2} X_{T_1}' X_{T_1} D_T^{-1/2}\right)^{-1} = \mathrm{O_p}(1)$, $\gamma^{0\prime}(X_{T_1^0} - X_{T_1})' P_{T_1} (X_{T_1^0} - X_{T_1})\gamma^0 = \mathrm{O_p}(g_2^2 T^{-1}) = \mathrm{o_p}(h_2)$ because $|\lambda - \lambda^0| = \mathrm{O}(T^{-1/2})$. Therefore, we have $(XX) = h_2 + \mathrm{o_p}(h_2)$ for $T_1 > T_1^0$ and we can similarly show that $(XX) = h_1 + \mathrm{o_p}(h_1)$ for $T_1 < T_1^0$. This implies that $(XX) = |T_1 - T_1^0|^3 \mathrm{O}(1)$ if μ_b^0 is fixed, which proves Lemma 1.3.

Next we consider (XU) treating the two cases $T_1^0 > T_1$ and $T_1^0 < T_1$. We start with the following results: (1) $\left(D_T^{-1/2} X_{T_1}' X_{T_1} D_T^{-1/2}\right) = \Omega_1 + \mathrm{o}(1)$ where

$$\Omega_1 = \begin{bmatrix} \frac{4}{\lambda^0} & -\frac{6}{(\lambda^0)^2} & \frac{2}{\lambda^0} & \frac{6}{(\lambda^0)^2} \\ -\frac{6}{(\lambda^0)^2} & \frac{12}{(\lambda^0)^3} & -\frac{6}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3} \\ \frac{2}{\lambda^0} & -\frac{6}{(\lambda^0)^2} & \frac{4}{\lambda^0(1-\lambda^0)} & 6\frac{1-2\lambda^0}{(\lambda^0)^2(1-\lambda^0)^2} \\ \frac{6}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3} & 6\frac{1-2\lambda^0}{(\lambda^0)^2(1-\lambda^0)^2} & 12\frac{3(\lambda^0)^2-3\lambda^0+1}{(\lambda^0)^3(1-\lambda^0)^3} \end{bmatrix} \tag{A.7}$$

and

$$\Omega_1^{-1} = \begin{bmatrix} 1 & \frac{1}{2} & 1-\lambda^0 & \frac{(1-\lambda^0)^2}{2} \\ \frac{1}{2} & \frac{1}{3} & \frac{1-(\lambda^0)^2}{2} & \frac{(1-\lambda^0)^2(\lambda^0+2)}{6} \\ 1-\lambda^0 & \frac{1-(\lambda^0)^2}{2} & 1-\lambda^0 & \frac{(1-\lambda^0)^2}{2} \\ \frac{(1-\lambda^0)^2}{2} & \frac{(1-\lambda^0)^2(\lambda^0+2)}{6} & \frac{(1-\lambda^0)^2}{2} & \frac{(1-\lambda^0)^3}{3} \end{bmatrix}.$$

(2) $T^{-1}D_T^{-1/2}X'_{T_1}U \Rightarrow \sigma\xi_1$ where

$$\xi_1 \equiv \begin{bmatrix} \int_0^1 W(r)\,\mathrm{d}r \\ \int_0^1 rW(r)\,\mathrm{d}r \\ \int_{\lambda^0}^1 W(r)\,\mathrm{d}r \\ \int_{\lambda^0}^1 (r-\lambda^0)W(r)\,\mathrm{d}r \end{bmatrix} = \begin{bmatrix} \int_0^1 (1-r)\,\mathrm{d}W(r) \\ \int_0^1 \frac{1-r^2}{2}\,\mathrm{d}W(r) \\ \int_0^{\lambda^0}(1-\lambda^0)\,\mathrm{d}W(r) + \int_{\lambda^0}^1 (1-r)\,\mathrm{d}W(r) \\ \int_0^{\lambda^0}\frac{(1-\lambda^0)^2}{2}\,\mathrm{d}W(r) + \int_{\lambda^0}^1 \frac{(1-\lambda^0)^2-(r-\lambda^0)^2}{2}\,\mathrm{d}W(r) \end{bmatrix}. \tag{A.8}$$

(3) When $T_1^0 < T_1$, define $u_t = u_{T_1^0} + v_k$, we have

$$\begin{aligned} T^{-1/2}\sum_{t=T_1^0+1}^{T_1}[\mu_b^0 + \beta_b^0(t-T_1^0)]u_t &= T^{-1/2}\sum_{k=1}^{T_1-T_1^0}(\mu_b^0+\beta_b^0 k)(u_{T_1^0}+v_k) \\ &= T^{-1/2}g_2 u_{T_1^0} + T^{-1/2}\sum_{k=1}^{T_1-T_1^0}(\mu_b^0+\beta_b^0 k)v_k \\ &= \sigma W(\lambda^0)g_2 + \mathrm{o_p}(g_2). \end{aligned}$$

(4) When $T_1^0 < T_1$,

$$\begin{aligned} &T^{-1/2}\sum_{t=T_1^0+1}^{T_1}\big[\mu_b^0 + \beta_b^0(t-T_1^0\big)x(T_1)'_t D_T^{-1/2} \\ &\quad = T^{-1}g_2[1 \quad \lambda^0 \quad 0 \quad 0] + \mathrm{o_p}(g_2/T). \end{aligned}$$

Using the above results, we have when $T_1^0 < T_1$,

$$\begin{aligned} &\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'(I-P_{T_1})U \\ &\quad = \sum_{t=T_1^0+1}^{T_1}[\mu_b^0+\beta_b^0(t-T_1^0)]u_t \\ &\qquad - \left[\sum_{t=T_1^0+1}^{T_1}[\mu_b^0+\beta_b^0(t-T_1^0)]x(T_1)'_t D_T^{-1/2}\right] \\ &\qquad \times \left(D_T^{-1/2}X'_{T_1}X_{T_1}D_T^{-1/2}\right)^{-1} D_T^{-1/2}X'_{T_1}U \\ &\quad = T^{1/2}g_2\sigma\big[W(\lambda^0) - \big[1 \quad \lambda^0 \quad 0 \quad 0\big]\Omega_1\xi_1 + \mathrm{o_p}(1)\big] \\ &\quad = T^{1/2}g_2\sigma\xi_3 + \mathrm{o_p}(T^{1/2}g_2), \end{aligned}$$

where $\xi_3 \equiv \int_0^{\lambda^0} [(3r^2 - 2r\lambda^0)/(\lambda^0)^2]\,\mathrm{d}W(r) \sim \mathrm{N}(0, 2\lambda^0/15)$. As in results (3) and (4) above, we can show that when $T_1^0 > T_1$,

$$T^{-1/2} \sum_{t=T_1+1}^{T_1^0} [\mu_b^0 + \beta_b^0(t - T_1^0)]u_t = \sigma W(\lambda^0)g_1 + \mathrm{o_p}(g_1)$$

$$T^{-1/2} \sum_{t=T_1+1}^{T_1^0} [\mu_b^0 + \beta_b^0(t - T_1^0)]x(T_1)_t' D_T^{-1/2} = T^{-1} g_1 [1 \quad \lambda^0 \quad 1 \quad 0] + \mathrm{o_p}(T^{-1}g_1).$$

Hence, when $T_1^0 > T_1$,

$$\begin{aligned}
&\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'(I - P_{T_1})U \\
&\quad = -\sum_{t=T_1+1}^{T_1^0} [\mu_b^0 + \beta_b^0(t - T_1^0)]u_t \\
&\qquad + \left[\sum_{t=T_1+1}^{T_1^0} [\mu_b^0 + \beta_b^0(t - T_1^0)]x(T_1)_t' D_T^{-1/2}\right] \left(D_T^{-1/2} X_{T_1}' X_{T_1} D_T^{-1/2}\right)^{-1} D_T^{-1/2} X_{T_1}' U \\
&\quad = T^{1/2} g_1 \sigma[-W(\lambda^0) + [1 \quad \lambda^0 \quad 1 \quad 0]\Omega_1 \xi_1] + \mathrm{o_p}(T^{1/2} g_1) \\
&\quad = T^{1/2} g_1 \sigma \xi_4 + \mathrm{o_p}(T^{1/2} g_1),
\end{aligned}$$

where $\xi_4 \equiv \int_{\lambda^0}^1 [(r-1)(3r - 2\lambda^0 - 1)/(1-\lambda^0)^2]\,\mathrm{d}W(r) \sim \mathrm{N}(0, 2(1-\lambda^0)/15)$. Therefore,

$$(XU) = T^{1/2}\sigma \left\{ \begin{array}{ll} g_2\xi_3 & \text{if } T_1^0 < T_1 \\ g_1\xi_4 & \text{if } T_1^0 > T_1 \end{array} + \mathrm{o_p}(1) \right\}.$$

This implies that $(XU) = |T_1 - T_1^0|^{3/2}\mathrm{O_p}(1)$. Last consider (UU). We have

$$\begin{aligned}
&U'(P_{T_1^0} - P_{T_1})U \\
&\quad = U'(X_{T_1^0} - X_{T_1}) D_T^{-1/2} (D_T^{-1/2} X_{T_1^0}' X_{T_1^0} D_T^{-1/2})^{-1} D_T^{-1/2} X_{T_1^0}' U \\
&\qquad + U' X_{T_1} D_T^{-1/2} (D_T^{-1/2} X_{T_1}' X_{T_1} D_T^{-1/2})^{-1} D_T^{-1/2} (X_{T_1}' X_{T_1} - X_{T_1^0}' X_{T_1^0}) D_T^{-1/2} \\
&\qquad \times (D_T^{-1/2} X_{T_1^0}' X_{T_1^0} D_T^{-1/2})^{-1} D_T^{-1/2} X_{T_1^0}' U \\
&\qquad + U' X_{T_1} D_T^{-1/2} (D_T^{-1/2} X_{T_1}' X_{T_1} D_T^{-1/2})^{-1} D_T^{-1/2} (X_{T_1^0} - X_{T_1})' U.
\end{aligned}$$

Note that

$$T^{-1/2}U'\left(C_{T_1^0}-C_{T_1}\right)=(T_1-T_1^0)[W(\lambda^0)+o_p(1)],$$

$$T^{-3/2}U'\left(B_{T_1^0}-B_{T_1}\right)=\left(T_1-T_1^0\right)\left[\int_{\lambda^0}^1 W(r)\,dr+o_p(1)\right].$$

Hence $U'(X_{T_1^0}-X_{T_1})D_T^{-1/2}=(T_1-T_1^0)[\sigma\xi_2'+o_p(1)]$ where $\xi_2'\equiv[0,0,W(\lambda^0),\int_{\lambda^0}^1 W(r)\,dr]$. For the second term, we have $D_T^{-1/2}(X'_{T_1}X_{T_1}-X'_{T_1^0}X_{T_1^0})D_T^{-1/2}=-(T_1-T_1^0)T^{-1}\Sigma_f$ with

$$\Sigma_f\equiv\begin{bmatrix}0 & 0 & 1 & 1-\lambda^0\\ 0 & 0 & \lambda^0 & \frac{1-(\lambda^0)^2}{2}\\ 1 & \lambda^0 & 1 & 1-\lambda^0\\ 1-\lambda^0 & \frac{1-(\lambda^0)^2}{2} & 1-\lambda^0 & (1-\lambda^0)^2\end{bmatrix}.$$

So the second term is equivalent in large samples to $-(T_1-T_1^0)T\sigma^2[\xi_1'\Omega_2\xi_1+o_p(1)]$ where

$$\Omega_2\equiv\Omega_1^{-1}\Sigma_f\Omega_1^{-1}=\begin{bmatrix}-\frac{4}{(\lambda^0)^2} & \frac{12}{(\lambda^0)^3} & -\frac{2}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3}\\ \frac{12}{(\lambda^0)^3} & -\frac{36}{(\lambda^0)^4} & \frac{12}{(\lambda^0)^3} & \frac{36}{(\lambda^0)^4}\\ -\frac{2}{(\lambda^0)^2} & \frac{12}{(\lambda^0)^3} & 4\,\frac{2\lambda^0-1}{(\lambda^0)^2(1-\lambda^0)^2} & \frac{12}{(\lambda^0)^3}\,\frac{3(\lambda^0)^2-3\lambda^0+1}{(\lambda^0-1)^3}\\ -\frac{12}{(\lambda^0)^3} & \frac{36}{(\lambda^0)^4} & \frac{12}{(\lambda^0)^3}\,\frac{3(\lambda^0)^2-3\lambda^0+1}{(\lambda^0-1)^3} & \frac{36}{(\lambda^0)^4}\,\frac{4(\lambda^0)^3-6(\lambda^0)^2+4\lambda^0-1}{(\lambda^0-1)^4}\end{bmatrix}.$$

Collecting these results, $(UU)=(T_1-T_1^0)T\sigma^2[2\xi_2'\Omega_1\xi_1-\xi_1'\Omega_2\xi_1+o_p(1)]$. This implies that $(UU)=|T_1-T_1^0|O_p(T)$. We can now prove Theorem 5, concerning the asymptotic expansion. Define a stochastic process $V^*(n;T,\lambda^0,\mu_b^0,\beta_b^0,\sigma)$ on the set of integers as follows: $V^*(0)=0$, $V^*(n)=V_1(n)$ for $n<0$ and $V^*(n)=V_2(n)$ for $n>0$, with

$$V_1(n;T,\lambda^0,\mu_b^0,\beta_b^0,\sigma)=\sum_{k=n+1}^{0}\left[\mu_b^0+\beta_b^0k\right]^2+2\sigma\xi_4T^{1/2}\sum_{k=n+1}^{0}[\mu_b^0+\beta_b^0k]+n\sigma^2T[2\xi_2'\Omega_1\xi_1-\xi_1'\Omega_2\xi_1],$$

$$V_2(n;T,\lambda^0,\mu_b^0,\beta_b^0,\sigma)=\sum_{k=1}^{n}[\mu_b^0+\beta_b^0k]^2+2\sigma\xi_3T^{1/2}\sum_{k=1}^{n}\left[\mu_b^0+\beta_b^0k\right]+n\sigma^2T[2\xi_2'\Omega_1\xi_1-\xi_1'\Omega_2\xi_1].$$

Then the finite sample distribution of $(\hat{T}_1 - T_1^0)$ can be approximated by $(\hat{T}_1 - T_1^0) \sim \arg\min_n V^*(n; T, \lambda^0, \mu_b^0, \beta_b^0, \sigma)$.

We now consider the standard limiting distribution. Define $m_T = (T_1 - T_1^0)/\sqrt{T}$. We can show that both h_1 and h_2 are asymptotically equivalent to $T^{3/2}(\beta_b^0)^2|m_T|^3/3$ and both g_1 and g_2 are asymptotically equivalent to $Tm^2\beta_b^0/2$, therefore

$$T^{-3/2}(XX) = (\beta_b^0)^2|m_T|^3/3 + o_p(1),$$

$$2T^{-3/2}(XU) = m_T^2\sigma\beta_b^0 \begin{cases} \xi_3 + o_p(1) & \text{if } m>0, \\ \xi_4 + o_p(1) & \text{if } m<0, \end{cases}$$

$$T^{-3/2}(UU) = m_T\sigma^2[2\xi_2'\Omega_1\xi_1 - \xi_1'\Omega_2\xi_1] + o_p(1).$$

Define $Z^*(m; \lambda^0, \beta_b^0, \sigma)$ as follows: $Z^*(0) = 0$, $Z^*(m) = Z_1(m)$ for $m<0$ and $Z^*(m) = Z_2(m)$ for $m>0$, with

$$Z_1(m; \lambda^0, \beta_b^0, \sigma) = (\beta_b^0)^2|m_T|^3/3 + m_T^2\sigma\beta_b^0\xi_4 + m_T\sigma^2[2\xi_2'\Omega_1\xi_1 - \xi_1'\Omega_2\xi_1] + o_p(1),$$

$$Z_2(m; \lambda^0, \beta_b^0, \sigma) = (\beta_b^0)^2|m_T|^3/3 + m_T^2\sigma\beta_b^0\xi_3 + m_T\sigma^2[2\xi_2'\Omega_1\xi_1 - \xi_1'\Omega_2\xi_1] + o_p(1).$$

By the continuous mapping theorem, we have $m_T^* = (\hat{T}_1 - T_1^0)/\sqrt{T} \to^{d} \arg\min_m Z^*(m; \lambda^0, \beta_b^0, \sigma)$.

A.3.2. The other parameters

Following the proof in Model I.a, we have

$$\begin{aligned}
&T^{-1}D_T^{1/2}(\hat{\gamma} - \gamma^0) \\
&\quad = (D_T^{-1/2}X'_{\hat{T}_1}X_{\hat{T}_1}D_T^{-1/2})^{-1}[T^{-1}D_T^{-1/2}X'_{\hat{T}_1}(X_{T_1^0} - X_{\hat{T}_1})\gamma^0 + T^{-1}D_T^{-1/2}X'_{\hat{T}_1}U] \\
&\quad = \Omega_1 \left[\beta_b^0 \frac{\hat{T}_1 - T_1^0}{\sqrt{T}} \begin{bmatrix} 1-\lambda^0 \\ \frac{1-(\lambda^0)^2}{2} \\ 1-\lambda^0 \\ \frac{(1-\lambda^0)^2}{2} \end{bmatrix} + \sigma \begin{bmatrix} \int_0^1 W(r)\,\mathrm{d}r \\ \int_0^1 rW(r)\,\mathrm{d}r \\ \int_{\lambda^0}^1 W(r)\,\mathrm{d}r \\ \int_{\lambda^0}^1 (r-\lambda^0)W(r)\,\mathrm{d}r \end{bmatrix} \right] + o_p(1) \\
&\quad = \beta_b^0 \frac{\hat{T}_1 - T_1^0}{\sqrt{T}} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \sigma\Omega_1 \begin{bmatrix} \int_0^1 W(r)\,\mathrm{d}r \\ \int_0^1 rW(r)\,\mathrm{d}r \\ \int_{\lambda^0}^1 W(r)\,\mathrm{d}r \\ \int_{\lambda^0}^1 (r-\lambda^0)W(r)\,\mathrm{d}r \end{bmatrix} + o_p(1).
\end{aligned}$$

Therefore, the limiting distribution of $\hat{\mu}_b$ depends on the limiting distribution of $\hat{T}_1 - T_1^0$ while the limiting distribution of the other parameters do not. It is easy to show that

$$\sigma\Omega_1\begin{bmatrix}\int_0^1 W(r)\,\mathrm{d}r\\ \int_0^1 rW(r)\,\mathrm{d}r\\ \int_{\lambda^0}^1 W(r)\,\mathrm{d}r\\ \int_{\lambda^0}^1 (r-\lambda^0)W(r)\,\mathrm{d}r\end{bmatrix}\sim \mathrm{N}\left(\begin{bmatrix}0\\0\\0\\0\end{bmatrix},\sigma^2\begin{bmatrix}\frac{2}{15}\lambda^0 & -\frac{1}{10} & -\frac{1}{30}\lambda^0 & \frac{1}{10}\\ -\frac{1}{10} & \frac{6}{5\lambda^0} & -\frac{1}{10} & -\frac{6}{5\lambda^0}\\ -\frac{1}{30}\lambda^0 & -\frac{1}{10} & \frac{2}{15} & 0\\ \frac{1}{10} & -\frac{6}{5\lambda^0} & 0 & \frac{6}{5(1-\lambda^0)\lambda^0}\end{bmatrix}\right)$$

from which the results stated in Theorem 6.3.a follow.

A.4. Model II.b: local disjoint broken trend with $I(0)$ errors

A.4.1. Asymptotics for break date

Again, we only cover the main arguments of the proof for the limiting distribution. From the result for the rate of convergence, the following pertains to the set $D(C)=\{T_1 : |T_1-T_1^0|<C\}$ and accordingly we have $|\lambda-\lambda^0|=\mathrm{O}(T^{-1})$ for $\lambda=T_1/T$. From the analysis for Model II.a, we know that $(XX)=h_1+\mathrm{o_p}(h_1)$ for $T_1<T_1^0$ and $(XX)=h_2+\mathrm{o_p}(h_2)$ for $T_1>T_1^0$. Next, consider the term (XU). First if $T_1>T_1^0$, then

$$\begin{aligned}\gamma^{0\prime}(X_{T_1^0}-X_{T_1})'(I-P_{T_1})U = &\sum_{t=T_1^0+1}^{T_1}[\mu_b^0+\beta_b^0(t-T_1^0)]u_t\\ &-\left[\sum_{t=T_1^0+1}^{T_1}[\mu_b^0+\beta_b^0(t-T_1^0)]x(T_1)_t' D_T^{-1/2}\right]\\ &\times\left(D_T^{-1/2}X_{T_1}'X_{T_1}D_T^{-1/2}\right)^{-1}D_T^{-1/2}X_{T_1}'U\end{aligned}$$

Note that, (1) $\sum_{t=T_1^0+1}^{T_1}[\mu_b^0+\beta_b^0(t-T_1^0)]u_t=\mathrm{O_p}(|T_1-T_1^0|^{3/2})$; (2) $\sum_{t=T_1^0+1}^{T_1}[\mu_b^0+\beta_b^0(t-T_1^0)]x(T_1)_t'D_T^{-1/2}=\mathrm{O_p}(|T_1-T_1^0|^2T^{-1/2})$; and (3) $D^{-1/2}X_{T_1}'U=\mathrm{O_p}(1)$. Since we look in a set where $|T_1-T_1^0|<C$ and $|\lambda-\lambda^0|=\mathrm{O}(T^{-1})$, $\gamma^{0\prime}(X_{T_1^0}-X_{T_1})'P_{T_1}U$ is dominated by $\gamma^{0\prime}(X_{T_1^0}-X_{T_1})'U$ asymptotically, and $(XU)=|T_1-T_1^0|^{3/2}\mathrm{O_p}(1)$. Consider the case $T_1^0>T_1$ for (XU),

$$\begin{aligned}\gamma^{0\prime}(X_{T_1^0}-X_{T_1})'(I-P_{T_1})U = &-\sum_{t=T_1+1}^{T_1^0}[\mu_b^0+\beta_b^0(t-T_1^0)]u_t\\ &+\left[\sum_{t=T_1+1}^{T_1^0}[\mu_b^0+\beta_b^0(t-T_1^0)]x(T_1)_t' D_T^{-1/2}\right]\\ &\times(D_T^{-1/2}X_{T_1}'X_{T_1}D_T^{-1/2})^{-1}D_T^{-1/2}X_{T_1}'U.\end{aligned}$$

Again we can show that $\gamma^{0\prime}(X_{T_1^0} - X_{T_1})' P_{T_1} U$ is dominated by $\gamma^{0\prime}(X_{T_1^0} - X_{T_1})' U$ asymptotically. Therefore,

$$(XU) = \begin{cases} \sum_{t=T_1^0+1}^{T_1} [\mu_b^0 + \beta_b^0 (t - T_1^0)] u_t + o_p(1) & \text{if } T_1 > T_1^0, \\ 0 & \text{if } T_1 = T_1^0, \\ -\sum_{t=T_1+1}^{T_1^0} [\mu_b^0 + \beta_b^0 (t - T_1^0)] u_t + o_p(1) & \text{if } T_1 < T_1^0. \end{cases}$$

Last, consider (UU). Note that

$$T^{-1/2} U'(C_{T_1^0} - C_{T_1}) = T^{-1/2} \sum_{t=\min\{T_1, T_1^0\}+1}^{\max\{T_1, T_1^0\}} u_t = T^{-1/2} |T_1 - T_1^0|^{1/2} O_p(1)$$

$$T^{-3/2} U'(B_{T_1^0} - B_{T_1}) = |T_1 - T_1^0| T^{-1} O_p(1).$$

Hence, $U'(X_{T_1^0} - X_{T_1}) D_T^{-1/2} = |T_1 - T_1^0|^{1/2} O_p(T^{-1/2})$. Then following the same arguments as for Model I.b, $(UU) = |T_1 - T_1^0|^{1/2} O_p(T^{-1/2})$. Following Bai (1997), we define a stochastic process $S^*(m)$ on the set of integers as follows: $S^*(0) = 0$, $S^*(m) = S_1(m)$ for $m < 0$ and $S^*(m) = S_2(m)$ for $m > 0$, with

$$S_1(m) = \sum_{k=m+1}^{0} (\mu_b^0 + \beta_b^0 k)^2 - 2 \sum_{k=m+1}^{0} (\mu_b^0 + \beta_b^0 k) u_k, \quad m = -1, -2, \ldots,$$

$$S_2(m) = \sum_{k=1}^{m} (\mu_b^0 + \beta_b^0 k)^2 + 2 \sum_{k=1}^{m} (\mu_b^0 + \beta_b^0 k) u_k, \quad m = 1, 2, \ldots .$$

Under the assumption that $\{u_t\}$ is strictly stationary and has a continuous distribution, the rest of the proof is similar to that of Bai (1997, p. 562) and, hence, omitted.

A.4.2. The other parameters

Similar to Model II.a, we have

$$\begin{aligned} D_T^{1/2}(\hat{\gamma} - \gamma^0) &= (D_T^{-1/2} X'_{\hat{T}_1} X_{\hat{T}_1} D_T^{-1/2})^{-1} [D_T^{-1/2} X'_{\hat{T}_1} (X_{T_1^0} - X_{\hat{T}_1}) \gamma^0 + D_T^{-1/2} X'_{\hat{T}_1} U] \\ &= \Omega_1 \left[\beta_b^0 (\hat{T}_1 - T_1^0) T^{1/2} \begin{bmatrix} 1 - \lambda^0 \\ \frac{1 - (\lambda^0)^2}{2} \\ 1 - \lambda^0 \\ \frac{(1 - \lambda^0)^2}{2} \end{bmatrix} + \sigma \begin{bmatrix} \int_0^1 \mathrm{d}W(r) \\ \int_0^1 r \, \mathrm{d}W(r) \\ \int_{\lambda^0}^1 \mathrm{d}W(r) \\ \int_{\lambda^0}^1 (r - \lambda^0) \, \mathrm{d}W(r) \end{bmatrix} \right] \\ &\quad + o_p(1) \end{aligned}$$

$$= \beta_b^0(\hat{T}_1 - T_1^0)T^{1/2}\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \sigma\Omega_1 \begin{bmatrix} \int_0^1 \, \mathrm{d}W(r) \\ \int_0^1 r\,\mathrm{d}W(r) \\ \int_{\lambda^0}^1 \, \mathrm{d}W(r) \\ \int_{\lambda^0}^1 (r-\lambda^0)\,\mathrm{d}W(r) \end{bmatrix} + \mathrm{o_p}(1).$$

The results of Theorem 6.4.a follow using (A.7) and (A.8).

A.5. Model III.a—disjoint broken trend with $I(1)$ errors

A.5.1. Asymptotics for break date

Note that in Model III, the regressors are $[\iota, \mathbf{t}, C, B^{dj}]$. Assuming again that $T_1 > T_1^0$,

$$x(T_1^0)_t - x(T_1)_t = \begin{cases} [0\ 0\ 0\ 0] & \text{if } t \leqslant T_1^0 \\ [0\ 0\ 1\ t] & \text{if } T_1^0 + 1 \leqslant t \leqslant T_1 \\ [0\ 0\ 0\ 0] & \text{if } t \geqslant T_1 + 1. \end{cases}$$

First consider the term (XX). We have

$$\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'(X_{T_1^0} - X_{T_1})\gamma^0 = \sum_{t=T_1^0+1}^{T_1} \left[(\beta_b^0)^2 t^2 + 2\beta_b^0\mu_b^0 t + (\mu_b^0)^2\right]$$
$$= |T_1 - T_1^0|\mathrm{O}(T^2)$$

$$\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'P_{T_1}(X_{T_1^0} - X_{T_1})\gamma^0$$
$$= \gamma^{0\prime}(X_{T_1^0} - X_{T_1})'X_{T_1}D_T^{-1/2}(D_T^{-1/2}X'_{T_1}X_{T_1}D_T^{-1/2})^{-1}D_T^{-1/2}X'_{T_1}(X_{T_1^0} - X_{T_1})\gamma^0.$$

It is obvious that $D_T^{-1/2}X'_{T_1}X_{T_1}D_T^{-1/2} = \mathrm{O}(1)$. The first two rows of $D_T^{-1/2}X'_{T_1}(X_{T_1^0} - X_{T_1})\gamma^0$ are zero while the last two rows are

$$\begin{bmatrix} T^{-1/2}\sum_{t=T_1^0+1}^{T_1} \mu_b^0 \\ T^{-3/2}\sum_{t=T_1^0+1}^{T_1} \beta_b^0 t^2 \end{bmatrix} = \begin{bmatrix} \mu_b^0|T_1 - T_1^0|\mathrm{O}(T^{-1/2}) \\ \beta_b^0|T_1 - T_1^0|\mathrm{O}(T^{1/2}) \end{bmatrix}.$$

Hence $\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'P_{T_1}(X_{T_1^0} - X_{T_1})\gamma^0 = |T_1 - T_1^0|^2\mathrm{O}(T)$. Therefore $(XX) = |T_1 - T_1^0|\mathrm{O}(T^2) + |T_1 - T_1^0|^2\mathrm{O}(T)$.

Then consider (XU). Let $D_T = diag(T, T^3, T, T^3)$. We have

$$\gamma^{0\prime}(X_{T_1^0} - X_{T_1})'U = \sum_{t=T_1^0+1}^{T_1} [\mu_b^0 u_t + \beta_b^0 t u_t] \leqslant |T_1 - T_1^0|\mathrm{O_p}(T^{3/2})$$

$$\gamma^{0\prime}(X_{T_1^0} - X_{T_1})' P_{T_1} U$$
$$= \gamma^{0\prime}(X_{T_1^0} - X_{T_1})' X_{T_1} D_T^{-1/2} (D_T^{-1/2} X'_{T_1} X_{T_1} D_T^{-1/2})^{-1} D_T^{-1/2} X'_{T_1} U.$$

Since $D_T^{-1/2} X'_{T_1} U = \mathrm{O_p}(T)$, $D_T^{-1/2} X'_{T_1} X_{T_1} D_T^{-1/2} = \mathrm{O}(1)$ and $\gamma^{0\prime}(X_{T_1^0} - X_{T_1})' X_{T_1} D_T^{-1/2} = |T_1 - T_1^0| \mathrm{O}(T^{1/2})$, we have $\gamma^{0\prime}(X_{T_1^0} - X_{T_1})' P_{T_1} U \leqslant |T_1 - T_1^0| \mathrm{O_p}(T^{3/2})$. Therefore $(XU) \leqslant |T_1 - T_1^0| \mathrm{O_p}(T^{3/2})$.

Last, consider (UU). Since $D_T^{-1/2} X'_{T_1} U = \mathrm{O_p}(T)$, $D_T^{-1/2} X'_{T_1^0} U = \mathrm{O_p}(T)$, $D_T^{-1/2} X'_{T_1^0} X_{T_1^0} D_T^{-1/2} = \mathrm{O}(1)$, $D_T^{-1/2} X'_{T_1^0} X_{T_1^0} D_T^{-1/2} = \mathrm{O}(1)$, the first two rows of $D_T^{-1/2}(X_{T_1^0} - X_{T_1})' U$ are zero while the last two rows of $D_T^{-1/2}(X_{T_1^0} - X_{T_1})' U$ are equal to

$$\begin{bmatrix} T^{-1/2} \sum_{t=T_1^0}^{T_1} u_t \\ T^{-3/2} \sum_{t=T_1^0+1}^{T_1} t u_t \end{bmatrix} \leqslant |T_1 - T_1^0| \mathrm{O_p}(1).$$

Moreover, it is easy to show that $D_T^{-1/2}\left(X'_{T_1} X_{T_1} - X'_{T_1^0} X_{T_1^0} \right) D_T^{-1/2} = |T_1 - T_1^0| \mathrm{O}(T^{-1})$. Therefore, $U'(P_{T_1^0} - P_{T_1}) U \leqslant |T_1 - T_1^0| \mathrm{O_p}(T)$. This proves Lemma 1.5.

Consider the issue of consistency. From Lemma 1, the term (XX) dominates whatever the order of $|T_1 - T_1^0|$. Hence, minimizing $\mathrm{SSR}(T_1)$ is equivalent to minimizing (XX). Since the latter is positive, it must converge to 0 to ensure that inequality (4) is satisfied. Accordingly, $(XX) = \mathrm{o_p}(1)$, and $|\hat{\lambda} - \lambda^0| = \mathrm{o_p}(T^{-3})$.

A.5.2. The other parameters

We have

$$T^{-1} D_T^{1/2} (\hat{\gamma} - \gamma^0)$$
$$= \left(D_T^{-1/2} X'_{\hat{T}_1} X_{\hat{T}_1} D_T^{-1/2} \right)^{-1} \left[T^{-1} D_T^{-1/2} X'_{\hat{T}_1} (X_{T_1^0} - X_{\hat{T}_1}) \gamma^0 + T^{-1} D_T^{-1/2} X'_{\hat{T}_1} U \right]$$

Note that $T^{-1} D_T^{-1/2} X'_{\hat{T}_1} (X_{T_1^0} - X_{\hat{T}_1}) \gamma^0 = |\hat{T}_1 - T_1^0| \mathrm{O}(T^{-1/2}) = \mathrm{o_p}(1)$, hence we have

$$T^{-1} D_T^{-1/2} X'_{\hat{T}_1} (X_{T_1^0} - X_{\hat{T}_1}) \gamma^0 + T^{-1} D_T^{-1/2} X'_{\hat{T}_1} U = T^{-1} D_T^{-1/2} X'_{\hat{T}_1} U + \mathrm{o_p}(1)$$
$$\Rightarrow \sigma \left[\int_0^1 W(r)\,\mathrm{d}r, \int_0^1 r W(r)\,\mathrm{d}r, \int_{\lambda^0}^1 W(r)\,\mathrm{d}r, \int_{\lambda^0}^1 r W(r)\,\mathrm{d}r \right]'$$

which is Normally distributed with variance–covariance matrix $\sigma^2 \Sigma_g$ and

$$\Sigma_g \equiv \begin{bmatrix} \frac{1}{3} & \frac{5}{24} & \frac{2-3(\lambda^0)^2+(\lambda^0)^3}{6} & \frac{5+3(\lambda^0)^4-8(\lambda^0)^3}{24} \\ \frac{5}{24} & \frac{2}{15} & \frac{5+(\lambda^0)^4-6(\lambda^0)^2}{24} & \frac{4+(\lambda^0)^5-5(\lambda^0)^3}{30} \\ \frac{2-3(\lambda^0)^2+(\lambda^0)^3}{6} & \frac{5+(\lambda^0)^4-6(\lambda^0)^2}{24} & \frac{1-3(\lambda^0)^2+2(\lambda^0)^3}{3} & \frac{5+9(\lambda^0)^4-8(\lambda^0)^3-6(\lambda^0)^2}{24} \\ \frac{5+3(\lambda^0)^4-8(\lambda^0)^3}{24} & \frac{4+(\lambda^0)^5-5(\lambda^0)^3}{30} & \frac{5+9(\lambda^0)^4-8(\lambda^0)^3-6(\lambda^0)^2}{24} & \frac{2+3(\lambda^0)^5-5(\lambda^0)^3}{15} \end{bmatrix}.$$

Note that this implies that the distribution of $\hat{\gamma}$ is independent of the limiting distribution of $\hat{T}_1 - T_1^0$. Now,

$$D_T^{-1/2} X'_{\hat{T}_1} X_{\hat{T}_1} D_T^{-1/2} \to_{\mathrm{p}} \Sigma_h \equiv \begin{bmatrix} 1 & \frac{1}{2} & 1-\lambda^0 & \frac{1-(\lambda^0)^2}{2} \\ \frac{1}{2} & \frac{1}{3} & \frac{1-(\lambda^0)^2}{2} & \frac{1-(\lambda^0)^3}{3} \\ 1-\lambda^0 & \frac{1-(\lambda^0)^2}{2} & 1-\lambda^0 & \frac{1-(\lambda^0)^2}{2} \\ \frac{1-(\lambda^0)^2}{2} & \frac{1-(\lambda^0)^3}{3} & \frac{1-(\lambda^0)^2}{2} & \frac{1-(\lambda^0)^3}{3} \end{bmatrix} \tag{A.9}$$

and

$$(D_T^{-1/2} X'_{\hat{T}_1} X_{\hat{T}_1} D_T^{-1/2})^{-1} \to_{\mathrm{p}} \Sigma_h^{-1}$$
$$\equiv \begin{bmatrix} \frac{4}{\lambda^0} & -\frac{6}{(\lambda^0)^2} & -\frac{4}{\lambda^0} & \frac{6}{(\lambda^0)^2} \\ -\frac{6}{(\lambda^0)^2} & \frac{12}{(\lambda^0)^3} & \frac{6}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3} \\ -\frac{4}{\lambda^0} & \frac{6}{(\lambda^0)^2} & 4\frac{4(\lambda^0)^2-2\lambda^0+1}{\lambda^0(1-\lambda^0)^3} & -6\frac{4(\lambda^0)^2-3\lambda^0+1}{(\lambda^0)^2(1-\lambda^0)^3} \\ \frac{6}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3} & -6\frac{4(\lambda^0)^2-3\lambda^0+1}{(\lambda^0)^2(1-\lambda^0)^3} & 12\frac{3(\lambda^0)^2-3\lambda^0+1}{(\lambda^0)^3(1-\lambda^0)^3} \end{bmatrix}. \tag{A.10}$$

Therefore, $T^{-1} D_T^{1/2}(\hat{\gamma} - \gamma^0)$ is asymptotically Normally distributed with zero mean and variance-covariance matrix $\sigma^2 \Sigma_h^{-1} \Sigma_g \Sigma_h^{-1}$ which is as stated in Theorem 6.5.

A.6. Model III.b—disjoint broken trend with $I(0)$ errors

A.6.1. Asymptotics for break date

Model III.b is same as III.a except that the error term u_t is $I(0)$, hence again $(XX) = |T_1 - T_1^0| \mathrm{O}(T^2) + |T_1 - T_1^0|^2 \mathrm{O}(T)$. Now consider (XU), we have

$$\gamma^{0\prime}(X_{T_1^0} - X_{T_1})' U = \sum_{t=T_1^0+1}^{T_1} \left[\mu_b^0 u_t + \beta_b^0 t u_t\right] \leqslant |T_1 - T_1^0| \mathrm{O_p}(T),$$

$$\gamma^{0\prime}(X_{T_1^0} - X_{T_1})' P_{T_1} U$$
$$= \gamma^{0\prime}(X_{T_1^0} - X_{T_1})' X_{T_1} D_T^{-1/2} \left(D_T^{-1/2} X'_{T_1} X_{T_1} D_T^{-1/2}\right)^{-1} D_T^{-1/2} X'_{T_1} U.$$

Since $D_T^{-1/2} X'_{T_1} U = \mathrm{O_p}(1)$, $D_T^{-1/2} X'_{T_1} X_{T_1} D_T^{-1/2} = \mathrm{O_p}(1)$ and $\gamma^{0\prime}(X_{T_1^0} - X_{T_1})' X_{T_1} D_T^{-1/2} = |T_1 - T_1^0| \mathrm{O}(T^{1/2})$, we have $\gamma^{0\prime}(X_{T_1^0} - X_{T_1})' P_{T_1} U \leqslant |T_1 - T_1^0| \mathrm{O_p}(T^{1/2})$. Therefore $(XU) \leqslant |T_1 - T_1^0| \mathrm{O_p}(T)$.

Last consider the term (UU). Since $D_T^{-1/2} X'_{T_1} U = \mathrm{O_p}(1)$, $D_T^{-1/2} X'_{T_1^0} U = \mathrm{O_p}(1)$, $D_T^{-1/2} X'_{T_1^0} X_{T_1^0} D_T^{-1/2} = \mathrm{O}(1)$, $D_T^{-1/2} X'_{T_1} X_{T_1} D_T^{-1/2} = \mathrm{O}(1)$, the first two rows of $D_T^{-1/2}(X_{T_1^0} - X_{T_1})' U$ are zero while the last two rows of $D_T^{-1/2}(X_{T_1^0} - X_{T_1})' U$

are equal to

$$\begin{bmatrix} T^{-1/2}\sum_{t=T_1^0}^{T_1} u_t \\ T^{-3/2}\sum_{t=T_1^0+1}^{T_1} tu_t \end{bmatrix} \leqslant |T_1 - T_1^0| \mathrm{O_p}(T^{-1/2}).$$

Moreover, it is easy to show that $D_T^{-1/2}\left(X'_{T_1}X_{T_1} - X'_{T_1^0}X_{T_1^0}\right)D_T^{-1/2} = |T_1 - T_1^0|\mathrm{O}(T^{-1})$. Therefore, $(UU) \leqslant |T_1 - T_1^0|\mathrm{O_p}(T^{-1/2})$. This proves Lemma 1.6. The remaining steps to show that $|\hat{\lambda} - \lambda^0| = \mathrm{o_p}(T^{-3})$ are exactly as for Model III.a.

A.6.2. The other parameters

We have

$$D_T^{1/2}(\hat{\gamma} - \gamma^0) = (D_T^{-1/2}X'_{\hat{T}_1}X_{\hat{T}_1}D_T^{-1/2})^{-1}\left[D_T^{-1/2}X'_{\hat{T}_1}(X_{T_1^0} - X_{\hat{T}_1})\gamma^0 + D_T^{-1/2}X'_{\hat{T}_1}U\right]$$

Note that $D_T^{-1/2}X'_{\hat{T}_1}(X_{T_1^0} - X_{\hat{T}_1})\gamma^0 = |\hat{T}_1 - T_1^0|\mathrm{O}(T^{1/2}) = \mathrm{o_p}(1)$, hence we have

$$D_T^{-1/2}X'_{\hat{T}_1}(X_{T_1^0} - X_{\hat{T}_1})\gamma^0 + D_T^{-1/2}X'_{\hat{T}_1}U = D_T^{-1/2}X'_{\hat{T}_1}U + \mathrm{o_p}(1)$$

$$\Rightarrow \sigma \begin{bmatrix} \int_0^1 \mathrm{d}W(r) \\ \int_0^1 r\,\mathrm{d}W(r) \\ \int_{\lambda^0}^1 \mathrm{d}W(r) \\ \int_{\lambda^0}^1 r\,\mathrm{d}W(r) \end{bmatrix} \sim \mathrm{N}(0, \sigma^2\Sigma_h)$$

with Σ_h defined by (A.9). Hence, $D_T^{1/2}(\hat{\gamma} - \gamma^0) \to^{\mathrm{d}} \mathrm{N}(0, \Sigma_h^{-1})$ with Σ_h^{-1} defined by (A.10). The distribution of $\hat{\gamma}$ is independent of the limiting distribution of $\hat{T}_1 - T_1^0$.

References

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

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

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

Bai, J., 1994. Least squares estimation of a shift in linear processes. Journal of Time Series Analysis 15, 453–472.

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

Bai, J., Lumsdaine, R.L., Stock, J.H., 1998. Testing for and dating breaks in integrated and cointegrated time series. Review of Economic Studies 65, 395–432.

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

Bai, J., Perron, P., 2003. Computation and estimation of multiple structural change models. Journal of Applied Econometrics 18, 1–22.

Bhattacharya, P.K., 1994. Some aspects of change-point analysis. In: Carlstein, E., Müller, H.-G., Siegmund, D. (Eds.), Change Point Problems, Institute of Mathematical Statistics Lecture Notes—Monograph Series, vol. 23, Hayward, CA, pp. 28–56.

Chu, C.-S.J., White, H., 1992. A direct test for changing trend. Journal of Business and Economic Statistics 10, 189–199.

Csörgó, M., Horváth, L., 1997. Limit Theorems in Change-Point Analysis. Wiley, New York.

Feder, P.I., 1975. On asymptotic distribution theory in segmented regression problem: identified case. Annals of Statistics 3, 49–83.

Feinstein, C.H., 1972. National income expenditure and output of the United Kingdom. In: Richard Stone (Series Ed.), Studies in the National Income and Expenditure of the United Kingdom, vol. 6. Cambridge University Press, London.

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

Hansen, P.R., 2003. Structural changes in the cointegrated vector autoregressive model. Journal of Econometrics 114, 261–295.

Hansen, P.R., Johansen, S., 1999. Some tests for parameter constancy in cointegrated VAR-models. Econometrics Journal 2, 306–333.

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

Inoue, A., 1999. Test for cointegrating rank with a trend-break. Journal of Econometrics 90, 215–237.

Kormendi, R.C., Meguire, P., 1990. A multicountry characterization of the nonstationarity of aggregate output. Journal of Money, Credit and Banking 22, 77–93.

Krishnaiah, P.R., Miao, B.Q., 1988. Review about estimation of change points. In: Krishnaiah, P.R., Rao, C.R. (Eds.), Handbook of Statistics, vol. 7. Elsevier, New York.

Lumsdaine, R.L., Papell, D.H., 1997. Multiple trend breaks and the unit root hypothesis. Review of Economics and Statistics 79, 223–230.

Maddison, A., 1982. Phases of Capitalist Development. Oxford University Press, London.

Nyblom, J., 1989. Testing for the constancy of parameters over time. Journal of the American Statistical Association 84, 223–230.

Perron, P., 1989. The great crash, the oil price shock and the unit root hypothesis. Econometrica 57, 1361–1401.

Perron, P., 1991. A test for changes in a polynomial trend function for a dynamic time series. Manuscript, Department of Economics, Princeton University.

Perron, P., 1992. Trend, unit root and structural change: a multi-country study with historical data. Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp. 144–149.

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

Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75, 335–346.

Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. Annals of Statistics 20, 971–1001.

Romer, C., 1989. The prewar business cycle reconsidered: new estimates of gross national product, 1869–1928. Journal of Political Economy 97, 1–37.

Seo, B., 1998. Tests for structural change in cointegrated systems. Econometric Theory 14, 222–259.

Vogelsang, T.J., 1997. Wald-type tests for detecting shifts in the trend function of a dynamic time series. Econometric Theory 13, 818–849.

Vogelsang, T.J., Perron, P., 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., Andrews, D.W.K., 1992. Further evidence on the great crash, the oil price shock and the unit-root hypothesis. Journal of Business and Economic Statistics 10, 251–270.