Inference on a Structural Break in Trend with Fractionally Integrated Errors*

Seong Yeon Chang † Pierre Perron ‡
Xiamen University Boston University

April 30, 2014; Revised: September 20, 2015

Abstract

Perron and Zhu (2005) established the consistency, rate of convergence and the limiting distributions of parameter estimates in a linear time trend with a change in slope with or without a concurrent change in level. They considered the dichotomous cases whereby the errors are short-memory stationary, $I(0)$, or have an autoregressive unit root, $I(1)$. We extend their analysis to cover the more general case of fractionally integrated errors for values of $d^*$ in the interval $(-0.5, 1.5)$ excluding the boundary case 0.5. Our theoretical results uncover some interesting features. For example, when a concurrent level shift is allowed, the rate of convergence of the estimate of the break date is the same for all values of $d^*$ in the interval $(-0.5, 0.5)$. This feature is linked to the contamination induced by allowing a level shift, previously discussed by Perron and Zhu (2005). In all other cases, the rate of convergence is monotonically decreasing as $d^*$ increases. We also provide results about the so-called spurious break issue. Simulation experiments are provided to illustrate some of the theoretical results.

JEL Classification: C13, C18, C22

Keywords: Fractionally integrated process, Linear trend, Segmented trend, Spurious break, Structural change

---

*We thank the editor Robert Taylor, the co-editor, two anonymous referees, Dukpa Kim and Zhongjun Qu for constructive comments and suggestions. This paper is a revised version of parts of Seong Yeon Chang’s Ph.D. Dissertation at Boston University (Chang, 2014).

†Wang Yanan Institute for Studies in Economics, MOE Key Laboratory of Econometrics, Fujian Key Laboratory of Statistical Science, Xiamen University, Xiamen, Fujian 361005, China (sychang@xmu.edu.cn)

‡Department of Economics, Boston University, 270 Bay State Rd., Boston, MA 02215 (perron@bu.edu)
1 Introduction

Economic relationships are often subject to structural changes. Hence, testing for a structural break and estimating the break date have been important topics in both economics and statistics; see Perron (2006) for a review. To test for a structural break, or instability of the parameters, important contributions include Andrews (1993) and Andrews and Ploberger (1994). Bai (1994, 1997) showed that the break date can be estimated consistently by minimizing the sum of squared residuals (SSR) from the unrestricted model and derived the limiting distribution of the estimate of the break date, which can be applied to constructing confidence intervals for the true break date. Bai and Perron (1998, 2003) considered statistical inference related to multiple structural changes under general conditions.

In the literature, most of the work assumed that the regressors and the errors are short-memory stationary processes. Structural breaks in trend regressors and non-stationary processes are also important from a practical perspective. Perron (1989) showed that the Dickey and Fuller (1979) type unit root test is biased in favor of a non-rejection of the unit root null hypothesis when the process is trend stationary with a structural break in slope. Work related to changes in trend include the following. Feder (1975) considered estimating the joint points of polynomial type segmented regressions. Bai (1997) and Bai and Perron (1998) provided inference results with trending regressors. In order to obtain the limiting distribution, the trending regressors are assumed to be a function of $t/T$, say $g(t/T)$, with $T$ the sample size. Deng and Perron (2006) analyzed the consequences of specifying the trend function in scaled form when a structural break is involved. Bai et al. (1998) analyzed the limiting distribution of the estimated break date for multivariate time series with a change in slope. Chu and White (1992) suggested a testing procedure for a change in a trend function with short-memory stationary errors. Perron (1991) and Vogelsang (1997) considered testing a structural break in trend when the errors are either short-memory stationary, $I(0)$, or having an autoregressive unit root, $I(1)$. Vogelsang (1999) devised a test whose limiting distribution does not change depending on whether the noise component is $I(0)$ or $I(1)$. Recently, Perron and Yabu (2009) considered testing for structural changes in the trend function of a time series without any prior knowledge about whether the errors are $I(0)$ or $I(1)$. Their testing procedure adopts a quasi-feasible generalized least squares (GLS) approach that uses a super-efficient estimate of the sum of the autoregressive parameters $\alpha$ when $\alpha = 1$. Harvey et al. (2009) proposed a GLS-based trend break test that is asymptotically size robust with $I(0)$ or $I(1)$ errors. Sayginsoy and Vogelsang (2011) (SV, henceforth) suggested fixed-$b$ asymptotics-based slope change tests with either $I(0)$ or $I(1)$ errors. Since the limiting null distributions of the tests vary
depending on the structure of the noise component, i.e., $I(0)$ or $I(1)$, a scaling factor approach to align the two distributions for a given significant level has been adopted. Yang and Vogelsang (2011) applied the fixed-$b$ theory to a sup-LM type test in order to test a level shift. Interestingly, they found that there is a bandwidth such that the fixed-$b$ asymptotic critical value is the same for both $I(0)$ and $I(1)$ errors. With respect to the problem of estimating the break date of the change in the slope of a linear trend with or without a concurrent level shift, Perron and Zhu (2005) (PZ, henceforth) analyzed the consistency, rate of convergence and the limiting distributions of the parameter estimates when the errors are either $I(0)$ or $I(1)$. The results of PZ and Perron and Yabu (2009) have been used in Kim and Perron (2009) to provide unit root tests with improved power that allow for a change in the trend function under both the null and alternative hypotheses.

Fractionally integrated processes have been popular in the economics and statistics literature, in particular following the introduction of the ARFIMA processes by Granger and Joyeux (1980) and Hosking (1981). Kuan and Hsu (1998) considered a change in mean model and established the consistency and the rate of convergence of the least square estimate of the break date when the errors are fractionally integrated; see also Lavielle and Moulines (2000). They found that the convergence rate depends on the order of integration $d^*$. Moreover, when no such change in mean is present, the estimate of the break date obtained by minimizing the sum of squared residuals indicates a spurious break date when $d^* \in (0,0.5)$. Hsu and Kuan (2008) showed that the least square estimate of the break date in a mean change model is not consistent when the errors are fractionally integrated with $d^* \in (0.5,1.5)$, and that the spurious feature also occurs. Gil-Alana (2008) executed a set of Monte Carlo simulations to confirm that both the order of fractional integration and the break date can be estimated simultaneously by minimizing the SSR considering a range of grid values for $d^*$ and the break date $T_1$. In the context of testing for a structural change in the framework of fractionally integrated processes, the following work are relevant. Shao (2011) proposed a simple testing procedure to test for a level shift in a stationary long memory time series based on the self-normalization idea of Shao (2010). More recently, Iacone et al. (2013a) proposed a robust test for a slope change in trend when the order of fractional integration $d^*$ in the errors is located in an interval $[0,1.5)$ excluding the boundary case 0.5. Iacone et al. (2013b) considered the same problem, but the testing procedure is based on fixed-$b$ asymptotics developed by SV. By developing $d^*$-adaptive critical values, the proposed test is (asymptotically) size controlled and improves power when $d^* = 1$ compared to the test of SV. Iacone et al. (2014) analyzed a change in mean model and suggested a sup-Wald test based on fixed-$b$ asymptotics.

The main contribution of this paper is to extend PZ’s analysis to cover the more general case of fractionally integrated errors for values of $d^*$ in the interval $(-0.5,1.5)$ excluding the boundary
case 0.5. We establish the consistency, rate of convergence, and the limiting distributions of the parameter estimates in models when the trend function exhibits a slope change with or without a concurrent change in level. Our theoretical results uncover some interesting features. First, when a concurrent level shift is allowed, the rate of convergence of the estimate of the break fraction is the same for all values of $d^*$ in the interval $(-0.5, 0.5)$. This feature is linked to the contamination induced by allowing a level shift, previously discussed by PZ. In all other cases, the rate of convergence is monotonically decreasing as $d^*$ increases. Second, the coefficient of the slope change can be estimated consistently for all $d^* \in (-0.5, 0.5) \cup (0.5, 1.5)$ while the level shift coefficient is asymptotically unidentified for $d^* \in (0.5, 1.5)$. We extend Hsu and Kuan’s (2008) result to the slope change model with a concurrent level shift. Third, we also provide results about the so-called spurious break issue. For $d^* \in (0, 0.5) \cup (0.5, 1.5)$, it is very likely to estimate a spurious break when there is no break in the data generating process. Simulation experiments are provided to illustrate some of the theoretical results in the paper.

The structure of the paper is as follows. In section 2, we review fractionally integrated processes, fractional Brownian motion and useful related functional central limit theorems. Section 3 presents the models, the assumptions and a key inequality used throughout the proofs. Section 4 provides the main contributions related to the limit properties of the estimates: consistency (Section 4.1), rate of convergence (Section 4.2), limit distributions of the estimate of the break date (Section 4.3) and limit distributions of the estimates of the other parameters (Section 4.4). The problem of the possibility of a spurious break is discussed in Section 5. Section 6 provides brief concluding remarks. All technical derivations are relegated to an online Supplementary Material.

2 Fractionally Integrated Processes and Functional Central Limit Theorem

In this section, we briefly define fractionally integrated processes and review results to be used in subsequent developments. We follow the notation of Wang et al. (2003) and Robinson (2005). Define first

$$\Delta^{-a} = \sum_{j=0}^{\infty} \pi_j(a)L^j, \quad \pi_j(a) = \frac{\Gamma(j+a)}{\Gamma(a)\Gamma(j+1)}$$

(1)

where $L$ is the lag operator, $\Delta = 1 - L$ is the difference operator and $\Gamma$ is the Gamma function with $\Gamma(a) = \infty$ for $a = 0, -1, \ldots$, and $\Gamma(0)/\Gamma(0) = 1$. Let $\{\eta_t, t = 0, \pm 1, \ldots\}$ be a zero-mean short-memory covariance stationary process, with spectral density that is bounded and bounded away from zero. For $d \in (-0.5, 0.5)$,

$$\zeta_t = \Delta^{-d} \eta_t, \quad t = 0, \pm 1, \ldots, \quad (2)$$
is covariance stationary and invertible for \( d > -0.5 \). The truncated version of \( \zeta_t \) is defined as
\[
\zeta^\#_t = \zeta_t 1_{t \geq 1}, \quad t = 0, \pm 1, \ldots, \tag{3}
\]
where \( 1_A \) is the indicator function for the event \( A \). For an integer \( m \geq 0 \),
\[
u_t = \Delta^{-m} \zeta^\#_t, \quad t = 0, \pm 1, \ldots \tag{4}
\]
is called a type I \( I(m + d) \) process. A zero-mean short-memory covariance stationary process \( \eta_t \) can be represented as a one-sided moving average:
\[
\eta_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad t = 0, \pm 1, \ldots, \tag{5}
\]
where \( \psi_0 = 1, \sum_{j=0}^{\infty} \psi_j^2 < \infty \), and \( \epsilon_t, t = 0, \pm 1, \ldots \) are independent and identically distributed (i.i.d.) random variables with mean zero. Let \( D[0,1] \) be the space of functions on \([0,1]\) which are right continuous and have left limits, equipped with the Skorohod topology. Let \( \Rightarrow \) denote weak convergence in distribution under the Skorohod topology and \( \mathbb{P} \) convergence in probability. Denote by \( [a] \) the integer part of \( a \in \mathbb{R} \). The order of integration is \( d^* = m + d \) with \( m \in \mathbb{N}_0 - 1 \).

Wang et al. (2003) derived a functional central limit theorem (FCLT) for \( m \geq 0 \) which includes the non-stationary cases. To consider general non-stationary fractional processes, the following condition is required.

- **Condition A**: \( \psi_j, j \geq 0 \) in (5) satisfy \( \sum_{j=0}^{\infty} j^{1/2-d} |\psi_j| < \infty \) and \( \psi(1) \equiv \sum_{j=0}^{\infty} \psi_j \neq 0 \). Also, \( E(|\epsilon_0|^{\max(2,2/(1+2d))}) < \infty \).

We summarize their results insofar as they will be relevant for subsequent derivations.

**Lemma 1** (Wang et al., 2003, Theorem 2.2) Let \( u_t \) satisfy (4) with \( m = 0 \) and assume Condition A holds. Then, for \( d \in (-0.5, 0.5) \),
\[
\frac{1}{\kappa(d)T^{1/2+d}} \sum_{t=1}^{[Tr]} u_t \Rightarrow B_d(r), \tag{6}
\]
where \( \kappa(d)^2 = \{\psi(1)^2 \Gamma(1 - 2d) E(\epsilon_0^2)\}/\{(1 + 2d) \Gamma(1 + d) \Gamma(1 - d)\} \) and \( B_d(\cdot) \) is a type I fractional Brownian motion on \( D[0,1] \), i.e.,
\[
B_d(t) = \frac{1}{\Gamma(d+1)} \left\{ \int_{-\infty}^{0} [(t-s)^d - (-s)^d] dB(s) + \int_{0}^{d} (t-s)^d dB(s) \right\},
\]
with \( B(\cdot) \) a standard Brownian motion.

**Lemma 2** (Wang et al., 2003, Theorem 3.1) Let \( u_t \) satisfy (4) with \( m = 1 \) and assume Condition A holds. Then, for \( d \in (-0.5, 0.5) \), a) \( \kappa(d)T^{1/2+d}u_{[Tr]} \Rightarrow B_d(r) \), b) \( \kappa(d)T^{1/2+d+1} \sum_{t=1}^{[Tr]} u_t \Rightarrow \int_0^r B_d(s) ds \), c) \( \kappa^2(d) T^{2(d+1)} \sum_{t=1}^{[Tr]} u_t^2 \Rightarrow \int_0^r [B_d(s)]^2 ds \).
3 The Models

We consider the series of interest $y_t$ as consisting of a systematic part $f_t$ and a random component $u_t$, namely, $y_t = f_t + u_t$. For the noise component $u_t$, the following two assumptions hold.

- **Assumption A1**: $u_t$ is a type I $I(m + d)$ process defined by (1)-(5).
- **Assumption A2**: The conditions of Lemmas 1 and 2 are satisfied.

For the systematic part $f_t$, we consider two cases. The first, labeled Model I, specifies that $f_t$ is a first-order linear trend with a single change in slope. In this case, the trend is joined at the time of break and there is no concurrent level shift. The second, labeled Model II, specifies that $f_t$ is a first-order linear trend with a concurrent break in both intercept and slope. Let $\lambda = T_1/T$ denote a generic break fraction with a postulated break date $T_1$, whose true value is $T_1^0$.

- **Model I (Joint Broken Trend)**: The deterministic component $f_t$ is specified as
  \[ f_t = \mu_1 + \beta_1 t + \beta_b B_t, \]
  where $B_t$ is a dummy variable for the slope change defined by $B_t = t - T_1$ if $t > T_1$ and 0 otherwise. Hence, the slope coefficient changes from $\beta_1$ to $\beta_1 + \beta_b$ at time $T_1$. Note that the trend function is continuous at the time point $T_1$, hence the labeling of a “joint broken trend”.

- **Model II (Local Disjoint Broken Trend)**: The deterministic component is specified by
  \[ f_t = \mu_1 + \beta_1 t + \mu_b C_t + \beta_b B_t, \]
  where $C_t$ is a dummy variable for the level shift defined by $C_t = 1$ if $t > T_1$ and 0 otherwise. At the break date $T_1$, there is a slope change with a concurrent level shift. The magnitude of the level shift is $\mu_b$, which is asymptotically negligible compared to the level of the series $\mu_1 + \beta_1 T_1$, hence the labeling of a “local disjoint broken trend”.

In matrix notation, the models defined above can be specified as $Y = X_{T_1} \gamma + U$, where $Y = \begin{bmatrix} y_1, \ldots, y_T \end{bmatrix}'$, $U = \begin{bmatrix} u_1, \ldots, u_T \end{bmatrix}'$, $X_{T_1} = \begin{bmatrix} x(T_1)_1, \ldots, x(T_1)_T \end{bmatrix}'$, with $x(T_1)_t' = \begin{bmatrix} 1 & t & B_t \end{bmatrix}$ and $\gamma = \begin{bmatrix} \mu_1 & \beta_1 & \beta_b \end{bmatrix}'$, for Model I, while for Model II, $x(T_1)_t' = \begin{bmatrix} 1 & t & C_t & B_t \end{bmatrix}$ and $\gamma = \begin{bmatrix} \mu_1 & \beta_1 & \mu_b & \beta_b \end{bmatrix}'$. Note that the matrix $X_{T_1}$ depends on the candidate break date $T_1$. 

Remark 1 In the literature, the following high level assumption on the regressors is standard. 
\( D_T^{-1}\sum_{t=1}^{[T\lambda]} x(T_1), x(T_1)'| D_T^{-1} \xrightarrow{P} Q(\lambda) \) uniformly in \( \lambda \in [0,1] \) for some \( D_T \), where \( Q(\lambda) \) is a positive semi-definite, symmetric, and an absolutely continuous, monotonically increasing function of \( \lambda \). We do not introduce this assumption explicitly because it automatically holds for both Models I and II. Later, this high level assumption with assumptions A1 and A2 are used to obtain asymptotic results for the so-called spurious break issue.

The break date can be estimated by using a global least-squares criterion:
\[
\hat{T}_1 = \arg \min_{T_1 \in \Lambda} Y'(I - P_{T_1})Y
\]
where \( P_{T_1} \) is the matrix that projects on the range space of \( X_{T_1} \), i.e., \( P_{T_1} = X_{T_1}(X_{T_1}'X_{T_1})^{-1}X_{T_1}' \) and \( \Lambda = [\pi T, (1 - \pi)T], \) \( 0 < \pi < 1/2 \). With \( X_{\hat{T}_1} \) constructed using the estimate \( \hat{T}_1 \), the OLS estimate of \( \gamma \) is \( \hat{\gamma} = (X_{\hat{T}_1}'X_{\hat{T}_1})^{-1}X_{\hat{T}_1}'Y \) and the SSR, for an estimated break fraction \( \hat{\lambda} = \hat{T}_1/T \), is
\[
S(\hat{\lambda}) = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2 = \frac{1}{T} \sum_{t=1}^{T} (y_t - x(T_1)\hat{\gamma})^2 = Y'(I - P_{T_1})Y
\]
where \( P_{\hat{T}_1} \) is the projection matrix associated with \( X_{\hat{T}_1} \). We denote the true value of each parameter with superscript 0: \( \gamma^0 = [\mu_1^0 \beta_1^0 \mu_b^0] \) in Model I, \( \gamma^0 = [\mu_1^0 \beta_1^0 \mu_b^0 \beta_b^0] \) in Model II, \( T_0 \), and \( \lambda^0 = T_0/T \). Hence, the data generating process (DGP) is specified as
\[
Y = X_{T_1^0} \gamma^0 + U = \begin{bmatrix} x(T_{11}^0) & \ldots & x(T_{1T}^0) \end{bmatrix} \gamma^0 + U. \tag{7}
\]
Throughout, we assume that there is at least a change in slope as stated in the following assumption.

- **Assumption A3**: \( \beta_b^0 \neq 0 \) and \( \lambda^0 \in (\pi, 1 - \pi) \) for some \( \pi \in (0,1/2) \).

This assumption is required to ensure that there is a break in slope and that the pre and post break samples are asymptotically large enough to obtain consistent estimates of the unknown coefficients. This is a standard assumption needed to derive any useful asymptotic result.

As in PZ, a key inequality plays a crucial role in proving the asymptotic results. By construction, we have for all \( T \), \( S(\hat{\lambda}) \leq S(\lambda^0) \), or equivalently, \( Y'(I - P_{\hat{T}_1})Y \leq Y'(I - P_{T_0})Y \). Using (7), this inequality can be written as \( Y'(P_{T_0} - P_{\hat{T}_1})Y \leq 0 \), or equivalently,
\[
(\gamma^0 X_{T_1^0}' + U')(P_{T_0} - P_{\hat{T}_1})(X_{T_1^0} \gamma^0 + U)
= \gamma^0 X_{T_1^0}'(P_{T_0} - P_{\hat{T}_1}) X_{T_1^0} \gamma^0 + 2\gamma^0 X_{T_1^0}' P_{T_0} U + U' P_{T_0} U
= \gamma^0 (X_{T_0} - X_{\hat{T}_1})' (I - P_{\hat{T}_1}) (X_{T_0} - X_{\hat{T}_1}) \gamma^0 + 2\gamma^0 (X_{T_0} - X_{\hat{T}_1})' (I - P_{\hat{T}_1}) U + U' P_{T_0} U \leq 0
\equiv (\hat{X} \hat{X}) + 2(\hat{X} \hat{U}) + (\hat{U} \hat{U}) \leq 0 \tag{8}
\]
where we use the fact that \( X_{T_1^0}' P_{T_0} = X_{T_1^0}' \) and \( X_{T_1^0}' (I - P_{T_0}) = 0 \).
4 Asymptotic Results

We consider in turn the consistency, rate of convergence and limit distributions of the estimates, concentrating on the estimate of the break fraction.

4.1 Consistency

We show that \( \hat{\lambda} \) is a consistent estimate of \( \lambda^0 \) when the errors are fractionally integrated with parameter \( d^* \in (-0.5, 0.5) \cup (0.5, 1.5) \). The idea behind the proof is the following. Unless \( \hat{\lambda} \rightarrow \lambda^0 \), the first term in (8) would asymptotically dominate the others since it is positive provided the event \( \{T_1^0 = \hat{T}_1\} \) does not hold for all \( T \), which occurs with probability one. It means that the key inequality does not hold if \( \hat{\lambda} \) does not converge to \( \lambda^0 \) in probability, which leads to the desired contradiction. The following theorem states the consistency result.

**Theorem 1** Under Assumptions A1-A3, in Models I and II, \( \hat{\lambda} \rightarrow \lambda^0, \forall d^* \in (-0.5, 0.5) \cup (0.5, 1.5) \).

4.2 Rate of Convergence

The following theorem shows that the rate of the convergence of the estimate of the break fraction, \( \hat{\lambda} \), depends on the order of fractional integration \( d^* \). It also differs across the two models being faster with no concurrent level shift.

**Theorem 2** Under Assumptions A1-A3, for every \( d \in (-0.5, 0.5) \): 1) For Model I:

\[
\hat{\lambda} - \lambda^0 = \begin{cases} 
O_p(T^{-3/2+d}) & \text{if } m = 0 \\
O_p(T^{-1/2+d}) & \text{if } m = 1,
\end{cases}
\]

2) For Model II:

\[
\hat{\lambda} - \lambda^0 = \begin{cases} 
O_p(T^{-1}) & \text{if } m = 0 \\
O_p(T^{-1/2+d}) & \text{if } m = 1.
\end{cases}
\]

Theorem 2 implies that the rate of convergence is slower when allowing for a concurrent level shift, even if none is present, for \( d^* \in (-0.5, 0.5) \). It is, however, the same when \( d^* \in (0.5, 1.5) \). These results accord with those from PZ who considered \( I(0) \) and \( I(1) \) processes. For Models I and II with \( I(1) \) errors, \( \hat{\lambda} - \lambda^0 = O_p(T^{-1/2}) \). On the other hand, for Model I with \( I(0) \) errors, \( \hat{\lambda} - \lambda^0 = O_p(T^{-3/2}) \) and for Model II with \( I(0) \) errors, \( \hat{\lambda} - \lambda^0 = O_p(T^{-1}) \). PZ presented an intuitive explanation for the change in convergence rate induced by introducing a level shift. Briefly, a random deviation from a deterministic trend function is subject to be captured as if it were a level shift. Hence, it can have an effect on the precision of the estimate.
The results show that the rate of convergence is linearly decreasing as $d^*$ increases for all models except Model II for $d^* \in (-0.5, 0.5)$. The result for this latter case is quite interesting as the rate of convergence is the same for all $d^* \in (-0.5, 0.5)$. The explanation for this feature is again related to the contamination induced by allowing a concurrent level shift, which implies added noise. If the process is stationary, $d^* \in (-0.5, 0.5)$, this added noise dominates and renders the rate of convergence invariant to $d^*$. If the process is non-stationary, $d^* \in (0.5, 1.5)$, the noise is small compared to the signal and we are back essentially to the case with no concurrent level shift.

4.3 The Limiting Distribution of the Estimate of the Break Date

Given results about the consistency and the rate of convergence of the estimate of the break fraction $\hat{\lambda}$, we can now consider its limiting distribution. The results are stated in the following theorem.

**Theorem 3** Under Assumptions A1-A3, we have for every $d \in (-0.5, 0.5)$: 1) For Model I: a) if $m = 0$, $T^{3/2-d}(\hat{\lambda} - \lambda^0) = -4\kappa(d) \zeta/\left[\lambda^0(1-\lambda^0)\beta^0_b\right]$, b) if $m = 1$, $T^{1/2-d}(\hat{\lambda} - \lambda^0) = -4\kappa(d) \int_{\lambda^0}^{1} B^*_d(r)dr/\left[\lambda^0(1-\lambda^0)\beta^0_b\right]$, where

\[
\zeta = \int_{\lambda^0}^{1} dB_d(r) + \frac{1 - \lambda^0}{2} \int_{\lambda^0}^{1} dB_d(r) - \frac{3(1 - \lambda^0)}{2\lambda^0} \int_{\lambda^0}^{1} rdB_d(r) - \frac{3(2\lambda^0 - 1)}{2\lambda^0(1 - \lambda^0)} \int_{\lambda^0}^{1} (r - \lambda^0)dB_d(r),
\]

and

\[
\int_{\lambda^0}^{1} B^*_d(r)dr = \left[\int_{\lambda^0}^{1} B_d(r)dr + \frac{1 - \lambda^0}{2} \int_{\lambda^0}^{1} B_d(r)dr - \frac{3(1 - \lambda^0)}{2\lambda^0} \int_{\lambda^0}^{1} rdB_d(r)dr\right. \\
\left. - \frac{3(2\lambda^0 - 1)}{2\lambda^0(1 - \lambda^0)} \int_{\lambda^0}^{1} (r - \lambda^0)B_d(r)dr\right].
\]

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

\[
S_1(v) = \sum_{k=v+1}^{0} (\mu_b^0 + \beta_v^0 k)^2 - 2 \sum_{k=v+1}^{0} (\mu_b^0 + \beta_v^0 k)u_k, \quad v = -1, -2, \ldots,
\]

\[
S_2(v) = \sum_{k=1}^{v} (\mu_b^0 + \beta_v^0 k)^2 + 2 \sum_{k=1}^{v} (\mu_b^0 + \beta_v^0 k)u_k, \quad v = 1, 2, \ldots.
\]

If $u_t$ is strictly stationary with a continuous distribution, $T(\hat{\lambda} - \lambda^0) \Rightarrow \text{arg min}_v S^*(v)$. b) If $m = 1,$
define

\[ \xi_1 = \left( \int_0^1 B_d(r)dr, \int_0^1 r B_d(r)dr, \int_0^1 B_d(r)dr, \int_0^1 (r - \lambda^0)B_d(r)dr \right)' , \]

\[ \xi_2 = \left( 0,0,B_d(\lambda^0), \int_\lambda^0 B_d(r)dr \right)' , \]

\[ \xi_3 = \int_0^{\lambda^0} [(3r^2 - 2r\lambda^0)/(\lambda^0)^2]dB_d(r) , \]

\[ \xi_4 = \int_{\lambda^0}^1 [(r - 1)(3r - 2\lambda^0 - 1)/(1 - \lambda^0)^2]dB_d(r) , \]

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

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

Also define \( Z^*(v) \) as: \( Z^*(0) = 0 \), \( Z^*(v) = Z_1(v) \) for \( v < 0 \) and \( Z^*(v) = Z_2(v) \) for \( v > 0 \), with

\[ Z_1(v) = (\beta_0^0)^2|v|^3/3 + v^2\kappa(d)|\beta_0^0\xi_4 + v\kappa(d)^2[2\xi_2^0\Omega_1\xi_1 - \xi_1^0\Omega_2\xi_1] , \quad v < 0 , \]

\[ Z_2(v) = (\beta_0^0)^2|v|^3/3 + v^2\kappa(d)|\beta_0^0\xi_3 + v\kappa(d)^2[2\xi_2^0\Omega_1\xi_1 - \xi_1^0\Omega_2\xi_1] , \quad v > 0 . \]

Then, \( T^{1/2-d}(\bar{\lambda} - \lambda^0) \Rightarrow \arg\min_v Z^*(v) . \)

Theorem 3 implies that the limiting distributions have interesting qualitative differences across models. First, in Model I, even though the magnitude of the break is fixed, the limiting distributions of the estimate of the break fraction do not depend on the structure of the error process, except via \( \kappa(d) \) which is required to properly scale the distribution. This feature contrasts with results for stationary regressors. Bai (1997), among many others, showed that the limiting distribution of the estimate of the break fraction depends on the exact distributions of both the regressors and the errors in linear regression models with stationary regressors. To avoid this issue, the so-called shrinking shift framework was introduced, whereby the magnitude of break decreases as the sample
size increases. Theorem 3, however, shows that we do not have to rely on such a shrinking shift framework to obtain the limiting distributions with a joint-segmented trend.

Second, in Model II, the limiting distributions are functions of a two-sided random process in which many nuisance parameters are involved. In particular, when \( d^* \in (-0.5, 0.5) \), the limiting distribution depends on the exact distributions of the errors. On the other hand, for \( d^* \in (0.5, 1.5) \), the limiting distribution does not depend on the exact distribution of the errors. Hence, a confidence interval for the break date can be formed by estimating the nuisance parameters consistently and simulating the various functionals of the fractional Brownian motions.

Third, comparing Models I and II, we find that including a level shift component as a regressor, even if irrelevant, has an important effect on the asymptotic distributions. For illustrative purpose, assume that the DGP does not have a level shift, i.e., \( \mu_0 = 0 \). While Model I does not allow a level shift, Model II introduces a dummy variable \( C_t \) to incorporate an irrelevant level shift. From Theorem 2, we know that the rate of convergence of the estimated break fraction is slower in Model II when \( d^* \in (-0.5, 0.5) \). Furthermore, the asymptotic distributions are different across models. In Section 6, we provide simulation experiments to further analyze the implications of incorporating a level shift component.

4.4 The Limiting Distribution of the Estimates of the Other Parameters

We turn to the limiting distribution of the other parameter estimates in the models, that is, \((\hat{\mu}_1, \hat{\beta}_1, \hat{\beta}_b)\) for Model I, and \((\hat{\mu}_1, \hat{\mu}_b, \hat{\beta}_1, \hat{\beta}_b)\) for Model II.

**Theorem 4** Under assumption A1-A3, the following results hold for all \( d \in (-0.5, 0.5) \). 1) For Model I:

\[
\begin{bmatrix}
T^{1/2-d}(\hat{\mu}_1 - \mu_1^0) \\
T^{3/2-d}(\hat{\beta}_1 - \beta_1^0) \\
T^{3/2-d}(\hat{\beta}_b - \beta_b^0)
\end{bmatrix} \Rightarrow \Sigma_a^{-1}\Sigma_0 \quad \text{if } m = 0,
\]

\[
\begin{bmatrix}
T^{-1/2-d}(\hat{\mu}_1 - \mu_1^0) \\
T^{1/2-d}(\hat{\beta}_1 - \beta_1^0) \\
T^{1/2-d}(\hat{\beta}_b - \beta_b^0)
\end{bmatrix} \Rightarrow \Sigma_a^{-1}\Sigma_1 \quad \text{if } m = 1,
\]

where

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

10
\[
\Sigma_0 = \kappa(d) \left( \int_0^{\lambda_0} \left[ \begin{array}{c}
3(\lambda_0)^2 - 2\lambda_0 - 6\lambda_0 r + 6r \\
(\lambda_0)^2 - 2(\lambda_0)^2 r + 3r \\
- (\lambda_0)^2 (\lambda_0 + 3r)
\end{array} \right] dB_d(r) + \int_0^{1} \left[ \begin{array}{c}
-3(\lambda_0 + 1 - 2r) \\
-3(\lambda_0)^2 - 2 + 3\lambda_0 r + 4r \\
-2\lambda_0 + 4r - 2
\end{array} \right] dB_d(r) \right),
\]

and
\[
\Sigma_1 = \kappa(d) \left( \int_0^{\lambda_0} \left[ \begin{array}{c}
3(1-\lambda_0)^2 r^2 + (3\lambda_0 - 2)\lambda_0 r - (\lambda_0)^2 \\
(2\lambda_0)^2 r^2 - 2\lambda_0 (1-\lambda_0)^2 r - (\lambda_0)^2 \\
- (2\lambda_0)^2 (3\lambda_0 - 2)\lambda_0 r - (\lambda_0)^2
\end{array} \right] dB_d(r) + \int_0^{1} \left[ \begin{array}{c}
3\{r^2 - (1 + \lambda_0)r + \lambda_0\} \\
\lambda_0^2 \{r^2 - (1 + \lambda_0)r + \lambda_0\} \\
2\{r^2 - (1 + \lambda_0)r + \lambda_0\}
\end{array} \right] dB_d(r) \right).
\]

2) For Model II:
\[
\left[ \begin{array}{c}
T^{1/2-d}(\hat{\mu}_b - \mu_b^0) \\
T^{3/2-d}(\hat{\beta}_b - \beta_b^0)
\end{array} \right] \Rightarrow \kappa(d) \Omega_1 \left[ \begin{array}{c}
\int_0^{1} dB_d(r) \\
\int_0^{1} r dB_d(r) \\
\int_0^{1} dB_d(r)
\end{array} \right]
\]

Hence, \( \hat{\mu}_b \) is asymptotically unidentified and \( \hat{\mu}_b - \mu_b^0 \Rightarrow \beta_b^0 \arg\min_\nu S^*(\nu), \) as defined in Theorem 3:
\[
\left[ \begin{array}{c}
T^{-1/2-d}(\hat{\mu}_b - \mu_b^0) \\
T^{1/2-d}(\hat{\beta}_b - \beta_b^0)
\end{array} \right] \Rightarrow \kappa(d) \Omega_1 \left[ \begin{array}{c}
\int_0^{1} B_d(r) dr \\
\int_0^{1} r B_d(r) dr \\
\int_0^{1} B_d(r) dr
\end{array} \right]
\]

This implies that \( \hat{\mu}_b \) is asymptotically unidentified because \( T^{-1/2-d}(\hat{\mu}_b - \mu_b^0) - \beta_b^0 (\hat{T}_1 - T_1^0) \Rightarrow \xi_3 + \xi_4, \) where \( \xi_3 \) and \( \xi_4 \) are random variables defined in Theorem 3.

Note that except for the unidentified intercept shift \( \hat{\mu}_b, \) the other parameters, \( (\hat{\mu}_1, \hat{\beta}_1, \hat{\beta}_b) \), have the same stochastic order for Models I and II. As noted in PZ, the exact model specification does not matter if one wants to make asymptotic inference on these parameters.

5 Spurious Break

In this section, we consider the properties of the least square estimate of a structural break date when no break is present in the data generating process. Nunes et al. (1995) and Bai (1998) showed that the least square estimate of the break date can lead to a spurious break date when
the error is an \( I(1) \) process, in the sense that the estimate will not gather around either end of the sample. Kuan and Hsu (1998) considered a change in mean model for a fractionally integrated process with \( d^* \in (-0.5, 0.5) \) and showed that a spurious break can be estimated if \( d^* \in (0, 0.5) \).

Hsu and Kuan (2008) confirmed the possibility of estimating a spurious mean break if the series is a non-stationary fractionally integrated process, i.e., \( d^* \in (0.5, 1.5) \). Here, we consider the issue of spurious breaks in the context of Model II with a disjoint-segmented trend. The data generating process is specified as:

\[ y_t = \mu_1^0 + \beta_1^0 t + u_t, \]

with \( u_t = \Delta^{-m} \xi_t^\# \) for \( t = 0, \pm 1, \pm 2, \ldots \) as defined in (4). Let \( S(T_1) \) denote the sum of squared residuals related to a generic break date \( T_1 \), that is, \( S(T_1) = Y'(I - P_{T_1})Y \). We have that \( \hat{T}_1 = \arg \min_{T_1} S(T_1) = \arg \min_{T_1} \{ S(T_1) - \sum_{t=1}^{T} u_t^2 \} \) because \( \sum_{t=1}^{T} u_t^2 \) is independent of \( T_1 \). If no structural change is allowed, then

\[
M_T(T_1) = - \left( S(T_1) - \sum_{t=1}^{T} u_t^2 \right) = \left( \sum_{t=1}^{T_1} x_t u_{t+1} \right) \left( \sum_{t=1}^{T_1} x_t u_t \right)^{-1} \left( \sum_{t=T_1+1}^{T} x_t u_{t+1} \right) \left( \sum_{t=T_1+1}^{T} x_t u_t \right)^{-1} \left( \sum_{t=1}^{T_1} x_t u_{t+1} \right)^{-1} \left( \sum_{t=1}^{T_1} x_t u_{t+1} \right) \left( \sum_{t=T_1+1}^{T} x_t u_{t+1} \right) \left( \sum_{t=T_1+1}^{T} x_t u_t \right)^{-1} \left( \sum_{t=1}^{T_1} x_t u_{t+1} \right),
\]

where \( x_t = (1, t) \). Let \( M_T^*(T_1) \) be the normalized version of \( M_T(T_1) \), that is,

\[
M_T^*(T_1) = T^{-2(d+m)} M_T(T_1) = T^{-(d+m)} \left( \sum_{t=1}^{T_1} x_t u_{t+1} \right) \left( \sum_{t=1}^{T_1} x_t u_t \right)^{-1} \left( \sum_{t=T_1+1}^{T} x_t u_{t+1} \right) \left( \sum_{t=T_1+1}^{T} x_t u_t \right)^{-1} \left( \sum_{t=1}^{T_1} x_t u_{t+1} \right) \left( \sum_{t=T_1+1}^{T} x_t u_{t+1} \right) \left( \sum_{t=T_1+1}^{T} x_t u_t \right)^{-1} \left( \sum_{t=1}^{T_1} x_t u_{t+1} \right),
\]

where \( D_{2T} = \text{diag}\{T^{1/2}, T^{3/2}\} \) and \( m \in \{0, 1\} \). Note that \( \tilde{T}_1 = \arg \max_{T_1} M_T(T_1) = \arg \max_{T_1} M_T^*(T_1) \) because the normalization \( T^{-2(d+m)} \) does not depend on \( T_1 \). We start with the case \( d^* \in (-0.5, 0.5) \) with \( m = 1 \). If assumptions A1-A2 hold, we have

\[
M_T^*(T_1) \Rightarrow M^*(\lambda) \equiv \kappa(d^2) \{ G(\lambda)'Q(\lambda)^{-1}G(\lambda) + [G(1) - \lambda G(\lambda)]'[Q(1) - \lambda Q(\lambda)]^{-1}[G(1) - \lambda G(\lambda)] \} \quad (9)
\]

where \( G(\lambda) = (B_d(\lambda), \int_0^\lambda r dB_d(r))' \), \( G(1) - \lambda G(\lambda) = (B_d(1) - B_d(\lambda), \int_0^1 r dB_d(r))' \),

\[
Q(\lambda) = \begin{pmatrix} \lambda & \lambda^2/2 \\ \lambda^2/2 & \lambda^3/3 \end{pmatrix}, \quad Q(1) - Q(\lambda) = \begin{pmatrix} 1 - \lambda & (1 - \lambda^2)/2 \\ (1 - \lambda^2)/2 & (1 - \lambda^3)/3 \end{pmatrix}.
\]
Taqqu (1977) showed that for \( d^* \in (-0.5, 0.5) \), \( B_d(t) \), \( t \in \mathbb{R} \) satisfies the following law of iterated logarithms: for some positive constant \( c_1 \),

\[
\limsup_{t \to \infty} \frac{B_d(t)}{(c_1 t^{1+2d^*} \log \log t)^{1/2}} = 1 \quad \text{almost surely (a.s.).}
\]

Since \( B_d(t) \) is self-similar with self-similarity parameter \( 0.5 + d^* \), for any \( c_2 > 0 \) it satisfies, \( B_d(t) \stackrel{d}{=} c_2^{-(0.5+d^*)} B_d(c_2t) \), where \( \stackrel{d}{=} \) denotes equality in distribution. Applying the law of iterated logarithms to \( B_d(1/t) \) and self-similarity, we have

\[
\limsup_{t \to 0} \frac{B_d(t)}{(c_2 t^{1+2d^*} \log \log(1/t))^{1/2}} = 1 \quad \text{a.s.}
\]

Then, for \( d^* \in (-0.5, 0] \),

\[
\limsup_{\lambda \to 0} \frac{B_d(\lambda)}{\sqrt{\lambda}} = \limsup_{\lambda \to 0} \sqrt{c \lambda^{2d^*} \log \log(1/\lambda)} = \infty \quad \text{a.s.}
\]

It is easy to verify that \( B_d(1) - B_d(\lambda) \) is also a fractional Brownian motion \( B_d(s) \) with \( s = 1 - \lambda \). For \( d^* \in (-0.5, 0] \),

\[
\limsup_{\lambda \to 1} \frac{B_d(1) - B_d(\lambda)}{\sqrt{1-\lambda}} = \limsup_{s \to 0} \frac{B_d(s)}{\sqrt{s}} = \infty \quad \text{a.s.}
\]

On the other hand, for \( d^* \in (0, 0.5) \),

\[
\limsup_{\lambda \to 0} \frac{B_d(\lambda)}{\sqrt{\lambda}} = \limsup_{\lambda \to 1} \frac{B_d(1) - B_d(\lambda)}{\sqrt{1-\lambda}} = 0 \quad \text{a.s.}
\]

Since the above are almost sure limits, we can define \( M^*(0) \) and \( M^*(1) \) as the almost sure limit of \( M^*(\lambda) \) as \( \lambda \to 0, 1 \), respectively. Hence, with probability 1,

\[
M^*(0) = M^*(1) = 12 \kappa(d)^2 \left\{ \frac{1}{3} B_d(1)^2 - \left( \int_0^1 r dB_d(r) \right) B_d(1) + \left( \int_0^1 r dB_d(r) \right)^2 \right\}.
\]

**Theorem 5** Under assumptions A1-A2, i) for \( d^* \in (-0.5, 0] \), \( \limsup_{\lambda \to 0} M^*(\lambda) = \limsup_{\lambda \to 1} M^*(\lambda) = \infty \) a.s.; ii) for \( d^* \in (0, 0.5) \), there exist some \( \lambda \in (0, 1) \) s.t. \( M^*(0) = M^*(1) < M^*(\lambda) \) a.s.

Theorem 5 implies that no spurious break is estimated if the order of fractional integration is a value in \((-0.5, 0]\). It is not the case for \( d^* \in (0, 0.5) \) as \( M^*(0) = M^*(1) \) is stochastically bounded while \( M^*(\lambda) \) can be arbitrarily large with \( \lambda \)’s close to either ends.

Now, consider the possibility of estimating a spurious break date when the errors are non-stationary, i.e., \( d^* \in (0.5, 1.5) \). Here, Theorem 1 in Bai (1998) is generalized to incorporate a non-stationary fractional process with a deterministic trend.

**Proposition 1** Under assumptions A1-A2, for \( d^* \in (0.5, 1.5) \), \( \sup_{\lambda \in (0,1)} M^*(\lambda) = O_p(1) \).
Proposition 1 implies that $M^*(\lambda)$ is stochastically bounded even when $\lambda \to 0$ or $\lambda \to 1$ for $d^* \in (0.5, 1.5)$. Of interest is the limit behavior of $M^*(\lambda)$ when $\lambda$ gets closer to either 0 or 1. From (9), $M^*(0) = \kappa(d)^2G(1)^{-1}Q(1)^{-1}G(1)$ using the fact that $G(\lambda)^{1/2}Q(\lambda)^{-1}G(\lambda) \to 0$ as $\lambda \to 0$. The latter follows from the fact that $\kappa(d)^2G(\lambda)^{1/2}Q(\lambda)^{-1}G(\lambda)$ is the limit of the first term in $M^*_T(T_1)$. Then,

$$T^{-(d+1)} \left( \sum_{t=1}^{T_1} x_t' u_t \right)' \left( \sum_{t=1}^{T_1} x_t' x_t \right)^{-1} \left( \sum_{t=1}^{T_1} x_t' u_t \right) T^{-(d+1)} \leq T^{-2(d+1)} \sum_{t=1}^{[T\lambda]} u_t^2$$

and $T^{-2(d+1)} \sum_{t=1}^{[T\lambda]} u_t^2 \to 0$ as $\lambda \to 0$. Thus, $M^*(0) = \lim_{\lambda \to 0} M^*(\lambda)$, thereby $M^*(\lambda)$ is continuous at $\lambda = 0$. Similarly, $M^*(1) = \kappa(d)^2G(1)^{1/2}Q(1)^{-1}G(1)$ is defined as the limit of $M^*(\lambda)$ as $\lambda \to 1$.

**Theorem 6** Under assumptions A1-A2, for $d^* \in (0.5, 1.5)$, with probability 1, $M^*(0) = M^*(1) < M^*(\lambda)$ for every $0 < \lambda < 1$.

Theorem 6 implies that $M^*(\lambda)$ cannot attain a maximum at zero or one almost surely.

6 Simulation Experiments

In this section, we provide simulation experiments to illustrate various theoretical results. We first assess whether the asymptotic distributions are good approximations to the finite sample distributions. We highlight the bimodality of the distribution induced by an irrelevant level shift included in the regression. We also illustrate the spurious break problem.

6.1 Finite Sample and Limiting Distributions

We start with simulations showing that the finite sample distributions of various estimates are well approximated by their asymptotic distributions. Of interest are three estimates: $\hat{T}_1$ (break date), $\hat{\beta}_b$ (slope change), and $\hat{\mu}_b$ (level shift). Throughout, we use 2,000 replications and two sample sizes $T = 200$ and 800. Whenever the asymptotic distributions are non-normal, we use simulations of the fractional Brownian motion and estimates of the various parameters to evaluate the probability density function using a kernel-based method applied to the simulated realizations.

We first consider the following DGP:

$$y_t = x(T_1^0)\gamma_0 + u_t = \mu_0^1 + \beta_0^1 t + \beta_0^0 B_t + u_t,$$

where $u_t = \Delta^{-m}(\zeta I_{t \geq 1})$, $\zeta_t = \Delta^{-d} \eta_t$, $\eta_t \sim i.i.d. N(0, \sigma^2)$ for $t = 0, \pm 1, \ldots, m \in \{0, 1\}$, and $B_t = (t - T_1^0)\mathbb{1}_{t \geq [T\lambda^0]}$ with $T_1^0 = [T\lambda^0]$. We set the various parameters at the following values: $\lambda^0 = 0.5$, $\mu_1^0 = 1.72$, $\beta_1^0 = 0.03$, $\beta^0_b = -0.02$, $\sigma^2 = 0.1$ for stationary case, and $\sigma = 0.01$ for
non-stationary case. The configurations are the same as those in PZ, chosen to obtain distributions that easily reveal the main features of interest. Using DGP (10), we consider two regression models: the joint broken trend (Model I) and the local disjoint broken trend (Model II).

Figure 1 presents the finite sample and asymptotic probability density function (pdf) of the normalized estimates of the break date and the slope change when the order of fractional integration $d^* = 0.2$ and $\sigma^2 = 0.1$. For Model I, the normalized estimate of the break date is given by $T^{1/2-d}(\hat{T}_1 - T_1^0)$ and the normalized estimate of the slope change is $T^{3/2-d}(\hat{\beta}_b - \beta_b^0)$. Simulation results pertaining to Model I are in the top panels. The results reveal that the finite sample distribution is well approximated by the asymptotic distribution for both estimates. For Model II, the normalized estimates $(\hat{T}_1 - T_1^0)$ and $T^{3/2-d}(\hat{\beta}_b - \beta_b^0)$ are considered and the results are presented in the bottom panels. Since we set $\mu_b^0 = 0$, Model II incorporates an irrelevant level shift. We find that the finite sample distribution of the estimate of the break date is clearly bimodal. Furthermore, the asymptotic distribution is a good approximation to the finite sample distribution when $T = 800$ but less so when $T = 200$. For the slope change, when $T = 200$, the finite sample distribution is right-skewed. However, as the sample size increases, the finite sample distribution approaches the limiting distribution.

Figure 2 presents a similar set of results for the non-stationary case. The DGP is still (10) but with $d^* = 1.2$ and $\sigma = 0.01$. For both Models I and II, the normalized statistics are $T^{-1/2-d}(\hat{T}_1 - T_1^0)$ and $T^{1/2-d}(\hat{\beta}_b - \beta_b^0)$. When the regression from Model I is used (which is well specified), the asymptotic distribution is a good approximation to the finite sample distribution for these two parameters. On the other hand, when the regression from Model II is used (which introduces an irrelevant level shift regressor), the asymptotic distribution of the estimated break fraction exhibits a minor bimodal pattern that is not present in the finite sample distribution when $T = 200$ or $T = 800$. However, increasing $T$ to 1600, the finite sample distribution indeed also exhibits a slight bimodal pattern and the approximation is satisfactory. For the slope change, the asymptotic approximation is still good.

Figure 3 presents the finite sample and asymptotic distribution of the estimate of the level shift $\hat{\mu}_b$ from the regression Model II. The DGP is given by (10) where $d^* = 0.2$ and $\sigma^2 = 0.1$ in panel (a), while $d^* = 1.2$ and $\sigma = 0.01$ in panel (b). The values of the other parameters remain as stated above. In panel (a), with $d^* = 0.2$, the distributions are clearly bimodal and the approximation is quite satisfactory with $T = 800$. We explain the feature in detail below. When the errors are non-stationary with $d^* = 1.2$, we only plot in panel (b) the finite sample distributions, which show little changes between $T = 200$ and 800.

**Remark 2** In Theorem 2, we show that introducing a level shift regressor reduces the rate of con-
vergence of the estimate of the break fraction when the order of fractional integration $d^*$ is in $(-0.5, 0.5)$ and that the rate of convergence is invariant in Model II. Furthermore, it induces bimodality in the distribution of this estimate (both in finite samples and in the limit). PZ (Section 5) provided an intuitive explanation for this phenomenon. Since the level shift regressor $C_t$ can categorize random departure from the trend line around the true break date as a level shift, this induces increased randomness in the estimate of the break date. This is referred to as a “contamination”. In the proof of Theorem 4.(2), we show that $$\hat{b} = \beta_0 (\hat{T}_1 - T_0^0) + o_p(1)$$ when $d^* \in (-0.5, 0.5)$. Accordingly, since $\hat{T}_1$ is contaminated by the level shift regressor $C_t$, it also has an effect on the estimate $\hat{b}$, referred to as a “feedback” effect. Because of the “feedback” effect, the true parameter of the level shift $\mu_b^0$ cannot be identified.

Figures 4 and 5 consider the case with a genuine level shift, i.e., $\mu_b^0 \neq 0$. The DGP is

$$y_t = x(T_1^0) t^0 + u_t = \mu_t^0 + \beta_0^0 t + \mu_b^0 C_t + \beta_b^0 B_t + u_t, \quad (11)$$

where $C_t = 1_{t \geq T^0}$. We consider the regression from Model II to estimate the parameters. In Figure 4, we present the results pertaining to the case where the errors are stationary with $d^* = 0.2$ and $\sigma^2 = 0.1$. We make a comparison between the finite sample distribution and the limiting distribution derived in Theorem 4. In panel (a), $\mu_b^0 = -0.1$ and the asymptotic distribution shows strong bimodality (the right mode being more important). Moreover, the asymptotic distribution is a good approximation to the finite sample distribution when the sample size is large, $T = 800$, but less so with $T = 200$. In panel (b), $\mu_b^0 = 0.3$ and the left mode clearly dominates. The asymptotic distribution approximates the finite sample distribution better compared to the case where $\mu_b^0 = -0.1$. In panel (c), we plot a set of the asymptotic distributions changing the value of $\mu_b^0$. As the absolute value of $\mu_b^0$ increases, one mode dominates the other and is more centered around 0. This implies that a large level shift is helpful in identifying the true break date.

Figure 5 presents similar results for non-stationary errors with $d^* = 1.2$ and $\sigma = 0.01$. We set $\mu_b^0 = 0.01$ in panel (a), $-0.02$ in panel (b), and 0.05 in panel (c). To understand the implications of the results in Figure 5, first note that the level shift parameter $\mu_b^0$ does not appear in the limiting distribution of the estimated break date in Model II for $m = 1$, i.e., when $d^* \in (0.5, 1.5)$. As the sample size increases, the magnitude of the level shift $\mu_b^0$ is relatively small compared to the level of the trend function. In the limit, the level shift effect is concealed by random variations in the non-stationary errors. Hence, the asymptotic distribution is an appropriate approximation when the magnitude of the level shift $\mu_b^0$ is relatively small. The adequacy of the approximation decreases as $\mu_b^0$ increases. For a large level shift, we can expect that $\mu_b^0$ would affect the limiting distribution. PZ suggested to use an asymptotic expansion under this circumstance (see their Theorem 5).
6.2 Spurious Break

We consider simulation experiments to illustrate the issue of a potential spurious break. The data generating process is specified by

\[ y_t = \mu_1^0 + \beta_1^0 t + u_t, \]

where \( u_t = \Delta^{-m}(\zeta_t1_{t\geq1}), \zeta_t = \Delta^{-d}\eta_t, \eta_t \sim i.i.d.N(0,\sigma^2) \) for \( t = 0, \pm1, \ldots \) and \( m \in \{0,1\} \). Without loss of generality, we set \( \mu_1^0 = \beta_1^0 = 0 \) and consider \( d^* \in \{-0.3,0.2,0.7,1.2\} \). The sample sizes used are \( T = 200 \) and \( 2000 \). For each value of \( d^* \), the results are obtained from 10,000 replications. We consider estimating the date of a structural break using Model II (locally disjoint broken trend). Figure 6(a) presents histograms of the estimates \( \hat{T}_1 \) when \( T = 200 \). For \( d^* = -0.3 \), the estimates are concentrated at the two end points (1 and \( T \)) indicating that the estimate of the break date is consistent and no spurious break feature is present, consistent with Theorem 5. For \( d^* = 0.2 \), the histogram of the estimate \( \hat{T}_1 \) spreads out across all admissible break dates with the exception of the end points. For \( d^* \in \{0.7,1.2\} \), the estimates of the break date \( \hat{T}_1 \) tend to cluster near the middle of the sample, which falsely indicates that there is a break in the sample. Figure 6(b) presents histograms of the estimate \( \hat{T}_1 \) with \( T = 2000 \). With this larger sample, the estimates often occur near the boundaries, though there is no mass at or very near 0 or 1 with \( d^* \in \{0.2,0.7,1.2\} \). Hence, the theoretical results are supported by the simulations.

These results reinforce the feature discussed in the literature to the effect that structural change and long memory imply similar features in the data, and it is difficult to distinguish one from the other at least in small samples. This suggests the importance of implementing a proper testing procedure for a structural break which should be robust to any a priori unknown order of integration. Recently, Harvey et al. (2009) and Perron and Yabu (2009) suggested testing procedures for a structural change in trend function designed to be robust to \( I(0) \) or \( I(1) \) errors. Iacone et al. (2013) presented a sup-Wald type test for a change in the slope of a trend function which is robust across fractional values of the order of integration.

7 Conclusion

In this paper, we establish the consistency, rate of convergence and limit distributions of parameter estimates in models where the trend function experiences a slope change at some unknown date, with or without a concurrent level shift, when the errors are fractionally integrated processes with the order of fractional integration \( d^* \in (-0.5,0.5) \cup (0.5,1.5) \). It is worth noting that introducing a level shift has a crucial effect on the asymptotic results. Our theoretical results uncover some interesting features. First, when a concurrent level shift is allowed, the rate of convergence of the estimate of
the break date is slower, and it is the same for all values of $d^* \in (-0.5, 0.5)$. This feature is linked to the contamination induced by allowing a level shift. In all other cases, the rate of convergence is monotonically decreasing as $d^*$ increases. Second, the level shift coefficient $\mu_b$ is asymptotically unidentified when the errors are non-stationary fractionally integrated processes, i.e., $d^* \in (0.5, 1.5)$ while the slope change coefficient $\beta_b$ can be estimated consistently for all $d^* \in (-0.5, 0.5) \cup (0.5, 1.5)$. Third, we also provide results about the so-called spurious break issue and show that it cannot occur in the limit when $d^*$ in the interval $(-0.5, 0]$. Lastly, via the simulation experiments, we confirm that an irrelevant level shift induces bimodality in the distribution of the break date estimates, but a relevant level shift improves the precision of the estimate.

The results in this paper can be useful for subsequent work. For instance, Lobato and Velasco (2007) considered efficient Wald test for a unit root against a fractionally integrated process with unknown order. However, their procedure does not allow a break under both the null and alternative hypotheses. Accordingly, an interesting avenue would be to extend the Kim and Perron (2009) unit root testing procedure that allows a structural change in the trend function under both the null and alternative hypotheses. Just as the results of Perron and Zhu (2005) and Perron and Yabu (2009) were useful to achieve this task, one could use our results and those of Iacone et al. (2013) to extend the test of Lobato and Velasco (2007). This is currently the object of ongoing research.

Notes

1. The restriction that $d^* \neq 0.5$ is standard in the long memory literature (e.g. Iacone et al., 2013). Tanaka (1999) showed that the case with $d^* = 0.5$ needs to be treated separately from the case with $d^* \neq 0.5$.

2. To generate one dimensional fractional Brownian motion $B_d(t)$ on $t \in [0, 1]$, we use the MATLAB code “hurst.m” of Kroese and Botev (2013) that applies the Fast Fourier Transform (FFT) to a circulant covariance matrix.

3. For a set of statistics $\{x_i\}_{i=1,\ldots,n}$, the pdf at a value $x$ is estimated by $\hat{g}(x) = (Nh_x)^{-1} \sum_{i=1}^n K((x-x_i)/h_x)$ where $K(\cdot)$ is a kernel function and $h_x$ is the bandwidth. We use the standard normal kernel and $n = 2,000$. As mentioned in PZ, the cross-validation method for choosing the optimal bandwidth does not work well because the estimates of the break date are discrete integers. As a rule of thumb, the bandwidth is set to $h_x = 0.3\hat{\sigma}_x$ where $\hat{\sigma}_x$ is the estimated standard deviation of the sample statistics $\{x_i\}_{i=1,\ldots,n}$. 

18
References


Figure 1: Finite sample and asymptotic distributions in Models I and II with $d^* = 0.2$: $\mu_0^b = 0$. The statistics are normalized as follows: $T^{1/2-d}(\hat{T}_1 - T_1^0)$ for the break date in Model I (panel (a)) but $\hat{T}_1 - T_1^0$ in Model II (panel (c)); and $T^{3/2-d}(\hat{\beta}_b - \beta_0^b)$ for the slope change in both models (panels (b) and (d)). The finite sample distributions are obtained using $u_t = \zeta_t 1_{t \geq 1}$, $\zeta_t = \Delta^{-d} \eta_t$ and $\eta_t \sim i.i.d. N(0, \sigma^2)$ with 2,000 replications. The values of the parameters are set to $\lambda^0 = 0.5$, $\mu_0^1 = 1.72$, $\beta_0^1 = 0.03$, $\beta_0^b = -0.02$, $\sigma^2 = 0.1$. Because the limiting distributions are non-standard, we use 5,000 simulated values to construct the pdf.
Figure 2: Finite sample and asymptotic distributions in Models I and II with $d^* = 1.2$: $\mu^0_b = 0$. The statistics are normalized as follows: $T^{-1/2-d}(\hat{T}_1 - T^0_1)$ for both models (panels (a) and (c)); and $T^{1/2-d}(\hat{\beta}_b - \beta^0_b)$ for the slope change in both models (panels (b) and (d)). The finite sample distributions are obtained using $u_t = \zeta_t1_{t \geq 1}$, $\zeta_t = \Delta^{-d}\eta_t$ and $\eta_t \sim i.i.d. N(0, \sigma^2)$ with 2,000 replications. The values of the parameters are set to $\lambda^0 = 0.5, \mu^0_1 = 1.72, \beta^0_1 = 0.03, \beta^0_b = -0.02, \sigma = 0.01$. Because the limiting distributions are non-standard, we use 5,000 simulated values to construct the pdf.
Figure 3: Unidentified level shift in Model II: $\mu_0^b = 0$. The statistics are normalized as follows: $\hat{\mu}_b - \mu_0^b$ for $d^* = 0.2$ (panel (a)); and $T^{-1/2 - d}(\hat{\mu}_b - \mu_0^b)$ for $d^* = 1.2$ (panel (b)). The finite sample distributions are obtained using $u_t = \zeta_t 1_{t \geq 1}$, $\zeta_t = \Delta^{-d} \eta_t$ and $\eta_t \sim i.i.d. N(0, \sigma^2)$ with 2,000 replications. The values of the parameters are set to $\lambda_0 = 0.5, \mu_0^1 = 1.72, \beta_0^0 = 0.03, \beta_0^b = -0.02$. Moreover, $\sigma^2 = 0.1$ if $d^* = 0.2$ and $\sigma = 0.01$ if $d^* = 1.2$. Because the limiting distributions are non-standard, we use 5,000 simulated values to construct the pdf in panel (a).
Figure 4: Finite sample and asymptotic distributions in Model II with $d^* = 0.2$: $\mu_b^0 \neq 0$. The statistic for the break date is normalized as $\hat{T}_1 - T_1^0$. The finite sample distributions are obtained using $u_t = \zeta_t 1_{t \geq 1}$, $\zeta_t = \Delta - d \eta_t$ and $\eta_t \sim i.i.d. N(0, \sigma^2)$ with 2,000 replications. The values of the parameters are set to $\lambda^0 = 0.5, \mu_1^0 = 1.72, \beta_1^0 = 0.03, \beta_b^0 = -0.02, \sigma^2 = 0.1$. Because the limiting distributions are non-standard, we use 5,000 simulated values to construct the pdf. In Panel (a), where $\mu_b = -0.1$, 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 of $\hat{T}_1 - T_1^0$ varying $\mu_b^0$. 
Figure 5: Finite sample and asymptotic distributions in Model II with $d^* = 1.2$: $\mu_b^0 \neq 0$. The statistic for the break date is normalized as $T^{-0.5-d}(\hat{T}_1 - T_0^0)$. The finite sample distributions are obtained using $u_t = \zeta_t1_{t \geq 1}$, $\zeta_t = \Delta^{-d}\eta_t$ and $\eta_t \sim i.i.d.N(0, \sigma^2)$ with 2,000 replications. The values of the parameters are set to $\lambda^0 = 0.5, \mu^0_1 = 1.72, \beta^0_1 = 0.03, \beta^0_b = -0.02, \sigma = 0.01$. Because the limiting distributions are non-standard, we use 5,000 simulated values to construct the pdf. In Panel (a), where $\mu_b = 0.01$, we compare the finite sample distributions for $T = 200$ and 800 against the limiting distribution; in Panel (b), we let $\mu_b^0 = -0.02$; in Panel (c), $\mu_b^0 = 0.05$. 
Figure 6: Empirical distributions of $\hat{T}_1$ when there is no change
Supplementary Material to “Inference on a Structural Break in Trend with Fractionally Integrated Errors”

Seong Yeon Chang\(^\dagger\) and Pierre Perron\(^\ddagger\)

\(^\dagger\)Wang Yanan Institute for Studies in Economics (WISE), Xiamen University
\(^\ddagger\)Department of Economics, Boston University

We consider the proofs of Theorems 1-4 for Models I and II separately, for ease of exposition. The proofs of Theorems 5-6 and Proposition 1 related to a spurious break are also provided. Note that here and throughout the text, “\(\rightarrow\)” denotes uniform convergence of a sequence of non-random elements, “\(\mathbb{P}\)” convergence in probability, and “\(\Rightarrow\)” weak convergence in the space \(\mathcal{D}[0, 1]\) under the Skorohod topology. We use the label \(O(T^a)\) and \(O_p(T^a)\) in its strict sense, i.e., it cannot imply \(o(T^a)\) and \(o_p(T^a)\), respectively. All limit statements are taken as \(T \to \infty\). We start with the following lemma.

**Lemma A.1** Define

\[
\begin{align*}
(XX) &= \gamma^0(X_{T_1}^0 - X_{T_1})(I - P_{T_1})(X_{T_1}^0 - X_{T_1})
\gamma^0 \\
(XU) &= \gamma^0(X_{T_1}^0 - X_{T_1})(I - P_{T_1})U \\
(UU) &= U'(P_{T_1}^0 - P_{T_1})U.
\end{align*}
\]

Under Assumptions A1-A3, the following results hold for all \(d \in (-0.5, 0.5)\) uniformly over all generic \(T_1 \in [\pi T, (1 - \pi)T]\) for some arbitrarily small \(\pi\) such that \(\lambda^0 \in [\pi, 1 - \pi]\).

1) Model I: a) if \(m = 0\):

\[
\begin{align*}
(XX) &= |T_1 - T_1^0|^2 O(T) \\
(XU) &= |T_1 - T_1^0| O_p(T^{1/2 + d}) \\
(UU) &= |T_1 - T_1^0| O_p(T^{-1/2 + d}).
\end{align*}
\]

b) if \(m = 1\):

\[
\begin{align*}
(XX) &= |T_1 - T_1^0|^2 O(T) \\
(XU) &= |T_1 - T_1^0| O_p(T^{3/2 + d}) \\
(UU) &= |T_1 - T_1^0| O_p(T^{1+2d}).
\end{align*}
\]

2) For Model II: a) if \(m = 0\):

\[
\begin{align*}
(XX) &= |T_1 - T_1^0|^3 O(1) \\
(XU) &= |T_1 - T_1^0|^{3/2 + d} O_p(1) \\
(UU) &= |T_1 - T_1^0|^{1/2 + d} O_p(T^{-1/2 + d}).
\end{align*}
\]

b) if \(m = 1\)

\[
\begin{align*}
(XX) &= |T_1 - T_1^0|^3 O(1) \\
(XU) &= |T_1 - T_1^0|^2 O_p(T^{1/2 + d}) \\
(UU) &= |T_1 - T_1^0| O_p(T^{1+2d}).
\end{align*}
\]
A.1 Results for Model I

Model I can be represented in matrix notation as

\[ Y = X_{T_1} \gamma + U = \begin{bmatrix} \mu_1 \\ t \ B_{T_1} \end{bmatrix} + U \]

where \( Y = (y_1, \ldots, y_T)' \), \( U = (u_1, \ldots, u_T)' \), \( t = (1, \ldots, 1)' \), \( t = (1, 2, \ldots, T)' \), \( B_{T_1} = (B_1, \ldots, B_T)' \), and \( \gamma = (\mu_1, \beta_1, \beta_b)' \). Note that the matrix \( X_{T_1} \) depends on the candidate break date \( T_1 \). In the proof, we only consider the case \( T_1 > T_0 \). It is straightforward to apply the same arguments to the case where \( T_1 < T_0 \). For \( T_1 > T_0 \), let

\[
\bar{e}_b(t) = \begin{cases} 
0 & \text{if } 1 \leq t \leq T_1^0 \\
\frac{t-T_1^0}{T_1 - T_1^0} & \text{if } T_1^0 < t < T_1 \\
1 & \text{if } T_1 \leq t \leq T,
\end{cases}
\]

and for \( T_1 = T_0 \), let

\[
\tilde{e}_b(t) = e_b(t) = \begin{cases} 
0 & \text{if } 1 \leq t \leq T_1^0 \\
1 & \text{if } T_1^0 < t \leq T.
\end{cases}
\]

For ease of notation, we suppress \( t \) and write \( \bar{e}_b \). With this notation, we can write

\[
(X_{T_1}^0 - X_{T_1}) \gamma^0 = \beta_b^0 (T_1 - T_1^0) \bar{e}_b.
\]

Note that \( \bar{e}_b([T_r]) \) converges to a continuous function \( f_{\bar{e}_b}(r) \) over \([0, 1] \) defined by, for \( \lambda > \lambda^0 \),

\[
f_{\bar{e}_b}(r) = \begin{cases} 
0 & \text{if } 0 \leq r \leq \lambda^0 \\
\frac{r-\lambda^0}{\lambda - \lambda^0} & \text{if } \lambda^0 < r < \lambda \\
1 & \text{if } \lambda \leq r \leq 1,
\end{cases}
\]

and by, for \( \lambda = \lambda^0 \),

\[
f_{\bar{e}_b}(r) = f_{\bar{e}_b}(r) = \begin{cases} 
0 & \text{if } 0 \leq r \leq \lambda^0 \\
1 & \text{if } \lambda^0 < r \leq 1.
\end{cases}
\]

Pertaining to the proof of Lemma A.1, we first consider the term \((XX)\). We have

\[
(XX) = \gamma^0 (X_{T_1}^0 - X_{T_1})'(I - P_{T_1})(X_{T_1}^0 - X_{T_1}) \gamma^0 = (T_1 - T_1^0)^2 (\beta_b^0)^2 \bar{e}_b(I - P_{T_1}) \bar{e}_b
\]

where the second equality holds because the first two columns of \((X_{T_1}^0 - X_{T_1})\) are zeros by construction. Note that \( \bar{e}_b(I - P_{T_1}) \bar{e}_b \) is the sum of squared residuals from a regression \( \bar{e}_b \) on \([i \ t \ B_{T_1}] \). Define

\[
S_T = \bar{e}_b(I - P_{T_1}) \bar{e}_b.
\]
Next, consider the continuous time least-squares projection of the function $f_{ib}(r)$ on $(1, r, f_B(r))$, where $f_B(r) = (r - \lambda)1_{r \geq \lambda}$. Let $(\hat{\alpha}, \hat{\beta}, \hat{\psi})$ denote the estimates of the coefficients and let $S_\infty$ denote the resulting SSR. From the definition of a Riemann integral, $T^{-1}S_T \rightarrow S_\infty$, where

$$ S_\infty = \int_0^1 \left( f_{ib}(r) - \hat{\alpha} - \hat{\beta}r - \hat{\psi}f_B(r) \right)^2 dr. $$

If $\hat{\alpha} = \hat{\beta} = 0$, then $S_\infty > 0$ from the definitions of $f_{ib}(r)$ and $f_B(r)$. Otherwise, we have

$$ S_\infty \geq \int_0^{\min\{\lambda, \lambda^0\}} \left( f_{ib}(r) - \hat{\alpha} - \hat{\beta}r - \hat{\psi}f_B(r) \right)^2 dr = \int_0^{\min\{\lambda, \lambda^0\}} (\hat{\alpha} + \hat{\beta}r)^2 dr > 0 $$

where the equality holds due to the definitions on $f_{ib}(r)$ and $f_B(r)$ and the fact that both $\lambda$ and $\lambda^0$ are bounded away from zero. Moreover, $S_\infty$ is bounded uniformly in $\lambda \in (0, 1)$ since

$$ S_\infty < \int_0^1 (f_{ib}(r) + |\hat{\alpha}| + |\hat{\beta}|r + |\hat{\psi}|f_B(r))^2 dr < \sup_{\lambda \in (0, 1)} (1 + |\hat{\alpha}| + |\hat{\beta}| + |\hat{\psi}|)^2 \equiv K, $$

where the second inequality holds due to the definitions on $f_{ib}(r)$ and $f_B(r)$ and using the fact that $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\psi}$ (which are implicitly functions of $\lambda$) are bounded uniformly in $\lambda \in (0, 1)$ whose proof is straightforward but tedious and, hence, omitted. Therefore, $S_T = O(T)$ uniformly in $T_1 \in [T\pi, T(1 - \pi)]$. Intuitively, $S_\infty$ is bounded and positive being a continuous time SSR. Accordingly, $$(XX) = (T_1 - T_1^0)^2(\beta_b^0)^2O(T)$$

uniformly in $T_1 \in [T\pi, T(1 - \pi)]$. Next, we consider the term $(XU)$. We have

$$(XU) = \gamma^0(XT_1^0 - XT_1)'(I - P_{T_1})U = \beta_b^0(T_1 - T_1^0)\gamma_t^0(I - P_{T_1})U.$$  

Define $\tilde{f}_{ib}(r)$ as the projection residuals from a least-squares regression of $f_{ib}(r)$ on $(1, r, f_B(r))$. Under assumptions A1-A3, by the continuous mapping theorem, we have in $\lambda \in (0, 1)$

$$T^{-(d+1)/2} \tilde{T}_{ib}(I - P_{T_1})U \Rightarrow \kappa(d) \int_0^1 \tilde{f}_{ib}(r)dB_d(r) \quad \text{if } m = 0.$$  

Define $F_{ib}^*(r) = \int_0^r \tilde{f}_{ib}(s)ds$. By the properties of orthogonal projections and the result for $(XX)$,

$$F_{ib}^*(1) = \int_0^1 \tilde{f}_{ib}(r)dr = \int_0^1 \left( f_{ib}(r) - \hat{\alpha} - \hat{\beta}r - \hat{\psi}f_B(r) \right) dr = 0,$$

hence,

$$T^{-(d+3/2)} \tilde{T}_{ib}(I - P_{T_1})U \Rightarrow \kappa(d) \int_0^1 \tilde{f}_{ib}(r)B_d(r)dr = \kappa(d) \left[ [B_d(r)F_{ib}^*(r)]_0 - \int_0^1 F_{ib}^*(r)dB_d(r) \right]$$

$$= -\kappa(d) \left( \int_0^1 F_{ib}^*(r)dB_d(r) \right) \quad \text{if } m = 1.$$  

3
Note that, by the continuous mapping theorem, for $T_1 = [\lambda T]$ and any $M > 0$,

$$P\left( \sup_{\lambda \in (0, 1)} |T^{-(d+1)/2}T_{t_b}^i(I - P_{T_1})U| > M \right) \Rightarrow P\left( \sup_{\lambda \in (0, 1)} \kappa(d) \int_0^1 \bar{f}_{t_b}(r)d_B(r) > M \right)$$

if $m = 0$ and

$$P\left( \sup_{\lambda \in (0, 1)} |T^{-(d+3)/2}T_{t_b}^i(I - P_{T_1})U| > M \right) \Rightarrow P\left( \sup_{\lambda \in (0, 1)} | - \kappa(d) \left( \int_0^1 F_{t_b}^*(r)d_B(r) \right) | > M \right)$$

if $m = 1$. We show the uniform boundedness of $\int_0^1 \bar{f}_{t_b}(r)d_B(r)$ and $\int_0^1 F_{t_b}^*(r)d_B(r)$ in $\lambda \in (0, 1)$.

$$\sup_{\lambda \in (0, 1)} \int_0^1 \bar{f}_{t_b}(r)d_B(r) \leq \sup_{\lambda \in (0, 1)} \int_0^1 |\bar{f}_{t_b}(r)|d_B(r) < \sqrt{K} \int_0^1 d_B(r) = O_p(1),$$

and

$$\sup_{\lambda \in (0, 1)} \int_0^1 F_{t_b}^*(r)d_B(r) \leq \sup_{\lambda \in (0, 1)} \int_0^1 |F_{t_b}^*(r)|d_B(r) \leq \sup_{\lambda \in (0, 1)} \int_0^1 |\bar{f}_{t_b}(s)|ds d_B(r) \leq \sqrt{K} \int_0^1 d_B(r) = O_p(1),$$

where $K = \sup_{\lambda \in (0, 1)}(1 + |\hat{\alpha}| + |\hat{\beta}| + |\hat{\psi}|)^2$ as defined above. Therefore, $\int_0^1 \bar{f}_{t_b}(r)d_B(r) = O_p(1)$ and $\int_0^1 F_{t_b}^*(r)d_B(r) = O_p(1)$ uniformly in $\lambda \in [\pi, 1 - \pi]$, which implies

$$T_{t_b}^i(I - P_{T_1})U = \begin{cases} 
O_p(T^{d+1/2}) & \text{if } m = 0, \\
O_p(T^{d+3/2}) & \text{if } m = 1,
\end{cases}$$

uniformly in $\lambda \in [\pi, 1 - \pi]$. Accordingly,

$$(XU) = \begin{cases} 
\beta^0_0(T_1 - T_1^0)O_p(T^{d+1/2}) & \text{if } m = 0, \\
\beta^0_0(T_1 - T_1^0)O_p(T^{d+3/2}) & \text{if } m = 1,
\end{cases}$$

uniformly in $\lambda \in [\pi, 1 - \pi]$. Finally, we consider the term $(UU)$. Define $D_T = diag(T^{1/2+d}, T^{3/2+d}, T^{3/2+d})$ with $d \in (-0.5, 0.5)$. We have

$$(UU) = U'(P_{T_1^0} - P_{T_1})U = U'(X_{T_1^0}^r(X_{T_1^0}^r)^{-1}X_{T_1^0}^r - X_{T_1^0}(X_{T_1^0}^rX_{T_1^0})^{-1}X_{T_1^0}^r)U$$

$$= U'(X_{T_1^0}^r - X_{T_1^0})D_T^{-1}[D_T^{-1}X_{T_1^0}^rX_{T_1^0}D_T^{-1}]^{-1}D_T^{-1}X_{T_1^0}^rU$$

$$+ U'X_{T_1^0}^rD_T^{-1}[D_T^{-1}X_{T_1^0}^rX_{T_1^0}D_T^{-1}]^{-1}D_T^{-1}[X_{T_1^0}^rX_{T_1^0} - X_{T_1^0}^rX_{T_1^0}]D_T^{-1}[D_T^{-1}X_{T_1^0}^rX_{T_1^0}D_T^{-1}]^{-1}D_T^{-1}X_{T_1^0}^rU$$

$$+ U'X_{T_1^0}^rD_T^{-1}[D_T^{-1}X_{T_1^0}^rX_{T_1^0}D_T^{-1}]^{-1}D_T^{-1}(X_{T_1^0}^r - X_{T_1^0})U.$$
Also, from Lemma 2, we have for \( m = 1 \),

\[
T^{-\frac{d+3}{2}} \sum_{t=1}^{T} tu_t \Rightarrow \kappa(d)[B_d(1) - \int_{0}^{1} B_d(r)dr] = \kappa(d) \int_{0}^{1} r dB_d(r).
\]

In addition, it is easy to show that

\[
T^{-\frac{d+3}{2}} \sum_{t=1}^{T} u_t \Rightarrow \kappa(d) \int_{0}^{1} B_d(r)dr,
\]

\[
T^{-\frac{d+5}{2}} \sum_{t=1}^{T} tu_t \Rightarrow \kappa(d) \int_{0}^{1} r dB_d(r).
\]

with each result holding uniformly in \( \lambda \in (0, 1) \). We next consider each term in \((UU)\).

1. \( D_T^{-1} X'_T X_T D_T^{-1} \) and \( D_T^{-1} X'_T X_T D_T^{-1} \) are \( O(T^{-2d}) \) uniformly in \( \lambda \).

2. When \( m = 0 \),

\[
D_T^{-1} X'_T U = \begin{bmatrix}
T^{-\frac{d+1}{2}} \sum_{t=1}^{T} u_t \\
T^{-\frac{d+3}{2}} \sum_{t=1}^{T} tu_t \\
T^{-\frac{d+3}{2}} \sum_{t=T_1+1}^{T}(t - T_1)u_t
\end{bmatrix}
\Rightarrow \begin{bmatrix}
\kappa(d) B_d(1) \\
\kappa(d) \int_{0}^{1} r dB_d(r) \\
\kappa(d) \int_{\lambda}^{1} (r - \lambda) dB_d(r)
\end{bmatrix}
\]

and \( D_T^{-1} X'_T U \) and \( D_T^{-1} X'_T U \) are \( O_p(1) \) uniformly in \( \lambda \) since

\[
P \left( \sup_{\lambda \in (0,1)} |T^{-\frac{d+3}{2}} \sum_{t=\lfloor \lambda T \rfloor +1}^{T} (t - \lfloor \lambda T \rfloor)u_t| > M \right) \Rightarrow P \left( \sup_{\lambda \in (0,1)} |\kappa(d) \int_{\lambda}^{1} (r - \lambda) dB_d(r)| > M \right)
\]

and \( \sup_{\lambda \in (0,1)} |\kappa(d) \int_{\lambda}^{1} (r - \lambda) dB_d(r)| \) has a well-defined bounded distribution. When \( m = 1 \),

\[
T^{-1} D_T^{-1} X'_T U = \begin{bmatrix}
T^{-\frac{d+3}{2}} \sum_{t=1}^{T} u_t \\
T^{-\frac{d+5}{2}} \sum_{t=1}^{T} tu_t \\
T^{-\frac{d+5}{2}} \sum_{t=T_1+1}^{T}(t - T_1)u_t
\end{bmatrix}
\Rightarrow \begin{bmatrix}
\kappa(d) \int_{0}^{1} B_d(r)dr \\
\kappa(d) \int_{0}^{1} r dB_d(r)dr \\
\kappa(d) \int_{\lambda}^{1} (r - \lambda) B_d(r)dr
\end{bmatrix}
\]
and $D_T^{-1}X_t'U$ and $D_T^{-1}X_T'_{T_1}U$ are $O_p(T)$ uniformly in $\lambda$, since

$$P\left(\sup_{\lambda \in (0,1)} |T^{-(d+5/2)} \sum_{t=\lfloor \lambda T \rfloor + 1}^T (t - [\lambda T])u_t| > M\right) \Rightarrow P\left(\sup_{\lambda \in (0,1)} |\kappa(d) \int_\lambda^1 (r - \lambda)B_d(r)dr| > M\right)$$

and $\sup_{\lambda \in (0,1)} |\kappa(d) \int_\lambda^1 (r - \lambda)B_d(r)dr|$ has a well-defined bounded distribution.

3. $U'(X_{T_2} - X_{T_1})D_T^{-1}$. It suffices to consider the third column of $(X_{T_2} - X_{T_1})$ because the first two columns are zeros. We have

$$T^{-(d+1/2+m)}U'(B_{T_1}^0 - B_{T_1}) = T^{-(d+1/2+m)} \sum_{t=T_1}^T (t - T_1^0)u_t$$

$$+ T^{-(d+1/2+m)}(T_1 - T_1^0) \sum_{t=T_1}^T u_t$$

$$= |T_1 - T_1^0|O_p(1) \quad \text{for } m \in \{0, 1\}.$$  

4. $D_T^{-1}[X_t'X_{T_1} - X_{T_1}']D_T^{-1}$. As noted earlier, it suffices to consider the terms in which $B_{T_1}$ and $B_{T_1}^0$ are involved. We have:

$$B_{T_1}'B_{T_1}^0 - B_{T_1}B_{T_1} = |T_1 - T_1^0|O(T^2),$$

$$B_{T_1}'t - B_{T_1}t = |T_1 - T_1^0|O(T^2),$$

$$B_{T_1}'t - B_{T_1}t = |T_1 - T_1^0|O(T),$$

hence, we have

$$D_T^{-1}[X_t'X_{T_1} - X_{T_1}']D_T^{-1} = |T_1 - T_1^0|O(T^{-(1+2d)}), \quad \text{for } m \in \{0, 1\}.$$  

Based on the results 1-4,

$$(UU) = \begin{cases} |T_1 - T_1^0|O_p(T^{-1+2d}) & \text{if } m = 0, \\ |T_1 - T_1^0|O_p(T^{1+2d}) & \text{if } m = 1, \end{cases}$$

uniformly in $\lambda \in [\pi, 1 - \pi]$. This completes the proof of Lemma A.1 for Model I.

A.1.1 Proof of Consistency (Theorem 1)

From the proof of Lemma A.1, we know that for Model I, if $m = 0,$

$$(\hat{X}\hat{X}) = (T_1^0 - \hat{T}_1)^2O(T)$$

$$(\hat{X}\hat{U}) = (T_1^0 - \hat{T}_1)O_p(T^{1/2+d})$$

$$(\hat{U}\hat{U}) = |T_1^0 - \hat{T}_1|O_p(T^{-1+2d}),$$

and if $m = 1,$
and, if \( m = 1 \),
\[
\begin{align*}
(\hat{XX}) &= (T_1^0 - \hat{T}_1)^2O(T) \\
(\hat{XU}) &= (T_1^0 - \hat{T}_1)O_p(T^{3/2+d}) \\
(\hat{UU}) &= |T_1^0 - \hat{T}_1|O_p(T^{1+2d}),
\end{align*}
\]
uniformly in \( \lambda \in [\pi, 1-\pi] \). \( \hat{XX} \) is positive provided the event \( \{T_1^0 = \hat{T}_1\} \) does not hold for all \( T \), which is the case if \( \hat{\lambda} \) does not converge in probability to \( \lambda^0 \). We consider the proof for \( m = 0 \) (the proof for \( m = 1 \) is similar). Suppose that \( \hat{\lambda} \) does not converge in probability to \( \lambda^0 \). Then, the results above imply that \( \hat{XX} = O_p(T^3) \), \( \hat{XU} = O_p(T^{3/2+d}) \), and \( \hat{UU} = O_p(T^{2d}) \) for \( \lambda \in (-0.5, 0.5) \).

Therefore, for sufficiently large \( T \), the term \( \hat{XX} \) dominates the others with some probability. It implies that the key inequality \( \hat{XX} + 2\hat{XU} + \hat{UU} \leq 0 \) cannot hold with probability 1. Since this inequality is valid for all \( T \), we have a contradiction. Hence, we can conclude that \( \hat{\lambda} \xrightarrow{P} \lambda^0 \).

### A.1.2 Rate of Convergence (Theorem 2)

Consider the set
\[
V(\epsilon) = \{T_1 : |T_1 - T_1^0| < \epsilon T, \forall \epsilon > 0\}.
\]

From the consistency of \( \hat{T}_1 \) in Theorem 1, \( \Pr(\hat{T}_1 \in V(\epsilon)) \to 1 \) as \( T \to \infty \). Hence, it suffices to consider the behavior of \( S(T_1) \) for all \( T_1 \in V(\epsilon) \). Consider another set \( V_c(\epsilon) \) defined by
\[
V_c(\epsilon) = \{T_1 : |T_1 - T_1^0| < \epsilon T \quad \text{and} \quad |T_1 - T_1^0| > CT^{-1/2+d+m}, \forall \epsilon > 0, \forall d \in (-0.5, 0.5)\},
\]
for \( m = \{0, 1\} \). Note that \( V_c(\epsilon) \subseteq V(\epsilon) \). Since \( S(\hat{T}_1) \leq S(T_1^0) \) with probability 1, it suffices to show that \( \hat{T}_1 \notin V_c(\epsilon) \) by showing that for each \( \eta > 0 \), there exists a constant \( C > 0 \) such that
\[
\Pr\left( \min_{T_1 \in V_c(\epsilon)} \{S(T_1) - S(T_1^0)\} \leq 0 \right) < \eta. \tag{A.1}
\]

Equation (A.1) is equivalent to
\[
\Pr\left( \min_{T_1 \in V_c(\epsilon)} \{(XX) + 2(XU) + (UU)\} \leq 0 \right) < \eta.
\]

Based on the results derived in Lemma A.1, we can apply the following normalizations to these three terms in the set \( V_c(\epsilon) \). If \( m = 0 \), then
\[
\begin{align*}
\frac{(XX)}{|T_1 - T_1^0|T^{1/2+d}} &= \frac{|T_1 - T_1^0|^2O(T)}{|T_1 - T_1^0|T^{1/2+d}} > \frac{CT^{-1/2+d}O(T)}{T^{-1/2+d}T} > aC + o(1), \\
\frac{(XU)}{|T_1 - T_1^0|T^{1/2+d}} &= \frac{|T_1 - T_1^0|dO_p(T^{1/2+d})}{|T_1 - T_1^0|T^{1/2+d}} = O_p(1), \\
\frac{(UU)}{|T_1 - T_1^0|T^{1/2+d}} &= \frac{|T_1 - T_1^0|dO_p(T^{-1/2+d})}{|T_1 - T_1^0|T^{1/2+d}} = o_p(1).
\end{align*}
\]
If \( m = 1 \), then
\[
\frac{(XX)}{|T_1 - T_1^0|^2 + d} = \frac{|T_1 - T_1^0|^2 O(T)}{|T_1 - T_1^0|^2 + d} > \frac{CT^{1/2+d}O(T)}{T^{1/2+d}T} > aC + o(1),
\]
\[
\frac{(XU)}{|T_1 - T_1^0|^2 + d} = \frac{|T_1 - T_1^0|O_p(T^{3/2+d})}{|T_1 - T_1^0|^2 + d} = O_p(1),
\]
\[
\frac{(UU)}{|T_1 - T_1^0|^2 + d} = \frac{|T_1 - T_1^0|O_p(T^{1+2d})}{|T_1 - T_1^0|^2 + d} = o_p(1),
\]
where \( a \) is a positive constant. Here, we simply use the fact that \( |T_1 - T_1^0| < \epsilon T \) and \( |T_1 - T_1^0| > CT^{-1/2+d+m} \) in \( V_0(\epsilon) \). For a given \( \epsilon \), we can choose a constant \( C \) that is large enough to satisfy (A.1). Therefore, \( T_1 \) cannot be in \( V_0(\epsilon) \), which implies that for every \( \epsilon > 0 \), there exists a \( C > 0 \) such that \( \Pr((\hat{\lambda} - \lambda_0) \geq CT^{-3/2+d+m}) < \epsilon \) for sufficiently large \( T \).

A.1.3 Limiting Distribution of the Estimate of the Break Date

Consider first the case with \( m = 1 \). Define the set \( D(C) = \{ T_1 : |T_1 - T_1^0| < CT^{1/2+d} \} \), for some positive number \( C \), and \( m_T = T^{-1/2-d}|T_1 - T_1^0| \). We analyze
\[
\arg\min_{T_1 \in D(C)} [S(T_1) - S(T_1^0)].
\]
For \( T_1 \in D(C) \), we have \( |T_1 - T_1^0| = O(T^{1/2+d}) \). Hence, \( (XX) = |T_1 - T_1^0|^2 O(T) = O(T^{2+2d}) \), \( (XU) = |T_1 - T_1^0| O_p(T^{3/2+d}) = O_p(T^{2+2d}) \) and \( (UU) = |T_1 - T_1^0| O_p(T^{1+2d}) = O_p(T^{3/2+3d}) \). Then, for \( d \in (-0.5, 0.5) \),
\[
\arg\min_{T_1 \in D(C)} [S(T_1) - S(T_1^0)] = \arg\min_{T_1 \in D(C)} [(XX) + 2(XU) + (UU)]/T^{2+2d}
\]
\[
= \arg\min_{T_1 \in D(C)} [(XX)/T^{2+2d} + 2(XU)/T^{2+2d} + o_p(1)],
\]
hence we only need to consider the first two terms. Note that on the set \( D(C) \), \( |\lambda - \lambda_0| = O(T^{-1/2+d}) \) for \( d \in (-0.5, 0.5) \). Using this fact, we can derive the following results that will subsequently be used:

\[
T^{2d}D_T X_T' X_T D_T^{-1} = \begin{bmatrix}
1 & 1/2 & (1 - \lambda_0)^2/2 \\
1/2 & 1/3 & (1 - \lambda_0)^2(2 + \lambda_0)/6 \\
(1 - \lambda_0)^2/2 & (1 - \lambda_0)^2(2 + \lambda_0)/6 & (1 - \lambda_0)^3/3
\end{bmatrix} + o(1)
\]
\[
\equiv \Sigma_a + o(1),
\]
and the inverse is \( T^{-2d}(D_T^{-1} X_T' X_T D_T^{-1})^{-1} = \Sigma_a^{-1} + o(1) \) with
\[
\Sigma_a^{-1} = \begin{bmatrix}
(\lambda_0 + 3)/\lambda_0 & -3(\lambda_0 + 1)/(\lambda_0)^2 & 3/( (\lambda_0)^2(1 - \lambda_0) ) \\
-3(\lambda_0 + 1)/(\lambda_0)^2 & 3(3\lambda_0 + 1)/(\lambda_0)^3 & -3(2\lambda_0 + 1)/( (\lambda_0)^3(1 - \lambda_0) ) \\
3/( (\lambda_0)^2(1 - \lambda_0) ) & -3(2\lambda_0 + 1)/( (\lambda_0)^3(1 - \lambda_0) ) & 3/( (\lambda_0)^3(1 - \lambda_0) )
\end{bmatrix}.
\]
We have

\[
(XX) = (\beta_0^2)^2 (B_{T_1^0} - B_{T_1})' (I - P_{T_1})(B_{T_1^0} - B_{T_1})
= (\beta_0^2)^2 \{(B_{T_1^0} - B_{T_1})' (B_{T_1^0} - B_{T_1}) - (B_{T_1^0} - B_{T_1})' X_{T_1} D_T^{-1} (D_T^{-1} X_{T_1} X_{T_1} D_T^{-1})^{-1} D_T^{-1} X_{T_1} (B_{T_1^0} - B_{T_1})\}.
\]

Consider first the second term in \((XX)\).

\[
T^{-1}(B_{T_1^0} - B_{T_1})' X_{T_1} D_T^{-1} = |T_1 - T_1^0| T^{-1/2 - d} T^{-1/2 + d} t_b X_{T_1} D_T^{-1}
= m_T \left[ 1 - \lambda^0 \left( \frac{1 - \lambda^0}{2} \right) \left( \frac{1 - \lambda^0}{2} \right) \right] + o(1)
\]

where \(m_T = T^{-1/2 - d} |T_1 - T_1^0| \). Using the results above,

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

Hence,

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

and

\[
T^{-2-2d}(B_{T_0} - B_{T_1})' (B_{T_1^0} - B_{T_1}) = T^{-2-2d} |T_1 - T_1^0| \gamma t_b \gamma t_b = m_T^2 T^{-1} \gamma t_b + (1 - \lambda^0) m_T^2 + o(1).
\]

Combining (A.2) and (A.3), we obtain

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

Next,

\[
(XU) = \gamma (X_{T_1^0} - X_{T_1})' (I - P_{T_1}) U = \beta_0 (B_{T_1^0} - B_{T_1})' (I - P_{T_1}) U.
\]

We have,

\[
T^{-2-2d}(B_{T_1^0} - B_{T_1})' U = |T_1 - T_1^0| T^{-1/2 - d} T^{-3/2 - d} t_b t_b U
= m_T \kappa(d) \int_0^1 (1 - r/\lambda) B_d(r) dr + o_p(1),
\]

\[
T^{-2-2d}(B_{T_1^0} - B_{T_1})' X_{T_1} (X_{T_1} X_{T_1})^{-1} X_{T_1} U
= T^{-1}(B_{T_1^0} - B_{T_1})' X_{T_1} D_T^{-1} (D_T^{-1} X_{T_1} X_{T_1} D_T^{-1})^{-1} D_T^{-1} X_{T_1} U
= T^{-1}(B_{T_1^0} - B_{T_1})' X_{T_1} D_T^{-1} T^{-2d}(D_T^{-1} X_{T_1} X_{T_1} D_T^{-1})^{-1} T^{-1} D_T^{-1} X_{T_1} U,
\]

and

\[
T^{-1} D_T^{-1} X_{T_1} U = T^{-1} \left[ T^{-1/2 - d} \sum_{t=1}^T u_t T^{-3/2 - d} \sum_{t=1}^T t u_t T^{-3/2 - d} \sum_{t=T+1}^T (t - T_1) u_t \right]' \]
\[
= \left[ T^{-3/2 - d} \sum_{t=1}^T u_t T^{-5/2 - d} \sum_{t=1}^T t u_t T^{-5/2 - d} \sum_{t=T+1}^T (t - T_1) u_t \right]' + o_p(1).
\]


Hence, for $d \in (-0.5, 0.5)$ and $m = 1$, we have

$$T^{-2-2d}(B_{T_1^0} - B_{T_1})'(I - P_{T_1})U$$

$$= \kappa(d) \left[\int_{\lambda_0}^{1} B_d(r)dr + \frac{1 - \lambda_0}{2} \int_{0}^{1} B_d(r)dr - \frac{3(1 - \lambda_0)}{2\lambda_0} \int_{0}^{1} rB_d(r)dr - \frac{3(2\lambda_0 - 1)}{2\lambda_0(1 - \lambda_0)} \int_{\lambda_0}^{1} (r - \lambda_0)B_d(r)dr\right]x^0_mT + o_p(1)$$

$$= \kappa(d)\beta^0_b m_T \int_{\lambda_0}^{1} B^*_d(r)dr + o_p(1),$$

where $B^*_d(r)$ is the residuals function from a continuous time least-squares regression of $B_d(r)$ on \{1, r, (r - \lambda_0)1_{r > \lambda_0}\}. Therefore,

$$m^*_T = T^{-1/2-2d}|T_1 - T_1^0| = \arg\min_{m_T \in D(C)} [(XX)/T^{2+2d} + 2(XU)/T^{2+2d} + o_p(1)]$$

$$= \arg\min_{m_T \in D(C)} \left[m^2_T(\beta^0_b)^2 \frac{(1 - \lambda_0)^2}{4} + 2\kappa(d)m_T\beta^0_b \int_{\lambda_0}^{1} B^*_d(r)dr\right] + o_p(1)$$

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

$$m^*_T = T^{-1/2-2d}|T_1 - T_1^0| \Rightarrow -\frac{4\kappa(d) \int_{\lambda_0}^{1} B^*_d(r)dr}{\lambda_0(1 - \lambda_0)^2 \beta^0_b}.$$

Next, consider the case with $m = 0$. Define $m_T = T^{-1/2+2d}|T_1 - T_1^0|$ for this case. Note that $T^{-1/2-2d}U \Rightarrow \kappa(d) \int_{\lambda_0}^{1} dB_d(r)$. For $(XX)$, we have the same results as for $m = 1$. For $(XU)$, we have:

$$(XU) = \beta^0_b (B_{T_1^0} - B_{T_1})'(I - P_{T_1})U = T^{-1/2-2d} \beta^0_b m_T\gamma^0_U(I - P_{T_1})U$$

$$= T^{-1/2-2d} \beta^0_b m_T\gamma^0_U U - T^{-1/2-2d} \beta^0_b m_T\gamma^0_U X_{T_1} D_{T}^{-1}(D_{T}^{-1}X_{T_1}X_{T_1}D_{T}^{-1})^{-1}D_{T}^{-1}X_{T_1} U$$

$$= \beta^0_b m_T \kappa(d) \left[\int_{\lambda_0}^{1} dB_d(r) - \left[\frac{\lambda_0 - 1}{2} \frac{3(1 - \lambda_0)}{2\lambda_0} \frac{3(2\lambda_0 - 1)}{2\lambda_0(1 - \lambda_0)} \right] \left[\int_{\lambda_0}^{1} dB_d(r) + \int_{\lambda_0}^{1} rB_d(r)dr + \int_{\lambda_0}^{1} (r - \lambda_0)B_d(r)dr\right]\right] + o_p(1)$$

$$= \beta^0_b m_T \kappa(d) \left[\int_{\lambda_0}^{1} dB_d(r) - \left[\frac{\lambda_0 - 1}{2} \frac{3(1 - \lambda_0)}{2\lambda_0} \frac{3(2\lambda_0 - 1)}{2\lambda_0(1 - \lambda_0)} \right] \left[\int_{\lambda_0}^{1} dB_d(r) + \int_{\lambda_0}^{1} rB_d(r)dr + \int_{\lambda_0}^{1} (r - \lambda_0)B_d(r)dr\right]\right] + o_p(1)$$

$$\equiv \beta^0_b m_T \kappa(d) \zeta + o_p(1).$$

For $(UU)$, we know that $U$ is an $I(d)$ process with $d \in (-0.5, 0.5)$. It is easy to show that $U'((X_{T_1} - X_{T_1}D_{T}^{-1}|T_1 - T_1^0|O_p(T^{-1}), D_{T}^{-1}X_{T_1}U = O_p(1)$, and $D_{T}^{-1}X_{T_1}^2X_{T_1}D_{T}^{-1} = O_p(T^{-2d})$. Hence, $(UU) = |T_1 - T_1^0|O_p(T^{-1-2d})$ which is dominated by $(XU)$ asymptotically. The optimal $m^*_T$ is therefore given by

$$m^*_T = T^{3/2+2d}(\lambda - \lambda_0) \Rightarrow -\frac{4\kappa(d) \int_{\lambda_0}^{1} B^*_d(r)dr}{\beta^0_b \lambda(1 - \lambda_0)^2 \beta^0_b \lambda(1 - \lambda_0)^2}.$$
A.1.4 Limit Distributions of the Estimates of the Other Parameters

The OLS estimates of the regression coefficients $\gamma$ is

$$
\hat{\gamma} = (X_{T_1}^\prime X_{T_1})^{-1} X_{T_1}^\prime Y = (X_{T_1}^\prime X_{T_1})^{-1} X_{T_1}^\prime X_{T_1} \gamma^0 + (X_{T_1}^\prime X_{T_1})^{-1} X_{T_1}^\prime U
$$

Hence,

$$
D_T(\hat{\gamma} - \gamma^0) = (D_T^{-1} X_{T_1}^\prime X_{T_1} D_T^{-1})^{-1}[D_T^{-1} X_{T_1}^\prime (X_{T_1}^0 - X_{T_1}) \gamma^0 + D_T^{-1} X_{T_1}^\prime U].
$$

First, for $m = 0$,

$$
D_T^{-1} X_{T_1}^\prime (X_{T_1}^0 - X_{T_1}) \gamma^0 + D_T^{-1} X_{T_1}^\prime U = D_T^{-1} X_{T_1}^\prime \beta_b | T_1 - T_1 \gamma_b + D_T^{-1} X_{T_1}^\prime U
$$

$$
= D_T^{-1} X_{T_1}^\prime \beta_b | T_1 - T_1 [T_1^{1/2-d} \sigma_b T_1^{-1/2+d}] + D_T^{-1} X_{T_1}^\prime U = m_T \beta_b T_1^{-1/2+d} D_T^{-1} X_{T_1}^\prime \tilde{\gamma}_b + D_T^{-1} X_{T_1}^\prime U
$$

Thus,

$$
\Rightarrow \frac{-4\kappa(d) \zeta}{\lambda_0(1 - \lambda_0)} \left[ 1 - \frac{1 - (\lambda_0)^2}{2} \right] \kappa(d) \left[ \int_0^1 dB_d(r) \right]
$$

$$
= \kappa(d) \left( \int_0^\lambda \left[ \frac{3(\lambda_0)^2 - 2\lambda_0 - 6\lambda r + 6r}{(\lambda_0)^2} \right] dB_d(r) + \int_1^\lambda \left[ \frac{-3(\lambda_0)^2 - 3\lambda_0 - 2r}{1 - \lambda_0} \right] dB_d(r) \right) \equiv \Sigma_0.
$$

Since $T^{-2d}(D_T^{-1} X_{T_1}^\prime X_{T_1} D_T^{-1})^{-1} \rightarrow \Sigma_0^{-1}$, $T^{-2d} D_T(\hat{\gamma} - \gamma^0) \rightarrow \Sigma_0^{-1} \Sigma_0$.

Second, for $m = 1$,

$$
T^{-1} D_T(\hat{\gamma} - \gamma^0) = (D_T^{-1} X_{T_1}^\prime X_{T_1} D_T^{-1})^{-1}[T_1^{-1/2-d} \sigma_b T_1^{-1/2+d} + T_1^{-1} D_T^{-1} X_{T_1}^\prime U].
$$

Then, we have

$$
T^{-1} D_T^{-1} X_{T_1}^\prime (X_{T_1}^0 - X_{T_1}) \gamma^0 + T^{-1} D_T^{-1} X_{T_1}^\prime U
$$

$$
= D_T^{-1} X_{T_1}^\prime \beta_b | T_1 - T_1 [T_1^{-1/2-d} \sigma_b T_1^{-1/2+d}] + T^{-1} D_T^{-1} X_{T_1}^\prime U
$$

$$
= m_T \beta_b T_1^{-1/2+d} D_T^{-1} X_{T_1}^\prime \tilde{\gamma}_b + T^{-1} D_T^{-1} X_{T_1}^\prime U
$$

Thus,

$$
\Rightarrow -\kappa(d) \left( \int_0^\lambda B_d^*(r) dr \right) \left[ \frac{2(1 + \lambda_0)}{2(1 - \lambda_0)} \right] + \kappa(d) \left[ \int_0^1 dB_d(r) dr \right]
$$

$$
= \kappa(d) \left( \int_0^\lambda \left[ \frac{3(1 - \lambda_0)^2 + 3(\lambda_0 - 2\lambda_0)^2}{(\lambda_0)^2} \right] dB_d(r) + \int_1^\lambda \left[ \frac{3(\lambda_0 + 2\lambda_0 - 1) + 3}{2(1 - \lambda_0)} \right] dB_d(r) \right) \equiv \Sigma_1.
$$

11
Therefore,

\[ T^{-1-2d}D_T(\hat{\gamma} - \gamma^0) \Rightarrow \Sigma^{-1}\Sigma_1. \]

### A.2 Results for Model II

We now consider results for Model II. The proof of the consistency is similar to that for Model I. In any event, the relevant bound will be derived in the proof of the limit distribution.

#### A.2.1 Consistency (Theorem 1)

From Lemma A.1, for \( m = 0 \):

\[
\begin{align*}
(\hat{X}\hat{X}) &= (T_1^0 - \hat{T}_1)^3O(1) \\
(\hat{X}\hat{U}) &= (T_1^0 - \hat{T}_1)^{3/2+d}O_p(1) \\
(\hat{U}\hat{U}) &= |T_1^0 - \hat{T}_1|^{1/2+d}O_p(T^{-1/2+d}),
\end{align*}
\]

and for \( m = 1 \):

\[
\begin{align*}
(\hat{X}\hat{X}) &= (T_1^0 - \hat{T}_1)^3O(1) \\
(\hat{X}\hat{U}) &= (T_1^0 - \hat{T}_1)^2O_p(T^{1/2+d}) \\
(\hat{U}\hat{U}) &= |T_1^0 - \hat{T}_1|O_p(T^{1+2d}).
\end{align*}
\]

The proof of consistency is similar to that for Model I. Suppose that \( \hat{\lambda} \not\to \lambda \). Then, with \( m = 1 \), \( (\hat{X}\hat{X}) = O(T^3) \), \( (\hat{X}\hat{U}) = O_p(T^{5/2+d}) \) and \( (\hat{U}\hat{U}) = O_p(T^{2+2d}) \) for all \( d \in (-0.5, 0.5) \). With some positive probability, \( (\hat{X}\hat{X}) \) dominates the other two terms, so that this result cannot be compatible with the key inequality \( (\hat{X}\hat{X}) + 2(\hat{X}\hat{U}) + (\hat{U}\hat{U}) \leq 0 \). Hence, we have a contradiction and conclude that \( \hat{\lambda} \not\to \lambda \).

#### A.2.2 Rate of Convergence (Theorem 2)

We consider the set

\[
\tilde{V}_e(\epsilon) = \{T_1 : |T_1 - T_1^0| < \epsilon T \text{ and } |T_1 - T_1^0| > CT^{m(d+1/2)}, \forall \epsilon > 0, \forall d \in (-0.5, 0.5)\}
\]

where \( m \in \{0, 1\} \). Given the results in Lemma A.1, if \( m = 0 \):

\[
\begin{align*}
\frac{(XX)}{|T_1 - T_1^0|^{3/2+d}} &= \frac{|T_1 - T_1^0|^3O(1)}{|T_1 - T_1^0|^{3/2+d}} = \frac{|T_1 - T_1^0|^{3/2-d}O(1)}{aC + o(1)} > C^{3/2-d}O(1), \\
\frac{(XU)}{|T_1 - T_1^0|^{3/2+d}} &= \frac{|T_1 - T_1^0|^{3/2+d}O_p(1)}{|T_1 - T_1^0|^{3/2+d}} = O_p(1), \\
\frac{(UU)}{|T_1 - T_1^0|^{3/2+d}} &= \frac{|T_1 - T_1^0|^{1/2+d}O_p(T^{-1/2+d})}{|T_1 - T_1^0|^{3/2+d}} = o_p(1),
\end{align*}
\]
and if \( m = 1 \):

\[
\frac{(XX)}{|T_1 - T_0|^2T^{1/2+d}} = \frac{|T_1 - T_0|^3O(1)}{|T_1 - T_0|^2T^{1/2+d}} > \frac{CT^{1/2+d}O(1)}{T^{1/2+d}} = aC + o(1),
\]

\[
\frac{(XU)}{|T_1 - T_0|^2T^{1/2+d}} = \frac{|T_1 - T_0|^2O_p(T^{1/2+d})}{|T_1 - T_0|^2T^{1/2+d}} = O_p(1),
\]

\[
\frac{(UU)}{|T_1 - T_0|^2T^{1/2+d}} = \frac{|T_1 - T_0|^2O_p(T^{1+2d})}{|T_1 - T_0|^2T^{1/2+d}} = o_p(1)
\]

for every \( d \in (-0.5, 0.5) \) where \( a \) is a positive constant. We can claim that \( \hat{T}_1 \notin \tilde{V}_c(\epsilon) \) by choosing a sufficiently large \( C > 0 \) such that for any \( \eta > 0 \),

\[
\Pr \left( \min_{T_1 \in \tilde{V}_c(\epsilon)} \{S(T_1) - S(T_0)\} \leq 0 \right) < \eta.
\]

This completes the proof.

### A.2.3 Limit Distribution of the Estimate of the Break Date

Given the results in Theorem 2, we work on the set \( D_0(C) = \{T_1 : |T_1 - T_0| < C\} \) if \( m = 0 \) and \( D_1(C) = \{T_1 : |T_1 - T_0| < T^{1/2+d}C\} \) if \( m = 1 \), for some positive \( C \). In other words, for \( \lambda = T_1/T \), \( |\lambda - \lambda^0| = O_p(T^{-1}) \) with \( m = 0 \) and \( |\lambda - \lambda^0| = O_p(T^{-1/2+d}) \) with \( m = 1 \). In Model II,

\[
X_{T_1} = \begin{bmatrix} \iota & t & C_{T_1} & B_{T_1} \end{bmatrix}
\]

where \( \iota = (1, \ldots, 1)' \), \( t = (1, 2, \ldots, T)' \), \( C_{T_1} = (C_1, \ldots, C_T)' \), \( B_{T_1} = (B_1, \ldots, B_T)' \) and \( \gamma = (\mu_1, \beta_1, \mu_b, \beta_b)' \). For \( T_1 > T_0 \),

\[
C_{T_1}^0 - C_{T_1} = \begin{cases} 1 & \text{if } T_1^0 \leq t \leq T_1 \\ 0 & \text{otherwise}, \end{cases}
\]

and

\[
B_{T_1^0} - B_{T_1} - (T_1 - T_1^0)C_{T_1} = \begin{cases} t - T_1^0 & \text{if } T_1^0 \leq t \leq T_1 \\ 0 & \text{otherwise}. \end{cases}
\]

When \( T_1 < T_1^0 \),

\[
C_{T_1^0} - C_{T_1} = \begin{cases} -1 & \text{if } T_1 \leq t \leq T_1^0 \\ 0 & \text{otherwise}, \end{cases}
\]

and

\[
B_{T_1^0} - B_{T_1} - (T_1 - T_1^0)C_{T_1} = \begin{cases} -(t - T_1^0) & \text{if } T_1 \leq t \leq T_1^0 \\ 0 & \text{otherwise}. \end{cases}
\]
We shall use the following notations. For $T_1^0 > T_1$,

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

and for $T_1^0 < T_1$,

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

We first consider the term $(XX)$. Noting that $(T_1 - T_1^0)(I - P_{T_1})C_{T_1} = 0$, we have

$$(XX) = \gamma^0(X_{T_1^0} - X_{T_1})(I - P_{T_1})(X_{T_1^0} - X_{T_1})\gamma^0$$

$$= [(C_{T_1^0} - C_{T_1})\mu_b^0 + (B_{T_1^0} - B_{T_1} - (T_1 - T_1^0)C_{T_1})\beta_b^0]'(I - P_{T_1})$$

$$\times [(C_{T_1^0} - C_{T_1})\mu_b^0 + (B_{T_1^0} - B_{T_1} - (T_1 - T_1^0)C_{T_1})\beta_b^0]$$

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

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

where $D_T = \text{diag}(T^{1/2+d}, T^{3/2+d}, T^{1/2+d}, T^{3/2+d})$. Note that for $T_1 > T_1^0$,

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

$$= T^{-d}\sum_{t=T_1^0+1}^{T_1} [\mu_b^0 + \beta_b^0(t-T_1^0)][T^{-1/2} tT^{-3/2} 0 0]$$

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

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

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

$$\leq T^{-1/2-d}g_2[1 T_1^0/T 0 0] + g_2[T^{-1/2-d}|T_1 - T_1^0|/T][0 1 0 0]$$

$$= O_p(|g_2|T^{-1/2-d})$$
where the last step follows from the fact that $|T_1 - T_1^0|/T \xrightarrow{p} 0$ and

$$(D_T^{-1}X_{T_1}X_{T_1}D_T^{-1})^{-1} = O_p(T^{2d})$$

on both $D_0(C)$ and $D_1(C)$. Hence, the second term in $(XX)$ is such that

$$\gamma^0(X_{T_1^0} - X_{T_1})'P_{T_1}(X_{T_1^0} - X_{T_1})\gamma^0 = O_p(g_2^2T^{-1}) = o_p(h_2)$$

because $|\lambda - \lambda^0| = O_p(T^{-1})$ if $m = 0$ and $|\lambda - \lambda^0| = O_p(T^{-1/2+d})$ if $m = 1$ where $d \in (-0.5, 0.5)$. Therefore,

$$(XX) = \begin{cases} 
      h_2 + o_p(h_2) & \text{if } T_1 > T_1^0 \\
      h_1 + o_p(h_1) & \text{if } T_1 \leq T_1^0.
   \end{cases}$$

This implies that

$$(XX) = |T_1 - T_1^0|^2O(1)$$

since $\mu_b^0$ is fixed. Consider now the term $(XU)$. For $m = 1$, we have

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

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

$$- \left\{ \sum_{t=T_1^0+1}^{T_1} \left[ \mu_b^0 + \beta_b^0(t - T_1^0) \right] x(T_1)'(D_T^{-1}) (D_T^{-1}X_{T_1}X_{T_1}D_T^{-1})^{-1} D_T^{-1}X_{T_1}' U. \right\}$$

We consider each term of this expression.

1. Since $m = 1$, $u_t = \Delta^{-1} \xi^#_t = u_{T_1^0} + \sum_{i=T_1^0+1}^{t} \zeta_i = u_{T_1^0} + v_t$ for $t > T_1^0$. When $T_1^0 < T_1$,

$$T^{-1/2-d} \sum_{t=T_1^0+1}^{T_1} \left[ \mu_b^0 + \beta_b^0(t - T_1^0) \right] u_t = T^{-1/2-d} \sum_{k=1}^{T_1-T_1^0} \left[ \mu_b^0 + \beta_b^0k \right] (u_{T_1^0} + v_{T_1^0+k})$$

$$= g_2\kappa(d)B_d(\lambda^0) + o_p(g_2).$$

2. $T^{2d}D_T^{-1}X_{T_1}X_{T_1}D_T^{-1} = \Omega_1^{-1} + o(1)$, where

$$\Omega_1 = \begin{bmatrix} 
\frac{4}{\lambda^0} & -\frac{6}{(\lambda^0)^2} & \frac{2}{\lambda^0} \\
-\frac{6}{(\lambda^0)^2} & \frac{12}{(\lambda^0)^3} & -\frac{6}{(\lambda^0)^2} \\
\frac{2}{\lambda^0} & -\frac{6}{(\lambda^0)^2} & \frac{4}{\lambda^0(1-\lambda^0)} \\
\frac{6}{(\lambda^0)^2} & -\frac{12}{(\lambda^0)^3} & \frac{6(1-2\lambda^0)}{(\lambda^0)^2(1-\lambda^0)^2} \\
\frac{6}{(\lambda^0)^2} & \frac{6(1-2\lambda^0)}{(\lambda^0)^2(1-\lambda^0)^2} & \frac{12(3\lambda^0)^2-3\lambda^0+1}{(\lambda^0)^3} \end{bmatrix}, \quad (A.4)$$
We can show that when
\[T = 3,\]

Combining the results 1-4, we obtain that
\[
\frac{\beta_0}{\mu_b} + \frac{\beta_0}{\mu_b} = \frac{1}{T}
\]

using integration by parts.

4. When \( T_1^0 < T_1 \),
\[
\sum_{t=T_1^0+1}^{T_1} \left[ \mu_b^0 + \beta_0(t - T_1^0) \right] x(T_1) D_{\bar{t}}^{-1} \cdot
\]
\[
= T^{-1/2 - d} \sum_{t=T_1^0+1}^{T_1} \left[ \mu_b^0 + \beta_0(t - T_1^0) \right] \begin{bmatrix} 1 & t/T & 0 \end{bmatrix} \cdot
\]
\[
= T^{-1/2 - d} g_2 \begin{bmatrix} 1 & \lambda^0 & 0 \end{bmatrix} \cdot
\]

Combining the results 1-4, we obtain that
\[
(XU) = T^{1/2 + d} g_2 \kappa(d) \{ B_d(\lambda^0) - [1 \ \lambda^0 \ 0 \ 0] \Omega_1 \xi_1 + o_p(1) \}
\]
\[
= T^{1/2 + d} g_2 \kappa(d) \xi_3 + o_p(T^{1/2 + d} g_2).
\]

After some algebra, we have
\[
\xi_3 = B_d(\lambda^0) - [1 \ \lambda^0 \ 0 \ 0] \Omega_1 \xi_1 = \int_0^{\lambda^0} \left( \frac{3r^2 - 2\lambda^0 r}{(\lambda^0)^2} \right) dB_d(r).
\]

We can show that when \( T_1^0 > T_1 \),
\[
T^{-1/2 - d} \sum_{t=T_1^0+1}^{T_1} \left[ \mu_b^0 + \beta_0(t - T_1^0) \right] u_t = g_1 \kappa(d) B_d(\lambda^0) + o_p(g_1)
\]

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

3. \( T^{-1} D_{\bar{t}}^{-1} X_{\bar{t}} U \Rightarrow \kappa(d) \xi_1 \), where
\[
\xi_1 = \begin{bmatrix}
\int_0^1 B_d(r) dr \\
\int_0^1 rB_d(r) dr \\
\int_0^1 B_d(r) dr \\
\int_0^1 (r - \lambda^0) B_d(r) dr
\end{bmatrix} = \begin{bmatrix}
\int_0^1 (1 - r) dB_d(r) \\
\int_0^1 \frac{1 - r^2}{2} dB_d(r) \\
\int_0^1 (1 - \lambda^0) dB_d(r) + \int_0^1 (1 - r) dB_d(r) \\
\int_0^1 \frac{(1 - \lambda^0)^2}{2} dB_d(r) + \int_0^1 \frac{(1 - \lambda^0)^2 - (r - \lambda^0)^2}{2} dB_d(r)
\end{bmatrix}
\]

(A.5)
and
\[
\sum_{t=T_1^0}^{T_1} [\mu_0^t + \beta_0^t(t - T_1^0)]x(T_1)^tD_T^{-1} = T^{-1/2 - d}g_1[1 \quad \lambda^0 \quad 1 \quad 0] + o_p(g_1T^{-1/2 - d}).
\]
Hence,
\[
(XU) = T^{1/2 + d}g_1\kappa(d)\xi_4 + o_p(T^{1/2 + d}g_1)
\]
where
\[
\xi_4 = \int_{\lambda_0}^{1}[(r - 1)(3r - 2\lambda_0 - 1)/(1 - \lambda_0^2)]dB_d(r).
\]
These results imply that
\[
(XU) = \begin{cases} 
T^{1/2 + d}\kappa(d)g_2\xi_3 + o_p(1) & \text{if } T_1^0 < T_1 \\
T^{1/2 + d}\kappa(d)g_1\xi_4 + o_p(1) & \text{if } T_1^0 > T_1
\end{cases}
\]
and
\[
(XU) = |T_1 - T_1^0|^2O_p(T^{1/2 + d}).
\]
We finally consider the term \((UU)\). We have
\[
(UU) = U'\left(P_{T_1^0} - P_{T_1}\right)U
\]
\[
= U'(X_{T_1^0} - X_{T_1})D_T^{-1}(D_T^{-1}X_{T_1^0}'X_{T_1^0}D_T^{-1})^{-1}D_T^{-1}X_{T_1'}U
\]
\[
+ U'X_{T_1}D_T^{-1}(D_T^{-1}X_{T_1^0}'X_{T_1^0}D_T^{-1})^{-1}D_T^{-1}[X_{T_1}'X_{T_1} - X_{T_1^0}'X_{T_1^0}]D_T^{-1}(D_T^{-1}X_{T_1^0}'X_{T_1^0}D_T^{-1})^{-1}D_T^{-1}X_{T_1'}U
\]
\[
+ U'X_{T_1}D_T^{-1}(D_T^{-1}X_{T_1'}X_{T_1}D_T^{-1})^{-1}D_T^{-1}(X_{T_1^0} - X_{T_1})'U.
\]
We first have
\[
T^{-1/2 - d}U'(C_{T_1^0} - C_{T_1}) = \kappa(d)(T_1^0 - T_1)\int_{\lambda_0}^{1} B_d(r)dr + o_p(1),
\]
\[
T^{-3/2 - d}U'(B_{T_1^0} - B_{T_1}) = \kappa(d)(T_1^0 - T_1)\int_{\lambda_0}^{1} rB_d(r)dr + o_p(1).
\]
Hence,
\[
U'(X_{T_1^0} - X_{T_1})D_T^{-1} = (T_1 - T_1^0)[\kappa(d)\xi_2' + o_p(1)]
\]
where \(\xi_2' = [0 \quad 0 \quad \int_{\lambda_0}^{1} B_d(r)dr \quad \int_{\lambda_0}^{1} rB_d(r)dr]\). For the second term in \((UU)\), we have
\[
D_T^{-1}[X_{T_1'}X_{T_1} - X_{T_1^0}'X_{T_1^0}]D_T^{-1} = -(T_1 - T_1^0)T^{-1 - 2d}\Sigma_f + o_p(1)
\]
with
\[
\Sigma_f = \begin{bmatrix}
0 & 0 & 1 & 1 - \lambda^0 \\
0 & 0 & \lambda^0 & \frac{1 - (\lambda^0)^2}{2} \\
1 & \lambda^0 & 1 & 1 - \lambda^0 \\
1 - \lambda^0 & \frac{1 - (\lambda^0)^2}{2} & 1 - \lambda^0 & (1 - \lambda^0)^2
\end{bmatrix}.
\]
Hence,
\[ T^{1+2d} U' X_TD_T^{-1}(D_T^{-1} X_{T_1}' X_{T_1}D_T^{-1})^{-1} D_T^{-1}[X_T' X_T - X_{T_1}' X_{T_1}]D_T^{-1}(D_T^{-1} X_{T_1}' X_{T_1}D_T^{-1})^{-1} D_T^{-1} X_{T_1}' U \]
\[ = -(T_1 - T_1^0)T^{1+2d} \kappa(d)^2 [\xi_1' \Omega_2 \xi_1 + o_p(1)] \]

where
\[
\Omega_2 = \Omega_1 \Sigma_f \Omega_1 = \begin{bmatrix}
\frac{-4}{(\lambda')^2} & \frac{12}{(\lambda')^3} & \frac{-2}{(\lambda')^2} & \frac{-12}{(\lambda')^3} \\
\frac{12}{(\lambda')^3} & \frac{-36}{(\lambda')^4} & \frac{12}{(\lambda')^3} & \frac{36}{(\lambda')^4} \\
\frac{-2}{(\lambda')^2} & \frac{12}{(\lambda')^3} & \frac{4(\lambda'0-1)}{(\lambda')^4(1-\lambda')^2} & \frac{12(3(\lambda')^2-3\lambda'0+1)}{(\lambda')^4(1-\lambda')^3} \\
\frac{-12}{(\lambda')^3} & \frac{36}{(\lambda')^4} & \frac{12(3(\lambda')^2-3\lambda'0+1)}{(\lambda')^4(1-\lambda')^3} & \frac{36(4\lambda')^2-6(\lambda')^2+4\lambda'0-1)}{(\lambda')^4(1-\lambda')^4}
\end{bmatrix}
\]

Collecting the results above, we have
\[(UU) = (T_1 - T_1^0)T^{1+2d} \kappa(d)^2 [2\xi_2' \Omega_1 \xi_1 - \xi_1' \Omega_2 \xi_1 + o_p(1)].\]

This implies that with \( m = 1, \)
\[(UU) = |T_1 - T_1^0|O_p(T^{1+2d}).\]

Define \( m_T = (T_1 - T_1^0)T^{-1/2-d} \). It is easy to show that both \( h_1 \) and \( h_2 \) are asymptotically equivalent to \( T^{3/2+2d(\beta_0^0)^2} |m_T|^3/3 \) and both \( g_1 \) and \( g_2 \) are asymptotically equivalent to \( T^{1+2d} m_T^{-2} \beta_0^0/2 \), therefore
\[ T^{-3/2-3d}(XX) = (\beta_0^0)^2 |m_T|^3/3 + o_p(1), \]
\[ 2T^{-3/2-3d}(XU) = \begin{cases} \kappa(d)m_T^2 \beta_0^0 \xi_3 + o_p(1) & \text{if } m_T > 0 \\ \kappa(d)m_T^2 \beta_0^0 \xi_4 + o_p(1) & \text{if } m_T < 0 \end{cases} \]
\[ T^{-3/2-3d}(UU) = m_T \kappa(d)^2 [2\xi_2^2 \Omega_1 \xi_1 - \xi_1^2 \Omega_2 \xi_1] + o_p(1). \]

Define \( Z^*(v; \lambda, \beta_0^0, \kappa(d)) \) as follows: \( Z^*(0) = 0, Z^*(v) = Z_1(v) \) for \( v < 0 \) and \( Z^*(v) = Z_2(v) \) for \( v > 0 \), with
\[ Z_1(v; \lambda, \beta_0^0, \kappa(d)) = (\beta_0^0)^2 |v|^3/3 + v^2 \kappa(d) \beta_0^0 \xi_4 + v \kappa(d)^2 [2\xi_2^2 \Omega_1 \xi_1 - \xi_1^2 \Omega_2 \xi_1] + o_p(1), \]
\[ Z_2(v; \lambda, \beta_0^0, \kappa(d)) = (\beta_0^0)^2 |v|^3/3 + v^2 \kappa(d) \beta_0^0 \xi_3 + v \kappa(d)^2 [2\xi_2^2 \Omega_1 \xi_1 - \xi_1^2 \Omega_2 \xi_1] + o_p(1). \]

By the continuous mapping theorem, we have
\[ m_T^* = \arg \min_v Z^*(v; \lambda, \beta_0^0, \kappa(d)). \]

Now, consider the case with \( d \in (-0.5, 0.5) \) and \( m = 0 \). The following argument applies to the set
\[ D_0(C) = \{ T_1 : |T_1 - T_1^0| < C \} \]
in which we have \( |\lambda - \lambda_0| = O_p(T^{-1}) \) for \( \lambda = T_1/T \). As in the case with \( m = 1, \)
\[ (XX) = \begin{cases} h_1 + o_p(h_1) & \text{if } T_1 < T_1^0 \\ h_2 + o_p(h_2) & \text{if } T_1 > T_1^0 \end{cases} \]
If $T_1 > T_1^0$,

$$(XU) = \sum_{t=T_1^0+1}^{T_1} [\mu^0_b + \beta^0_b(t - T_1^0)]u_t$$

$$- \{ \sum_{t=T_1^0+1}^{T_1} [\mu^0_b + \beta^0_b(t - T_1^0)]x(T_1)^{D_{T^{-1}}}(D_{T^{-1}}X_{T_1}D_{T^{-1}})^{-1} D_{T^{-1}}X_{T_1}U \}.$$

We next consider each term of $(XU)$.

1. \[ \sum_{t=T_1^0+1}^{T_1} [\mu^0_b + \beta^0_b(t - T_1^0)]u_t \]

$$= \sum_{k=1}^{T_1-T_1^0} [\mu^0_b + \beta^0_b]u_{T_1^0+k}^{T_1-1} = O_p(|T_1 - T_1^0|^{3/2+d}).$$

2. \[ \sum_{t=T_1^0+1}^{T_1} \sum_{k=1}^{T_1-T_1^0} [\mu^0_b + \beta^0_b(t - T_1^0)]x(T_1)^D_{T^{-1}} \]

$$= T^{-1/2-d} \sum_{t=T_1^0+1}^{T_1} [\mu^0_b + \beta^0_b(t - T_1^0)][1 \ t/T \ 0 \ 0]$$

$$= T^{-1/2-d} \sum_{k=1}^{T_1-T_1^0} [\mu^0_b + \beta^0_b]k[1 \ (k + T_1^0)/T \ 0 \ 0]$$

$$= T^{-1/2-d} \sum_{k=1}^{T_1-T_1^0} [\mu^0_b + \beta^0_b]k[1 \ k/T \ 0 \ 0] + T^{-1/2-d} \sum_{k=1}^{T_1-T_1^0} [\mu^0_b + \beta^0_b]k[1 \ \lambda^0 \ 0 \ 0]$$

$$= O_p((T_1 - T_1^0)^2 T^{-1/2-d}).$$

3. \(D_{T^{-1}}X_{T_1}X_{T_1}D_{T^{-1}})^{-1} \approx O_p(T^{2d}).$$

4. \(D_{T^{-1}}X_{T_1}U = O_p(1).$$

Since we search in a set where \(|T_1 - T_1^0| < C\) for some \(C > 0\) and \(|\lambda - \lambda^0| = O_p(T^{-1})\), \(\gamma^0(X_{T_1^0} - X_{T_1})^' P_{T_1}U\) is dominated by \(\gamma^0(X_{T_1^0} - X_{T_1})^' U\) asymptotically. Hence,

$$(XU) = |T_1 - T_1^0|^{3/2+d} O_p(1).$$
We can derive the results for $T_1^0 > T_1$ in a similar way and we obtain,

$$(XU) = \begin{cases} 
\sum_{t=T_1^0+1}^{T_1} [p_0^0 + \beta_0(t - T_1^0)]u_t + o_p(1) & \text{if } T_1 > T_1^0 \\
0 & \text{if } T_1 = T_1^0 \\
-\sum_{t=T_1+1}^{T_1^0} [p_0^0 + \beta_0(t - T_1^0)]u_t + o_p(1) & \text{if } T_1 < T_1^0.
\end{cases}$$

Next, consider the term $(UU)$. We have

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

$$= T^{-1/2-d}|T_1 - T_1^0|^{1/2+d}|T_1 - T_1^0|^{-1/2-d} \max_{\min \{T_1, T_1^0\} + 1} \sum_{t=\min \{T_1, T_1^0\} + 1}^{\max \{T_1, T_1^0\}} u_t$$

$$= T^{-1/2-d}|T_1 - T_1^0|^{1/2+d}O_p(1),$$

and

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

Hence,

$$U'(X_{T_1^0} - X_{T_1})D_T^{-1} = |T_1 - T_1^0|^{1/2+d}O_p(T^{-1/2-d}).$$

Then following the same arguments as for Model I, we have

$$(UU) = |T_1 - T_1^0|^{1/2+d}O_p(T^{-1/2-d})O_p(T^{2d})O_p(1) = |T_1 - T_1^0|^{1/2+d}O_p(T^{-1/2+d}).$$

Following Bai (1997), we define a stochastic process $S^*(\nu)$ on the set of integers as follows:

$$S^*(\nu) = \begin{cases} 
S_1(\nu) & \text{if } \nu < 0 \\
0 & \text{if } \nu = 0 \\
S_2(\nu) & \text{if } \nu > 0
\end{cases}$$

with

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

$$S_2(\nu) = \sum_{k=1}^{\nu} (\mu_b + \beta_b k)^2 + 2 \sum_{k=1}^{\nu} (\mu_b + \beta_b k)u_k, \quad \nu = 1, 2, \ldots.$$
A.2.4 Limit Distributions of the Estimates of the Other Parameters

As for Model I, we use the facts that

\[ D_T(\hat{\gamma} - \gamma^0) = (D_T^{-1}X_{T_1}^rX_{T_1}D_T^{-1})^{-1}[D_T^{-1}X_{T_1}^r(X_{T_1} - X_{T_1})\gamma^0 + D_T^{-1}X_{T_1}^rU], \]

and

\[ T^{-2d}(D_T^{-1}X_{T_1}^rX_{T_1}D_T^{-1})^{-1} = \Omega_1 + o(1). \]

Hence, when \( m = 0 \), we obtain

\[ T^{-2d}D_T(\hat{\gamma} - \gamma^0) = \Omega_1 \begin{pmatrix} \beta_0^{|\hat{T}_1 - T_1^0|T^{1/2-d}} & 1 - \lambda^0 & \frac{1-(\lambda^0)^2}{2} & f_0^1 dB_d(r) & f_0^1 r dB_d(r) & f_0^1 dB_d(r) \\ 0 & \frac{1}{1 - \lambda^0} & \frac{(1 - \lambda^0)^2}{2} & f_0^1 dB_d(r) & f_0^1 r dB_d(r) & f_0^1 dB_d(r) \\ 0 & 1 & \frac{1}{1 - \lambda^0} & f_0^1 dB_d(r) & f_0^1 r dB_d(r) & f_0^1 dB_d(r) \\ 0 & 0 & 0 & f_0^1 dB_d(r) & f_0^1 r dB_d(r) & f_0^1 dB_d(r) \end{pmatrix} + o_p(1) \]

Note that the limiting distribution of \( \hat{\mu}_b \) depends on that of \(|\hat{T}_1 - T_1^0|\). Similarly, it is easy to show that, when \( m = 1 \),

\[ T^{-1-2d}D_T(\hat{\gamma} - \gamma^0) = \beta_0^{|\hat{T}_1 - T_1^0|T^{-1/2-d}} + \kappa(d)\Omega_1 \begin{pmatrix} f_0^1 B_d(r) \\ f_0^1 r B_d(r) \\ f_0^1 B_d(r) \\ f_0^1 dB_d(r) \end{pmatrix} + o_p(1). \]

Therefore, the limiting distribution of \( \hat{\mu}_b \) depends on that of \(|\hat{T}_1 - T_1^0|\) while the limiting distributions of the estimates of the other parameters do not. We can derive the result in Theorem 4 using (A.4) and (A.5).
A.3 Results for a Spurious Break

Proof of Theorem 5. Consider first the case with \( d^* \in (-0.5, 0] \). After some algebra, we have

\[
M^*(\lambda) = 12\kappa(d)^2 \left\{ \frac{B_d(\lambda)^2}{3\lambda} \left( \int_0^\lambda r dB_d(r) \right) B_d(\lambda) + \left( \int_0^\lambda r dB_d(r) \right)^2 \frac{1}{\lambda^3} \right. \\
+ \left. \frac{1 - \lambda^3}{3(1 - \lambda)^3} \frac{[B_d(1) - B_d(\lambda)]^2}{1 - \lambda} - \frac{1 - \lambda^2}{(1 - \lambda)^4} \left( \int_\lambda^1 r dB_d(r) \right) [B_d(1) - B_d(\lambda)] \right. \\
+ \left. \frac{1 - \lambda}{(1 - \lambda)^4} \left( \int_\lambda^1 r dB_d(r) \right)^2 \right\}
\]

Given that \( \lim_{\lambda \to 0} \int_0^\lambda r dB_d(r) = 0 \) and \( \lim_{\lambda \to 1} \int_0^1 r dB_d(r) = 0 \), we can show that \( \limsup_{\lambda \to 0} M^*(\lambda) = \limsup_{\lambda \to 1} M^*(\lambda) = \infty \) almost surely (a.s.) for \( d^* \in (-0.5, 0] \) due to the law of iterated logarithms for a fractional Brownian motion.

Second, for \( d^* \in (0, 0.5) \),

\[
M^*(1) - M^*(\lambda) \\
= 12\kappa(d)^2 \left\{ \frac{B_d(1)^2}{3} - \left( \int_0^1 r dB_d(r) \right) B_d(1) + \left( \int_0^1 r dB_d(r) \right)^2 \right. \\
- \frac{B_d(\lambda)^2}{3\lambda} + \left( \int_0^\lambda r dB_d(r) \right) B_d(\lambda) - \left( \int_0^\lambda r dB_d(r) \right)^2 \frac{1}{\lambda^3} \right. \\
+ \left. \frac{1 - \lambda^2}{(1 - \lambda)^4} \left( \int_\lambda^1 r dB_d(r) \right) [B_d(1) - B_d(\lambda)] \right. \\
+ \left. \frac{1 - \lambda}{(1 - \lambda)^4} \left( \int_\lambda^1 r dB_d(r) \right)^2 \right\}
\]

\[
= 12\kappa(d)^2 \left\{ \frac{B_d(1)^2}{3} - \frac{B_d(\lambda)^2}{3\lambda} - \frac{1 - \lambda^3}{3(1 - \lambda)^3} \frac{[B_d(1) - B_d(\lambda)]^2}{1 - \lambda} + A(\lambda) \right\}
\]

\[
< 12\kappa(d)^2 \left\{ \frac{B_d(1)^2}{3} - \frac{B_d(\lambda)^2}{3\lambda} - \frac{[B_d(1) - B_d(\lambda)]^2}{3(1 - \lambda)} + A(\lambda) \right\}
\]

\[
= 12\kappa(d)^2 \left\{ - \frac{(\lambda B_d(1) - B_d(\lambda))^2}{3(1 - \lambda)} + A(\lambda) \right\}
\]

where \( A(\lambda) = -(\int_0^1 r dB_d(r))B_d(1) + (\int_0^1 r dB_d(r))^2 + (\int_0^\lambda r dB_d(r))B_d(\lambda)\lambda^{-2} - (\int_0^\lambda r dB_d(r))^2 \lambda^{-3} + (1 + \lambda(1 - \lambda)^{-3}(\int_\lambda^1 r dB_d(r))[B_d(1) - B_d(\lambda)] - (1 - \lambda)^{-3}(\int_\lambda^1 r dB_d(r))^2 \) and the inequality holds for \( \lambda \in (0, 1) \). For \( \lambda \in (0, \epsilon) \), \( -(\int_0^\lambda r dB_d(r))^2 \lambda^{-3} \) dominates the other terms in \( A(\lambda) \) because of the law of iterated logarithms for a fractional Brownian motion and the order of \( \lambda \). Similarly, so does \(-(\int_1^\lambda r dB_d(r))^2 (1 - \lambda)^{-3} \) for \( \lambda \in (1 - \epsilon, 1) \). It implies that \( M^*(0) = M^*(1) < M^*(\lambda) \) for some \( \lambda \in (0, \epsilon) \cup (1 - \epsilon, 1) \), which completes the proof.

Proof of Proposition 1. For any arbitrary \( z \) and an arbitrary projection matrix \( P \), we have \( z'Pz \leq z'z \), then for \( d^* \in (0.5, 1.5) \), it follows from Lemma 2 that

\[
M^*_T(T_1) \leq T^{-2(d+1)} \sum_{t=1}^T u_t^2 \Rightarrow \kappa(d)^2 \int_0^1 [B_d(r)]^2 dr
\]

for all \( T_1 \in [1, T] \). Since \( M^*_T(T_1) \) is uniformly bounded in probability, its limit \( M^*(\lambda) \) is also uniformly bounded in probability for \( \lambda \in (0, 1) \).
Proof of Theorem 6. The proof is similar to Theorem 2 in Bai (1998), who also provided a useful lemma, which we now state. Define a matrix \( H \) as follows; for arbitrary positive definite matrices \( A \) and \( B \) with \( A > B \ (p \times p) \)

\[
H = \begin{pmatrix}
(A - B)^{-1} - A^{-1} & -(A - B)^{-1} \\
-(A - B)^{-1} & (A - B)^{-1} + B^{-1}
\end{pmatrix}.
\]

Let \( D = (A - B)^{-1} + B^{-1} > 0, (A - B)^{-1} - A^{-1} = (A - B)^{-1}D^{-1}(A - B)^{-1}, \) and

\[
C = \begin{pmatrix}
(A - B)D & I \\
0 & D^{-1}(A - B)^{-1}
\end{pmatrix}.
\]

We obtain

\[
C'HC = \text{diag}(D, 0) \geq 0.
\]

It implies that \( H \) is a positive semidefinite matrix because \( C \) has full rank. For arbitrary vectors \( x \) and \( y \ (p \times 1) \), let \( z = (x', y')' \), then we have

\[
-z'Hz = x'A^{-1}x - y'B^{-1}y - (x - y)'(A - B)^{-1}(x - y) \leq 0. \tag{A.6}
\]

\( M^*(0) - M^*(\lambda) \leq 0 \) if and only if

\[
G(1)'Q(1)G(1) - G(\lambda)'Q(\lambda)G(\lambda) - [G(1) - G(\lambda)]'[Q(1) - Q(\lambda)]^{-1}[G(1) - G(\lambda)] \leq 0.
\]

The above inequality follows from Bai’s lemma (A.6) by letting \( A = Q(1), B = Q(\lambda), x = G(1), \) and \( y = G(\lambda) \). Furthermore, \( M^*(0) - M^*(\lambda) < 0 \) if and only if \(-\xi'H\xi < 0\) where \( \xi = (G(1), G(\lambda))' \) with \( G(\lambda) = \left(\int_0^\lambda B_d(r)dr, \int_0^\lambda rB_d(r)dr\right)' \). Let \( \Xi \) be an orthogonal matrix such that \( \Xi'H\Xi = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4) \) with \( \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \), where \( \lambda_i \)'s are the eigenvalues of \( H \). Because \( H \geq 0 \) and \( H \neq 0 \), the maximum eigenvalue of \( H \) is positive. Moreover,

\[
-\xi'H\xi = -(\Xi\xi)'\text{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)(\Xi\xi) \leq -\varsigma^2\lambda_1
\]

where \( \varsigma \) is the first component of \( \Xi\xi \). Given that \( G(\lambda) \) has a continuous distribution for each \( \lambda \), so does \( \xi \). Hence, \( \Xi\xi \) is a vector of continuous random variables and \( P(\varsigma^2 = 0) = 0 \), which implies that \(-\varsigma^2\xi < 0\) with probability 1. We conclude that \(-\xi'H\xi < 0\), which completes the proof. ■

References

