Graduate School of Economics, Hitotsubashi University

Discussion Paper Series No. 2022-03

In-fill asymptotic distribution of the change point estimator when estimating breaks one at a time

Toshikazu Tayanagi and Eiji Kurozumi

September 2022

In-fill asymptotic distribution of the change point estimator when estimating breaks one at a time *

Toshikazu Tayanagi and Eiji Kurozumi
Department of Economics, Hitotsubashi University

September 27, 2022

Abstract

In this study, we investigate the least squares (LS) estimator of a structural change point by the in-fill asymptotic theory, which has been recently used by Jiang, Wang and Yu (2018, 2020), when the model with two structural changes is estimated as the model with only a one-time structural change. We, hence, show that the finite sample distribution of the estimator has four peaks, which is different from the classical long-span asymptotic distribution, which contains only one peak. Conversely, the in-fill asymptotic distribution of the estimator has four peaks and can approximate the finite sample distribution very well. We also demonstrate that the estimator is consistent in the in-fill asymptotic framework with a relatively large magnitude of the break. In the latter case, the finite sample distribution of the estimator has only one peak and is well approximated by both the in-fill and long-span asymptotic theory.

*We thank the participants at economics and statistics seminar held at Hitotsubashi university for helpful discussions and comments. All errors are our responsibility. Eiji Kurozumi's research was supported by JSPS KAKENHI Grant Number 19K01585. Address correspondence to Toshikazu Tayanagi , Department of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan; e-mail: ed215004@g.hit-u.ac.jp

1 Introduction

This study investigates the break point estimator by the least squares method under continuous record asymptotics or the in-fill asymptotic scheme developed by such as Phillips (1987), Perron (1991), and Jiang, Wang and Yu (2018), among others. We suppose the level-shifts model, that has two break points and consider the case where they are estimated one at a time.

The break point estimator has been investigated in the statistics and econometrics literature. For example, Hinkley (1970) and Yao (1987) investigated the change point estimator using the maximum likelihood method. Conversely, Bai (1994) applied the least squares method for the estimation. He showed, by assuming that the samples before and after the break increase proportionally to the whole sample size while the break fraction is fixed–sometimes called the long-span asymptotic scheme–that the break fraction estimator is consistent, unimodal, and symmetric. However, the corresponding finite sample distribution can be trimodal and asymmetric, particularly when the break size is relatively small.

To explain the discrepancy between the finite sample and asymptotic distributions, Jiang, Wang and Yu (2018, 2020) investigated the break point estimator of the continuous time model with a one-time break and of the corresponding discrete model under the other scheme, called the in-fill asymptotic scheme. Here, sampling frequency goes to infinity in the fixed interval, or equivalently, the sampling interval becomes zero.[1] Under the in-fill asymptotic scheme, Jiang, Wang and Yu (2018) derived the asymptotic distribution of the break point estimator for the level shift model, which can successfully replicate the important properties of finite sample distribution; trimodality and asymmetry. [2]

In this study, we extend the methodology of Jiang, Wang and Yu (2018) to the local level model with twice shifts. Although two change points exist in the model, they can be estimated one at a time as considered by Chong (1995) and Bai (1997). They then fit the model with a one-time break to the two breaks model and showed that the estimated break fraction is consistent with either of the true break fractions. Using the long-span asymptotic scheme, Bai (1997) derived the asymptotic distribution of this break point estimator, which is unimodal and asymmetric. However, as the later

[1]Yu (2014) and Zhou and Yu (2015) also used the same technique.

[2]Casini and Perron (2018, 2021) studied a feasible break point estimator of the continuous time model; however, we do not cover this in the study.

section demonstrates, the corresponding finite sample distribution of the estimator can have four modes and be asymmetric. Hence, we use the in-fill asymptotic scheme and investigate the same break point estimator. We will show that the in-fill asymptotic scheme can capture the important properties of the finite sample distribution such as four peaks and asymmetry when the level shifts shrink to zero at the rate of the square root of the sampling interval (called "small shift" in this paper). Additionally, we investigate the case where the break size shrinks to zero at a slower rate than in the above case (called "large shift" in this paper). We derive the asymptotic distribution of the break point estimator under the in-fill asymptotic scheme, which is the same as that obtained by Bai (1997); it is unimodal and asymmetric. We will demonstrate that when the break size is relatively large, the finite sample distribution of the estimator has the same properties. Therefore, both the in-full and long-span asymptotic schemes can explain the finite sample properties of the estimator in this case.

The rest of the paper is organized as follows. In Section 2, we review the long-span scheme and demonstrate discrepancy between the finite sample and limiting distributions. In Section 3, we derive the in-fill asymptotic distribution in the case of the "small shift". Conversely, Section 4 deals with the case of the "large shift." Concluding remarks are given in Section 5. All proofs of the theoretical results are relegated to Appendix.

2 Long-span Asymptotic Theory and Finite Sample Distributions

In this section, we briefly review the long-span asymptotic theory developed by Bai (1997) and compare the limiting distribution of the break point estimator with the finite sample distribution. Consider the level-shifts model given by

$$y_t = \begin{cases} \mu_1 + X_t & \text{for } 1 \leq t \leq k_1^0, \\ \mu_2 + X_t & \text{for } k_1^0 + 1 \leq t \leq k_2^0, \\ \mu_3 + X_t & \text{if } k_2^0 + 1 \leq t \leq T, \end{cases} \tag{2.1}$$

where $\{X_t\}$ is a linear process satisfying the conditions given in the next section. Chong (1995) and Bai (1997) proposed estimating break dates one at a time despite the two breaks and showed that the estimator is consistent with one of the two break fractions. Hence, they fit the model with a one-time break and the (first) break point estimator is given by

$$\hat{k} = \underset{k=1,\cdots,T-1}{\arg\min}\, \{S_T(k)\}. \tag{2.2}$$

where $S_T(k)$ is the sum of the squared residuals given by

$$S_T(k) = \sum_{t=1}^{k}(y_t - \bar{Y}_k)^2 + \sum_{t=k+1}^{T}(y_t - \bar{Y}_k^*)^2, \quad \text{where} \quad \bar{Y}_k = \frac{1}{k}\sum_{t=1}^{k} y_t \quad \text{and} \quad \bar{Y}_k^* = \frac{1}{T-k}\sum_{t=k+1}^{T} y_t.$$

Proposition 4 in Bai (1997) gives the limiting distribution of $\hat{k} - k_1^0$ when

$$\frac{\tau_1^0}{\tau_2^0}(\mu_1 - \mu_2)^2 > \frac{1-\tau_2^0}{1-\tau_1^0}(\mu_2 - \mu_3)^2. \tag{2.3}$$

Figure 1 demonstrates the histogram of the corresponding limiting distribution under Gaussian assumption with $\tau_1^0 = 0.33$, $\tau_2^0 = 0.67$, $\mu_1 = 0$, $\mu_2 = 4$, and $\mu_3 = 1$. The limiting distribution is unimodal and skewed. Conversely, the finite sample distribution with the same parameters with $T = 120$ (the number of replication is $10,000$) is asymmetric and has four modes (Figure 2). As the finite sample distribution is rather different from the asymptotic distribution, we use the in-fill asymptotic theory in the next section.

3 In-fill Asymptotic Distribution with Small Shifts

In this section, we develop the in-fill asymptotic theory and show that the limiting distribution developed in this section has four modes, as in the case of finite samples given in Figure 2.

Let $\{\tilde{y}_{th}\}$ for $t = 1, 2, \dots, T$ be the discrete time observations of a continuous time process on $[0, 1]$, where h is the sampling interval and $Th = 1$. By extending Jiang, Wang and Yu (2018), we consider the discretized model with twice level shifts given by the following:

$$\tilde{y}_{th} - \tilde{y}_{(t-1)h} = \begin{cases} \mu_1\sqrt{h} + \sqrt{h}X_t & \text{for } 1 \le t \le k_1^0, \\ \mu_2\sqrt{h} + \sqrt{h}X_t & \text{for } k_1^0 + 1 \le t \le k_2^0, \\ \mu_3\sqrt{h} + \sqrt{h}X_t & \text{for } k_2^0 + 1 \le t \le T, \end{cases}$$

where k_1^0 and k_2^0 are the break dates and μ_i for $i = 1$, 2, and 3 are defined below. Corresponding break fractions are given by $\tau_1^0 = k_1^0/T$ and $\tau_2^0 = k_2^0/T$, which we assume holds not only asymptotically but also in finite samples for simplicity. By letting $y_t = (\tilde{y}_{th} - \tilde{y}_{(t-1)h})/\sqrt{h}$, the model is expressed as

$$y_t = \begin{cases} \mu_1 + X_t & \text{for } 1 \le t \le k_1^0, \\ \mu_2 + X_t & \text{for } k_1^0 + 1 \le t \le k_2^0, \\ \mu_3 + X_t & \text{for } k_2^0 + 1 \le t \le T, \end{cases} \tag{3.1}$$

$$\text{where} \quad \mu_i = \left(\mu + \frac{\delta_i}{\varepsilon}\right)\sqrt{h}$$

for $i = 1$, 2, and 3. We allow for serial correlation in X_t by considering a linear process provided by

$$X_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j}, \quad \text{where} \quad a(1) = \sum_{j=0}^{\infty} a_j \neq 0 \quad \text{and} \quad \sum_{j=0}^{\infty} j|a_j| < \infty.$$

For the innovations $\{\epsilon_t\}$, we make the following assumption:

Assumption 1 *Innovations $\{\epsilon_t\}$ are martingale differences satisfying $E[\epsilon_t|\mathcal{F}_{t-1}] = 0$, $E[\epsilon_t^2] = \sigma^2$, and there exists a $\delta > 0$ such that $\sup_t E[|\epsilon_t|^{2+\delta}] < \infty$, where $\mathcal{F}_t$ is the σ-filed generated by ϵ_s for $s \leq t$.*

For two structural change points, the following assumption is made throughout the study.

Assumption 2 *(i) $\tau_i^0 \in (0,1)$ for $i = 1$ and 2 with $\tau_1^0 < \tau_2^0$. (ii) $\delta_1 \neq \delta_2$ and $\delta_2 \neq \delta_3$.*

Assumption 2(i) implies that the break fractions are distinct and k_1^0 and k_2^0 are not too close each other. In contrast, Assumption 2(ii) guarantees that the structural changes occurred twice and there are three regimes.

We investigate the asymptotic behavior of $\hat{k}$ given by (2.2), the (first) break point estimator considered by Bai (1997), wherein break dates are estimated one at a time. As shown by Bai (1994), the least squares estimator $\hat{k}$ can be expressed as follows:

$$\hat{k} = \arg\max_k \left\{ \left(\sqrt{T} V_T(k)\right)^2 \right\}, \quad \text{where} \quad V_T(k) = \sqrt{\frac{k(T-k)}{T^2}} \left(\bar{Y}_k^* - \bar{Y}_k\right). \tag{3.2}$$

To develop the in-fill asymptotic theory, we assume that the sampling interval h goes to 0, whereas ε in μ_i is different from 0.

Assumption 3 *$h \to 0$ and $\varepsilon \neq 0$ is fixed.*

As $Th = 1$, $h \to 0$ implies that $T \to \infty$. Under Assumption 3, the magnitudes of the breaks are of order $\sqrt{h} = 1/\sqrt{T}$ and thus, as discussed in Elliott and Mller (2007) and Jiang, Wang and Yu (2018), $\hat{\tau} = \hat{k}/T$ cannot be consistent with τ_1^0 and τ_2^0, as provided in the following theorem.[3]

[3]Bai (1997) proposed to estimate the second break point by splitting the sample at the first estimator $\hat{k}$. However, because the first break fraction estimator is inconsistent, there is no room for estimating the second break point in our setting, thus we focus on only the first break date estimator $\hat{k}$.

Theorem 1 *Under Assumptions 1–3, the in-fill asymptotic distribution of $\hat{\tau}$ is given by the following:*

$$\hat{\tau} \xrightarrow{d} \arg\max_{0<\tau<1} \left\{ \left(\sqrt{\tau(1-\tau)} J(\tau) \right)^2 \right\}, \tag{3.3}$$

$$\text{where} \quad J(\tau) = \begin{cases} J_1(\tau) & \text{if } \tau \le \tau_1^0, \\ J_2(\tau) & \text{if } \tau_1^0 < \tau \le \tau_2^0, \\ J_3(\tau) & \text{if } \tau_2^0 < \tau, \end{cases}$$

$$J_1(\tau) = \tilde{B}(\tau,\sigma) + \frac{\delta_1}{\varepsilon} - \frac{\tau_1^0 - \tau}{1-\tau}\frac{\delta_1}{\varepsilon} - \frac{\tau_2^0 - \tau_1^0}{1-\tau}\frac{\delta_2}{\varepsilon} - \frac{1-\tau_2^0}{1-\tau}\frac{\delta_3}{\varepsilon},$$

$$J_2(\tau) = \tilde{B}(\tau,\sigma) + \frac{\tau_1^0}{\tau}\frac{\delta_1}{\varepsilon} + \frac{\tau - \tau_1^0}{\tau}\frac{\delta_2}{\varepsilon} - \frac{\tau_2^0 - \tau}{1-\tau}\frac{\delta_2}{\varepsilon} - \frac{1-\tau_2^0}{1-\tau}\frac{\delta_3}{\varepsilon},$$

$$J_3(\tau) = \tilde{B}(\tau,\sigma) + \frac{\tau_1^0}{\tau}\frac{\delta_1}{\varepsilon} + \frac{\tau_2^0 - \tau_1^0}{\tau}\frac{\delta_2}{\varepsilon} + \frac{\tau - \tau_2^0}{\tau}\frac{\delta_3}{\varepsilon} - \frac{\delta_3}{\varepsilon},$$

$$\text{and} \quad \tilde{B}(\tau,\sigma) = \frac{1}{\tau}\sigma a(1) B(\tau) - \frac{1}{1-\tau}\sigma a(1)(B(1) - B(\tau)).$$

Remark 1 *Our model* (3.1) *becomes a one-time structural break model when $\tau_2^0 = 1$ ($k_2^0 = T$) and $\delta_3 = \delta_2$ ($\mu_3 = \mu_2$). In this case, if $\{X_t\}$ is an i.i.d. sequence as considered in Jiang, Wang and Yu (2018), we have $a(1) = 1$ and thus, by the straightforward calculation, we can show that*

$$\begin{aligned} \sqrt{\tau(1-\tau)} J_1(\tau) &= \sigma \left(\frac{B(\tau) - \tau B(1)}{\sqrt{\tau(1-\tau)}} - \frac{(1-\tau_1^0)\sqrt{\tau}}{\sqrt{1-\tau}} \frac{(\delta_2 - \delta_1)}{\sigma\varepsilon} \right) \\ &= \sigma \left(\frac{B(\tau_1^0 + u) - (\tau_1^0 + u)B(1)}{\sqrt{(\tau_1^0 + u)(1 - \tau_1^0 - u)}} - \frac{(1-\tau_1^0)\sqrt{\tau_1^0 + u}}{\sqrt{1 - \tau_1^0 - u}} \frac{\delta^*}{\sigma\varepsilon} \right) \end{aligned}$$

for $\tau \le \tau_1^0$, where the second equality holds by defining $\tau = \tau_1^0 + u$ and $\delta^* = \delta_2 - \delta_1$. Conversely, for $\tau > \tau_1^0$,

$$\begin{aligned} \sqrt{\tau(1-\tau)} J_2(\tau) &= \sigma \left(\frac{B(\tau) - \tau B(1)}{\sqrt{\tau(1-\tau)}} - \frac{\tau_1^0 \sqrt{(1-\tau)}}{\sqrt{\tau}} \frac{(\delta_2 - \delta_1)}{\sigma\varepsilon} \right) \\ &= \sigma \left(\frac{B(\tau_1^0 + u) - (\tau_1^0 + u)B(1)}{\sqrt{(\tau_1^0 + u)(1 - \tau_1^0 - u)}} - \frac{\tau_1^0 \sqrt{1 - \tau_1^0 - u}}{\sqrt{\tau_1^0 + u}} \frac{\delta^*}{\sigma\varepsilon} \right). \end{aligned}$$

These distributions are the same as those in Theorem 4.2 of Jiang, Wang and Yu (2018). Therefore, our result in Theorem 1 includes the in-fill asymptotic theory developed by Jiang, Wang and Yu (2018).

To investigate distributional property (3.3), we draw the histogram with the same parameters as in the previous section; $\tau_1^0 = 0.33$, $\tau_2^0 = 0.67$, $\sigma^2 = 1$, $a(1) = 1$, $\delta_1/\varepsilon = 0$, $\delta_2/\varepsilon = 4$, and $\delta_3/\varepsilon = 1$. Figure 3 shows that the limiting distribution has four modes around two break points and the ends of the samples. Hence, the in-full asymptotic distribution can successfully replicate the important property of the finite sample distribution in Figure 2. That is, in the case where break dates are estimated one at a time with relatively small shifts, the in-fill asymptotics is useful to capture the distributional property of the estimator in finite samples.

4 In-fill Asymptotic Distribution with Large Shifts

In this section, we investigate the asymptotic behavior of the break point estimator when the magnitudes of the breaks are larger than those given by Assumption 3. While break fraction estimator is inconsistent when the breaks are small under Assumption 3 (Theorem 1), we may expect that it would be consistent with either of the two break fractions for the larger breaks.

To consider the in-full asymptotic theory with large breaks, we make the following assumption:

Assumption 4 *(i)* $h \to 0$ *and* $\varepsilon \to 0$*. (ii)* $\varepsilon \log T \to 0$*. (iii)* $\sqrt{h}/\varepsilon \to 0$.

As the magnitudes of the breaks are given by $(\delta_2 - \delta_1)\sqrt{h}/\varepsilon$ and $(\delta_3 - \delta_2)\sqrt{h}/\varepsilon$, Assumption 4(i) implies the larger break sizes compared to Assumption 3, in which ε is fixed. However, we still assume the shrinking shifts of the breaks by (iii). Assumptions 4(ii) and (iii) impose the converging speed of ε; it goes to zero strictly faster than $1/\log T = -1/\log h$ but slower than $1/\sqrt{T} = \sqrt{h}$. This speed is essentially the same as that supposed in Bai (1997).

Whether $\hat{\tau}$ converges to τ_1^0 or τ_2^0 depends on the relative magnitude of the two breaks. Let $U_T(k/T) = S_T(k)/T$.

Assumption 5

$$\text{plim} \left(\frac{\sqrt{h}}{\varepsilon} \right)^{-2} [U_T(k_1^0/T) - U_T(k_2^0/T)] < 0.$$

Assumption 5 corresponds to Assumption B2 in Bai (1997). To obtain the probability limit given in Assumption 5, we note that $U_T(k_1^0/T)$ and $U_T(k_2^0/T)$ can be expressed as, from (6.3) and (6.11) in the appendix,

$$\begin{aligned}U_T(k_1^0/T) =&\frac{(k_2^0-k_1^0)}{T}\left(\frac{1}{T-k_1^0}(T-k_2^0)(\mu_2-\mu_3)\right)^2\\&+\frac{(T-k_2^0)}{T}\left(\frac{1}{T-k_1^0}(k_2^0-k_1^0)(\mu_3-\mu_2)\right)^2+\frac{1}{T}\sum_{t=1}^{T}X_t^2+R_{1T}(k_1^0),\\=&\frac{(T-k_2^0)(k_2^0-k_1^0)}{T(T-k_1^0)}\left(\frac{(\delta_2-\delta_3)\sqrt{h}}{\varepsilon}\right)^2+\frac{1}{T}\sum_{t=1}^{T}X_t^2+R_{1T}(k_1^0)\end{aligned}$$

and

$$U_T(k_2^0/T) =\frac{k_1^0(k_2^0-k_1^0)}{k_2^0T}\left(\frac{(\delta_1-\delta_2)\sqrt{h}}{\varepsilon}\right)^2+\frac{1}{T}\sum_{t=1}^{T}X_t^2+R_{2T}(k_2^0),$$

where $R_{1T}(k)$ and $R_{2T}(k)$ are defined by (6.4) and (6.12), which are shown to be $O_p(h/\varepsilon)$ in the appendix. Then, we have the following:

$$\begin{aligned}&\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2}\left(U_T(k_1^0/T)-U_T(k_2^0/T)\right)\\&=\frac{(T-k_2^0)(k_2^0-k_1^0)}{T(T-k_1^0)}(\delta_2-\delta_3)^2-\frac{k_1^0(k_2^0-k_1^0)}{k_2^0T}(\delta_1-\delta_2)^2+\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2}(R_{1T}(k_1^0)-R_{2T}(k_2^0))\\&\xrightarrow{p}\frac{(1-\tau_2^0)(\tau_2^0-\tau_1^0)}{1-\tau_1^0}(\delta_2-\delta_3)^2-\frac{\tau_1^0}{\tau_2^0}(\tau_2^0-\tau_1^0)(\delta_1-\delta_2)^2.\end{aligned}$$

Hence, Assumption 5 can be written as follows:

$$\frac{1-\tau_2^0}{1-\tau_1^0}(\delta_2-\delta_3)^2<\frac{\tau_1^0}{\tau_2^0}(\delta_1-\delta_2)^2, \tag{4.1}$$

which is equivalent to (2.3). Hence, Assumption 5 implies that the first break dominates the second one in terms of break magnitude and the relative time span of the regimes.

To derive the in-fill asymptotic distribution of $\hat{\tau}$, we first show that it is consistent with τ_1^0.

Proposition 1 *Under Assumptions 1, 2, 4, and 5, we have the following:*

$$\hat{\tau}-\tau_1^0=O_p(\varepsilon).$$

Note that because $\hat{\tau}$ is inconsistent when ε is fixed (Theorem 1), we observe that Assumption 5 is key for the consistency.

Proposition 1 relies on the inequality given by Assumption 5. If the inequality is in the opposite direction, $\hat{\tau}$ becomes consistent with τ_2^0. As discussed by Bai (1997), $\hat{\tau}$ will converge in probability to either τ_1^0 or τ_2^0 with each probability 1/2 if the equality holds.

As we obtained the consistency of $\hat{\tau}$ by Proposition 1, we can focus on the behavior of $S_T(k)$ only in the small neighborhood of τ_1^0. Let us define $D_T = \{k : T\eta \leq k \leq T\tau_2^0(1-\eta)\}$, where η is a small positive value such that $\tau_1^0 \in (\eta, T\tau_2^0(1-\eta))$ and $D_M = \left\{k : |k - k_1^0| \leq M\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2}\right\}$ for some given large value of M. Let us define the intersection D_T and complement of D_M as $D_{T,M^c} = \{k : T\eta \leq k \leq T\tau_2^0(1-\eta),\ |k - k_1^0| > M\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2}\}$. In the appendix, we show that $S_T(k)$ cannot be minimized on D_{T,M^c} with the probability approaching one. We then obtain the following proposition.

Proposition 2 *Under Assumptions 1, 2, 4, and 5, for every $\epsilon > 0$, there exists an $M < \infty$ independent of T such that, for all large T,*

$$P\left(T|\hat{\tau} - \tau_1^0| > M\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2}\right) < \epsilon.$$

Proposition 2 implies that

$$T\left(\frac{\sqrt{h}}{\varepsilon}\right)^2 (\hat{\tau} - \tau_1^0) = O_p(1)$$

and gives the asymptotic order to derive the limiting distribution. Note that because $Th = 1$, $\hat{\tau}$ approaches τ_1^0 at a rate of ϵ^2.

The in-fill asymptotic distribution of $\hat{\tau}$ is given by the following theorem.

Theorem 2 *Under Assumptions 1, 2, 4, and 5, we have*

$$T\left(\frac{\sqrt{h}}{\varepsilon}\right)^2 (\delta_2 - \delta_1)^2(\hat{\tau} - \tau_1^0) \xrightarrow{d} \sigma^2 a(1)^2 \underset{u \in (-\infty,\infty)}{\arg\min} \{\Gamma(u, \lambda_1)\},$$

where

$$\lambda_1 = \frac{1-\tau_2^0}{1-\tau_1^0}\frac{\delta_3 - \delta_2}{\delta_2 - \delta_1},$$
$$\Gamma(u,\lambda) = \begin{cases} 2B_1(u) + u(1-\lambda) & \text{if } u \geq 0, \\ 2B_2(-u) + |u|(1+\lambda) & \text{if } u < 0, \end{cases}$$

and $B_1(\cdot)$ and $B_2(\cdot)$ are two independent standard Brownian motions on $[0,\infty)$.

The in-fill asymptotic distribution derived in Theorem 2 is the same as the long-span asymptotic distribution obtained in Proposition 8 of Bai (1997). This limiting distribution is generally unimodal and asymmetric. Figure 4 shows histograms of the asymptotic distribution obtained in Theorem 2 for the same set of parameters as before. Moreover, the distribution in Figure 4 is unimodal and skewed to the left, which is rather different from Figure 2. However, as demonstrated by Figure 5, in which the data generating process is given by (3.1) for $\tau_1^0 = 0.33$, $\tau_2^0 = 0.67$, $\mu = 0$, $\delta_1 = 0$, $\delta_2 = 4$, $\delta_3 = 1$, $h = 1/T = 1/120$ and $\varepsilon = h^{1/4}$ (large shifts), the finite sample distribution can be approximated well by the in-fill asymptotic distribution with $\varepsilon \to 0$.

5 Concluding Remarks

In this study, we applied the in-fill asymptotic scheme to the level-shifts model when the break date is estimated by fitting the misspecified one-time break model. We first showed that the finite sample distribution of the break point estimator obtained by the least squares estimation has four modes, which is quite different from the traditional asymptotic distribution by Bai (1997). We then derived the in-fill asymptotic distribution with small shifts, which has four modes and is a better approximation to the finite distribution than that in the long-span asymptotic scheme. Additionally, by considering the model with large shifts, which still shrink to zero but at a slower rate than in the small shift case, the in-fill asymptotic distribution was shown to become the same as that derived in the long-span scheme. In this case, the break fraction estimator is consistent, unimodal, and asymmetric.

Although the finite sample properties of the break point estimator could be replicated using the in-fill asymptotic scheme, constructing the confidence set under the assumption of small shifts is difficult because the break fraction estimator is inconsistent, which results in the inconsistent estimation of the nuisance parameters required for constructing the confidence set. We may need to impose some mild regularity conditions on the model considered by Casini and Perron (2018, 2021), and this is our future research.

6 Appendix

Proof of Theorem 1: Following Bai (1994), (2.2) is expressed as

$$\hat{k} = \arg\min_{k}\{S_T(k)\} = \arg\max_{k}\left\{\left(\sqrt{T}V_T(k)\right)^2\right\}, \tag{6.1}$$

where

$$V_T(k) = \sqrt{\frac{k(T-k)}{T^2}}\left(\bar{Y}_k^* - \bar{Y}_k\right), \tag{6.2}$$

$$\bar{Y}_k = \frac{1}{k}\sum_{t=1}^{k} y_t, \quad \bar{Y}_k^* = \frac{1}{T-k}\sum_{t=k+1}^{T} y_t.$$

Under Assumptions 1 and 2, we have, by the functional central limit theorem (FCLT),

$$\frac{\sqrt{T}}{k}\sum_{t=1}^{k} X_t = \frac{T}{k}\frac{1}{\sqrt{T}}\sum_{t=1}^{k} X_t \Rightarrow \frac{1}{\tau}a(1)\sigma B(\tau),$$

$$\frac{\sqrt{T}}{T-k}\sum_{t=k+1}^{T} X_t = \frac{T}{T-k}\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T} X_t \Rightarrow \frac{1}{1-\tau}a(1)\sigma(B(1)-B(\tau)),$$

for $k = [\tau T]$ with a given $\tau \in (0,1)$, where $B(\cdot)$ is a standard Brownian motion on $[0,1]$. We show that the limiting distribution of (6.2) depends on the regime wherein k is located.

For $k = [\tau T] \le k_1^0$, we have, by the FCLT,

$$\begin{aligned}
&\sqrt{T}\left(\frac{1}{k}\sum_{t=1}^{k} y_t - \frac{1}{T-k}\sum_{t=k+1}^{T} y_t\right) \\
=&\sqrt{T}\left(\frac{1}{k}\sum_{t=1}^{k} y_t - \frac{1}{T-k}\left(\sum_{t=k+1}^{k_1^0} y_t + \sum_{t=k_1^0+1}^{k_2^0} y_t + \sum_{t=k_2^0+1}^{T} y_t\right)\right) \\
=&\frac{\sqrt{T}}{k}\sum_{t=1}^{k} X_t + \sqrt{T}\left(\mu + \frac{\delta_1}{\varepsilon}\right)\sqrt{h} \\
&- \frac{\sqrt{T}}{T-k}\left(\sum_{t=k+1}^{T} X_t + (k_1^0 - k)\left(\mu + \frac{\delta_1}{\varepsilon}\right)\sqrt{h} + (k_2^0 - k_1^0)\left(\mu + \frac{\delta_2}{\varepsilon}\right)\sqrt{h} + (T - k_2^0)\left(\mu + \frac{\delta_3}{\varepsilon}\right)\sqrt{h}\right) \\
=&\frac{\sqrt{T}}{k}\sum_{t=1}^{k} X_t - \frac{\sqrt{T}}{T-k}\sum_{t=k+1}^{T} X_t \\
&+ \frac{\delta_1}{\varepsilon}\sqrt{h}\sqrt{T} - \frac{\sqrt{T}}{T-k}(k_1^0 - k)\frac{\delta_1}{\varepsilon}\sqrt{h} - \frac{\sqrt{T}}{T-k}(k_2^0 - k_1^0)\frac{\delta_2}{\varepsilon}\sqrt{h} - \frac{\sqrt{T}}{T-k}(T - k_2^0)\frac{\delta_3}{\varepsilon}\sqrt{h} \\
\Rightarrow&\frac{1}{\tau}a(1)\sigma B(\tau) - \frac{1}{1-\tau}a(1)\sigma(B(1)-B(\tau)) + \frac{\delta_1}{\varepsilon} - \frac{\tau_1^0 - \tau}{1-\tau}\frac{\delta_1}{\varepsilon} - \frac{\tau_2^0 - \tau_1^0}{1-\tau}\frac{\delta_2}{\varepsilon} - \frac{1-\tau_2^0}{1-\tau}\frac{\delta_3}{\varepsilon}.
\end{aligned}$$

Similarly, for $k_1^0 < k \le k_2^0$, we have

$$
\begin{aligned}
&\sqrt{T}\left(\frac{1}{k}\sum_{t=1}^{k} y_t - \frac{1}{T-k}\sum_{t=k+1}^{T} y_t\right) \\
=&\sqrt{T}\left(\frac{1}{k}\left(\sum_{t=1}^{k_1^0} y_t + \sum_{t=k_1^0+1}^{k} y_t\right) - \frac{1}{T-k}\left(\sum_{t=k+1}^{k_2^0} y_t + \sum_{t=k_2^0+1}^{T} y_t\right)\right) \\
=&\frac{\sqrt{T}}{k}\left(\sum_{t=1}^{k} X_t + k\mu\sqrt{h} + k_1^0\frac{\delta_1\sqrt{h}}{\varepsilon} + (k-k_1^0)\frac{\delta_2\sqrt{h}}{\varepsilon}\right) \\
&-\frac{\sqrt{T}}{T-k}\left(\sum_{t=k+1}^{T} X_t + (T-k)\mu\sqrt{h} + (k_2^0-k)\frac{\delta_2\sqrt{h}}{\varepsilon} + (T-k_2^0)\frac{\delta_3\sqrt{h}}{\varepsilon}\right) \\
=&\frac{\sqrt{T}}{k}\sum_{t=1}^{k} X_t - \frac{\sqrt{T}}{T-k}\sum_{t=k+1}^{T} X_t \\
&+\frac{\sqrt{T}}{k}k_1^0\frac{\delta_1\sqrt{h}}{\varepsilon} + \frac{\sqrt{T}}{k}(k-k_1^0)\frac{\delta_2\sqrt{h}}{\varepsilon} - \frac{\sqrt{T}}{T-k}(k_2^0-k)\frac{\delta_2\sqrt{h}}{\varepsilon} - \frac{\sqrt{T}}{T-k}(T-k_2^0)\frac{\delta_3\sqrt{h}}{\varepsilon} \\
\Rightarrow&\frac{1}{\tau}a(1)\sigma B(\tau) - \frac{1}{1-\tau}a(1)\sigma(B(1)-B(\tau)) \\
&+\frac{\tau_1^0}{\tau}\frac{\delta_1}{\varepsilon} + \frac{\tau-\tau_1^0}{\tau}\frac{\delta_2}{\varepsilon} - \frac{\tau_2^0-\tau}{1-\tau}\frac{\delta_2}{\varepsilon} - \frac{1-\tau_2^0}{1-\tau}\frac{\delta_3}{\varepsilon},
\end{aligned}
$$

and for $k_2^0 < k$,

$$
\begin{aligned}
\sqrt{T}\left(\frac{1}{k}\sum_{t=1}^{k} y_t - \frac{1}{T-k}\sum_{t=k+1}^{T} y_t\right) =&\frac{\sqrt{T}}{k}\left(\sum_{t=1}^{k_1^0} y_t + \sum_{t=k_1^0+1}^{k_2^0} y_t + \sum_{t=k_2^0+1}^{k} y_t\right) - \frac{\sqrt{T}}{T-k}\sum_{t=k+1}^{T} y_t \\
=&\frac{\sqrt{T}}{k}\left(\sum_{t=1}^{k} X_t + k\mu\sqrt{h} + k_1^0\frac{\delta_1}{\varepsilon}\sqrt{h} + (k_2^0-k_1^0)\frac{\delta_2\sqrt{h}}{\varepsilon} + (k-k_2^0)\frac{\delta_3\sqrt{h}}{\varepsilon}\right) \\
&-\frac{\sqrt{T}}{T-k}\left(\sum_{t=k+1}^{T} X_t + (T-k)\mu\sqrt{h} + (T-k)\frac{\delta_3\sqrt{h}}{\varepsilon}\right) \\
\Rightarrow&\frac{1}{\tau}a(1)\sigma B(\tau) - \frac{1}{1-\tau}a(1)\sigma(B(1)-B(\tau)) \\
&+\frac{\tau_1^0}{\tau}\frac{\delta_1}{\varepsilon} + \frac{\tau_2^0-\tau_1^0}{\tau}\frac{\delta_2}{\varepsilon} + \frac{\tau-\tau_2^0}{\tau}\frac{\delta_3}{\varepsilon} - \frac{\delta_3}{\varepsilon}.
\end{aligned}
$$

Therefore, we have

$$
\hat{\tau} \xrightarrow{d} \underset{0<\tau<1}{\arg\max}\left\{\left(\sqrt{\tau(1-\tau)}J(\tau)\right)^2\right\},
$$

where $J(\tau)$ is defined in the main statement of Theorem 1.■

We need several lemmas to prove Propositions 1, 2, and Theorem 2.

Lemma 1 *Under Assumption 1, there exists an $M < \infty$ such that, for all i and all $j > i$*

(a) $\left|E\left[\left(\sum_{t=1}^{i} X_t\right)\left(\sum_{s=i+1}^{j} X_s\right)\right]\right| \leq M$,

(b) $\left|E\left[\frac{1}{j-i}\left(\sum_{t=i+1}^{j} X_t\right)^2\right]\right| \leq M$.

Proof: (a) and (b) are given by Lemma 11 and (A.12) in Bai (1997), respectively.■

Lemma 2 *Under Assumptions 1 and 2, there exists an $M < \infty$ such that, for $i = 1,\ 2,$ and 3,*

$$T|E[R_{iT}(k)] - E[R_{iT}(k_1^0)]| \leq \frac{|k_1^0 - k|}{T}M.$$

Proof: These relations are given by (A.14) in Bai (1997).■

Lemma 3 *Under Assumptions 1,2,4, and 5 in which $\varepsilon \to 0$ and $h \to 0$, the following relations hold:*

(a) $\sup_{1\leq k\leq T}\left|U_T(k/T) - E[U_T(k/T)] - T^{-1}\sum_{t=1}^{T}(X_t^2 - E[X_t^2])\right| = O_p(h/\varepsilon)$.

(b) *There exists $C > 0$ for all large T such that*

$$E[S_T(k)] - E[S_T(k_1^0)] \geq C(\sqrt{h}/\varepsilon)^2|k - k_1^0|.$$

Proof: (a) Let

$$A_{Tk} = \frac{1}{k}\sum_{t=1}^{k} X_t \quad \text{and} \quad A_{Tk}^* = \frac{1}{T-k}\sum_{t=k+1}^{T} X_t.$$

To prove (a), we consider the three cases; (i) $k \leq k_1^0$, (ii) $k_1^0 < k \leq k_2^0$, and (iii) $k_2^0 < k$. For $k \leq k_1^0$, it can be shown that

$$\left|U_T(k/T) - E[U_T(k/T)] - T^{-1}\sum_{t=1}^{T}(X_t^2 - E[X_t^2])\right| = |R_{1T}(k) - E[R_{1T}(k)]|\,,$$

where, as given by (A5) in Bai (1997),

$$
\begin{aligned}
U_T(k/T) =& \frac{1}{T} S_T(k) \\
=& \frac{(k_1^0 - k)}{T} a_{Tk}^2 + \frac{(k_2^0 - k_1^0)}{T} b_{Tk}^2 + \frac{(T - k_2^0)}{T} c_{Tk}^2 + \frac{1}{T} \sum_{t=1}^{T} X_t^2 + R_{1T}(k), \qquad (6.3) \\
R_{1T}(k) =& \frac{1}{T} \left[2a_{Tk} \sum_{t=k+1}^{k_1^0} X_t + 2b_{Tk} \sum_{t=k_1^0+1}^{k_2^0} X_t + 2c_{Tk} \sum_{t=k_2^0+1}^{T} X_t \right] \\
& - \frac{2}{T} [(k_1^0 - k) a_{Tk} + (k_2^0 - k_1^0) b_{Tk} + (T - k_2^0) c_{Tk}] A_{Tk}^* \\
& - \frac{k}{T} (A_{Tk})^2 - \frac{T-k}{T} (A_{Tk}^*)^2, \qquad (6.4) \\
a_{Tk} =& \frac{1}{T-k} \{ (T - k_1^0)(\mu_1 - \mu_2) + (T - k_2^0)(\mu_2 - \mu_3) \}, \qquad (6.5) \\
b_{Tk} =& \frac{1}{T-k} \{ (k_1^0 - k)(\mu_2 - \mu_1) + (T - k_2^0)(\mu_2 - \mu_3) \}, \qquad (6.6) \\
c_{Tk} =& \frac{1}{T-k} \{ (k_1^0 - k)(\mu_2 - \mu_1) + (k_2^0 - k)(\mu_3 - \mu_2) \}. \qquad (6.7)
\end{aligned}
$$

Note that a_{Tk}, b_{Tk}, and c_{Tk} are $O(\sqrt{h}/\varepsilon)$ from the definitions of μ_i for $i = 1$, 2, and 3, respectively.

To investigate $R_{1T}(k)$, we observe, by the FCLT, that

$$
\frac{1}{T} \sum_{t=k+1}^{k_1^0} X_t = O_p(T^{-1/2}), \quad \frac{1}{T} \sum_{t=k_1^0+1}^{k_2^0} X_t = O_p(T^{-1/2}), \quad \text{and} \quad \frac{1}{T} \sum_{t=k_2^0+1}^{T} X_t = O_p(T^{-1/2}).
$$

Thus, the first term on the right-hand side of $R_{1T}(k)$ in (6.4) is $O_p(T^{-1/2}\sqrt{h}/\varepsilon) = O_p(h/\varepsilon)$.

To evaluate the second term, because A_{Tk}^* is $O_p(T^{-1/2}) = O_p(h^{1/2})$ uniformly in $k \leq k_1^0$ by the FCLT, we can observe that the second term on the right-hand side of (6.4) is $O_p(h/\epsilon)$ uniformly in $k \leq k_1^0$.

For the third term, we use Hájek-Rényi inequality given in Proposition 1 in Bai (1994); for a given $\alpha > 0$, there exists some constant $C_1 > 0$ such that

$$
P \left(\sup_{k=1,\cdots,T} \frac{1}{\sqrt{k}} \left| \sum_{t=1}^{k} X_t \right| > \alpha \right) \leq \frac{C_1}{\alpha^2} \log T. \qquad (6.8)
$$

(6.8) implies that $\sup_k \frac{1}{\sqrt{k}} \left| \sum_{t=1}^{k} X_t \right| = O_p \left(\sqrt{\log T} \right)$ and thus

$$
\frac{k}{T} (A_{Tk})^2 = \frac{1}{T} \left(\frac{1}{\sqrt{k}} \sum_{t=1}^{k} X_t \right)^2 = \frac{1}{T} O_p \left(\left(\sqrt{\log T} \right)^2 \right) = o_p(h/\varepsilon)
$$

uniformly in $k \in [1, T]$ under Assumption 4.

Because $A^*_{Tk} = O_p(T^{-1/2})$, the fourth term on the right-hand side of (6.4) becomes

$$\frac{T-k}{T}\left(A^*_{Tk}\right)^2 = O_p(T^{-1}) = o_p(h/\varepsilon)$$

uniformly in $k \in [1, k_1^0]$.

Combining these results, we have

$$R_{1T}(k) = O_p(h/\varepsilon) \text{ uniformly in } k \leq k_1^0. \tag{6.9}$$

Next, we investigate $E[R_{1T}(k)]$. Note that the expectations of the first and second terms on the right-hand side of (6.4) are zero, while those of the last two terms are $O(T^{-1})$ by Lemma 1(b). Then, we observe that

$$E[R_{1T}(k)] = O(T^{-1}) = O(h) = o(h/\varepsilon) \text{ uniformly in } k \leq k_1^0. \tag{6.10}$$

From (6.9) and (6.10), we observe that

$$|R_{1T}(k) - E[R_{1T}(k)]| = O_p(h/\varepsilon) \text{ uniformly in } k \leq k_1^0.$$

In the case where $k_1^0 < k \leq k_2^0$, $U_T(k/T)$ can be expressed as (see (A.9) in Bai (1997)),

$$\begin{aligned} U_T(k/T) =& \frac{k_1^0(k-k_1^0)}{kT}(\mu_2-\mu_1)^2 + \frac{(k_2^0-k)(T-k_2^0)}{T(T-k)}(\mu_3-\mu_2)^2 \\ &+ \frac{1}{T}\sum_{t=1}^{T} X_t^2 + R_{2T}(k), \end{aligned} \tag{6.11}$$

where

$$\begin{aligned} R_{2T}(k) =& \frac{1}{T}\left[2d_{Tk}\sum_{t=1}^{k_1^0} X_t + 2e_{Tk}\sum_{t=k_1^0+1}^{k} X_t + 2f_{Tk}\sum_{t=k+1}^{k_2^0} X_t + 2g_{Tk}\sum_{t=k_2^0+1}^{T} X_t\right] \\ &- \frac{2}{T}\left[k_1^0 d_{Tk} + (k-k_1^0)e_{Tk}\right]A_{Tk} - \frac{2}{T}\left[(k_2^0-k)f_{Tk} + (T-k_2^0)g_{Tk}\right]A^*_{Tk} \\ &- \frac{k}{T}(A_{Tk})^2 - \frac{T-k}{T}(A^*_{Tk})^2, \end{aligned} \tag{6.12}$$

$$\begin{aligned} d_{Tk} =& [(k-k_1^0)/k](\mu_1-\mu_2), \\ e_{Tk} =& (k_1^0/k)(\mu_2-\mu_1), \\ f_{Tk} =& [(T-k_2^0)/(T-k)](\mu_2-\mu_3), \\ g_{Tk} =& [(k_2^0-k)/(T-k)](\mu_3-\mu_2), \end{aligned}$$

and we consider

$$\left| U_T(k/T) - E[U_T(k/T)] - T^{-1}\sum_{t=1}^{T}(X_t - E[X_t^2]) \right| = |R_{2T}(k) - E[R_{2T}(k)]|\,.$$

In exactly the same manner as in the case of $k \leq k_1^0$, we can show that (a) holds.

For $k_2^0 < k$, as given by (A.11) in Bai (1997),

$$U_T(k/T) = \frac{k_1^0}{T}h_{Tk}^2 + \frac{k_2^0 - k_1^0}{T}p_{Tk}^2 + \frac{k - k_2^0}{T}q_{Tk}^2 + \frac{1}{T}\sum_{t=1}^{T}X_t^2 + R_{3T}(k), \tag{6.13}$$

where

$$\begin{aligned} R_{3T}(k) =& \frac{1}{T}\left[2h_{Tk}\sum_{t=1}^{k_1^0}X_t + 2p_{Tk}\sum_{t=k_1^0+1}^{k_2^0}X_t + 2q_{Tk}\sum_{t=k_2^0+1}^{k}X_t\right] \\ &- \left(\frac{2}{T}[k_1^0 h_{Tk} + (k_2^0 - k_1^0)p_{Tk} + (k - k_2^0)q_{Tk}]\right)A_{Tk} \\ &- \frac{k}{T}(A_{Tk})^2 - \frac{(T-k)}{T}(A_{Tk}^*)^2, \qquad (6.14) \\ h_{Tk} =& \frac{1}{k}[(k - k_1^0)(\mu_1 - \mu_2) + (k - k_2^0)(\mu_2 - \mu_3)], \\ p_{Tk} =& \frac{1}{k}[k_1^0(\mu_2 - \mu_1) + (k - k_2^0)(\mu_2 - \mu_3)], \\ q_{Tk} =& \frac{1}{k}[k_1^0(\mu_2 - \mu_1) + k_2^0(\mu_3 - \mu_2)], \end{aligned}$$

and we consider

$$\left| U_T(k/T) - E[U_T(k/T)] - T^{-1}\sum_{t=1}^{T}(X_t - E[X_t^2]) \right| = |R_{3T}(k) - E[R_{3T}(k)]|\,.$$

The order of the preceding equality is obtained in the same manner.

(b) For $k \leq k_1^0$, the left-hand side of Lemma 3(b) becomes, as given by (A.17) in Bai (1997),

$$\begin{aligned} &E[S_T(k)] - E[S_T(k_1^0)] \\ =& \frac{k_1^0 - k}{(1 - k/T)(1 - k_1^0/T)}\left[(1 - k_1^0/T)(\mu_1 - \mu_2) + (1 - k_2^0/T)(\mu_2 - \mu_3)\right]^2 \\ &+ T\left(E[R_{1T}(k)] - E[R_{1T}(k_1^0)]\right). \end{aligned} \tag{6.15}$$

We first note that

$$(1 - k_1^0/T)(\mu_1 - \mu_2) + (1 - k_2^0/T)(\mu_2 - \mu_3) \neq 0,$$

which can be obtained in the same manner as (A.18) in Bai (1997). Therefore, there exists a $C > 0$ such that

$$\begin{aligned}
&\frac{k_1^0 - k}{(1-k/T)(1-k_1^0/T)}\left[(1-k_1^0/T)(\mu_1-\mu_2)+(1-k_2^0/T)(\mu_2-\mu_3)\right]^2\\
=&\frac{k_1^0 - k}{(1-\tau)(1-\tau_1^0)}\left(\frac{\sqrt{h}}{\varepsilon}\right)^2\left[(1-\tau_1^0)(\delta_1-\delta_2)+(1-\tau_2^0)(\delta_2-\delta_3)\right]^2\\
\geq&|k_1^0-k|\left(\frac{\sqrt{h}}{\varepsilon}\right)^2 C.
\end{aligned}$$

On the other hand, for the second term on the right-hand side of (6.15), we observe by Lemma 2 that

$$T|E[R_{1T}(k)-E[R_{1T}(k_1^0)]|\leq\frac{|k_1^0-k|}{T}M=|k_1^0-k|Mh=|k_1^0-k|\times o\left(\frac{h}{\varepsilon^2}\right).$$

Thus, we have

$$\begin{aligned}
E[S_T(k)]-E[S_T(k_1^0)]&\geq C\left(\frac{\sqrt{h}}{\varepsilon}\right)^2|k-k_1^0|-|k_1^0-k|Mh\\
&\geq C\left(\frac{\sqrt{h}}{\varepsilon}\right)^2|k-k_1^0|/2.
\end{aligned}$$

For $k_1^0 < k \leq k_2^0$, it can be shown that, as given by inequality below (A.20) in Bai (1997),

$$\begin{aligned}
E[S_T(k)]-E[S_T(k_1^0)]\geq&(k-k_1^0)\frac{k_2^0}{k}\left[\frac{k_1^0}{k_2^0}(\mu_2-\mu_1)^2-\frac{(T-k_2^0)}{(T-k_1^0)}(\mu_3-\mu_2)^2\right]\\
&+T\left[E[R_{2T}(k)]-E[R_{2T}(k_1^0)]\right].
\end{aligned}$$

Because

$$\frac{k_1^0}{k_2^0}(\mu_2-\mu_1)^2-\frac{(T-k_2^0)}{(T-k_1^0)}(\mu_3-\mu_2)^2=\frac{\tau_1^0}{\tau_2^0}(\mu_2-\mu_1)^2-\frac{(1-\tau_2^0)}{(1-\tau_1^0)}(\mu_3-\mu_2)^2>0$$

as implied by (4.1), and using Lemma 2, we have

$$\begin{aligned}
E[S_T(k)]-E[S_T(k_1^0)]\geq&(k-k_1^0)\left(\frac{\sqrt{h}}{\varepsilon}\right)^2\frac{\tau_2^0}{\tau}\left[\frac{\tau_1^0}{\tau_2^0}(\delta_2-\delta_1)^2-\frac{(1-\tau_2^0)}{(1-\tau_1^0)}(\delta_3-\delta_2)^2\right]\\
&+T\left[E[R_{2T}(k)]-E[R_{2T}(k_1^0)]\right]\\
\geq&(k-k_1^0)\left(\frac{\sqrt{h}}{\varepsilon}\right)^2C-M|k_1^0-k|h\\
\geq&(k-k_1^0)\left(\frac{\sqrt{h}}{\varepsilon}\right)^2C/2
\end{aligned}$$

for some $M > 0$ and $C > 0$.

For $k > k_2^0$, $S_T(k)$ becomes, from (6.13),

$$S_T(k) = k_1^0 h_{Tk}^2 + (k_2^0 - k_1^0) p_{Tk}^2 + (k - k_2^0) q_{Tk}^2 + \sum_{t=1}^{T} X_t^2 + TR_{3T}(k).$$

Similarly to the case $k < k_1^0$, it can be shown that

$$ES_T(k) - ES_T(k_1^0) \geq (k - k_1^0) \left(\frac{\sqrt{h}}{\varepsilon}\right)^2 C$$

for some constant C.■

Lemma 4 *Under Assumptions 1,2,4, and 5 in which $\varepsilon \to 0$ and $h \to 0$, for every ϵ, there exists an $M < \infty$ such that*

$$P\left(\min_{k \in D_{T,M^c}} \left\{S_T(k) - S_T(k_1^0)\right\} \leq 0\right) < \epsilon,$$

where $D_{T,M^c} = \left\{k : T\eta \leq k \leq T\tau_2^0(1-\eta),\ |k - k_1^0| > M\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2}\right\}$.

Proof: The proof proceeds similarly to that of Lemma 4 of Bai (1997). For $k \leq k_1^0$,

$$\begin{aligned}
&S_T(k) - S_T(k_1^0) \\
&= S_T(k) - E[S_T(k)] - (S_T(k_1^0) - E[S_T(k_1^0)]) + E[S_T(k)] - E[S_T(k_1^0)] \\
&\geq S_T(k) - E[S_T(k)] - (S_T(k_1^0) - E[S_T(k_1^0)]) + C|k - k_1^0| \left(\frac{\sqrt{h}}{\varepsilon}\right)^2 \qquad (6.16)
\end{aligned}$$

for some $C > 0$, where the inequality holds by Lemma 3(b). Then, $S_T(k) - S_T(k_1^0) \leq 0$ implies that

$$C\left(\frac{\sqrt{h}}{\varepsilon}\right)^2 \leq \frac{1}{|k - k_1^0|} \left|S_T(k) - E[S_T(k)] - \left(S_T(k_1^0) - E[S_T(k_1^0)]\right)\right|.$$

By Lemma 2, we have, for some $M < \infty$,

$$\begin{aligned}
&\left|S_T(k) - E[S_T(k)] - (S_T(k_1^0) - E[S_T(k_1^0)])\right| \\
&= \left|T(R_{1T}(k) - E[R_{1T}(k)]) - T(R_{1T}(k_1^0) - E[R_{1T}(k_1^0)])\right| \\
&\leq \left|T(R_{1T}(k) - R_{1T}(k_1^0))\right| + M\frac{|k - k_1^0|}{T}.
\end{aligned}$$

Therefore, it is sufficient to show that for every $\eta > 0$ and $\epsilon > 0$, there exists an $M > 0$ such that

$$P\left(\eta\left(\frac{\sqrt{h}}{\varepsilon}\right)^2 < \sup_{k \in D_{TM^c}} \left\{\frac{1}{|k - k_1^0|} |T(R_{1T}(k) - E[R_{1T}(k)])| + \frac{M}{T}\right\}\right) < \epsilon,$$

but, because $M/T = O(h) = o\left((\sqrt{h}/\varepsilon)^2\right)$, we shall show that

$$P\left(\sup_{k\in D_{TM^c}}\left\{\frac{1}{|k-k_1^0|}\left|T(R_{1T}(k)-E[R_{1T}(k)])\right|\right\} > \eta\left(\frac{\sqrt{h}}{\varepsilon}\right)^2\right) < \epsilon.$$

For $k \le k_1^0$, note that by (6.4),

$$\begin{aligned}
&T|R_{1T}(k)-R_{1T}(k_1^0)|\\
&=2a_{Tk}\sum_{t=k+1}^{k_1^0}X_t + 2(b_{Tk}-b_{Tk_1^0})\sum_{t=k_1^0+1}^{k_2^0}X_t + 2(c_{Tk}-c_{Tk_1^0})\sum_{t=k_2^0+1}^{T}X_t\\
&\quad -2(k_1^0-k)a_{Tk}A_{Tk}^* - 2(k_2^0-k_1^0)(b_{Tk}A_{Tk}^* - b_{Tk_1^0}A_{Tk_1^0}^*)\\
&\quad -2(T-k_2^0)(c_{Tk}A_{Tk}^* - c_{Tk_1^0}A_{Tk_1^0}^*)\\
&\quad +\left(k_1^0A_{Tk_1^0}^2 - kA_{Tk}^2\right) + \left((T-k_1^0)A_{Tk_1^0}^{*2} - (T-k)A_{Tk}^{*2}\right).
\end{aligned} \tag{6.17}$$

Subsequently, we shall show that each term on the right-hand side of (6.17) divided by $k_1^0 - k$ is of smaller order than $(\sqrt{h}/\varepsilon)^2$ uniformly.

As $|a_{Tk}| \le \frac{\sqrt{h}}{\varepsilon}C_1$ for some $C_1 < \infty$ uniformly, the first term on the right-hand side of (6.17) is evaluated as, by Hájek-Rényi inequality,

$$\begin{aligned}
&P\left(\sup_{k\le k_1^0 - M\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2}}\left|\frac{a_{Tk}}{k_1^0-k}\sum_{t=k+1}^{k_1^0}X_t\right| > \eta\left(\frac{\sqrt{h}}{\varepsilon}\right)^2\right)\\
&\le P\left(\sup_{k\le k_1^0 - M\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2}}\left\{\frac{1}{k_1^0-k}\left|\sum_{t=k+1}^{k_1^0}X_t\right|\right\} > \eta\left(\frac{\sqrt{h}}{\varepsilon}\right)C_1^{-1}\right) \le \frac{C_1^2C}{\eta^2M}
\end{aligned} \tag{6.18}$$

for some $C < \infty$. By taking a large value of M, the right-hand side of (6.18) becomes small.

For the second term on the right-hand side of (6.17), because $T - k_2^0 < T - k_1^0$, we have

$$\begin{aligned}
&|b_{Tk} - b_{Tk_1^0}|\\
&=\left|\frac{1}{T-k}\left\{(k_1^0-k)(\mu_2-\mu_1) + (T-k_2^0)(\mu_2-\mu_3)\right\} - \frac{1}{T-k_1^0}(T-k_2^0)(\mu_2-\mu_3)\right|\\
&=\left|\frac{k_1^0-k}{T-k}(\mu_2-\mu_1) + \frac{k-k_1^0}{(T-k)(T-k_1^0)}(T-k_2^0)(\mu_2-\mu_3)\right|\\
&\le\left|\frac{k_1^0-k}{T-k}(\delta_2-\delta_1)\frac{\sqrt{h}}{\varepsilon}\right| + \left|\frac{k_1^0-k}{T-k}(\delta_2-\delta_3)\frac{\sqrt{h}}{\varepsilon}\right|\\
&\le\left|\frac{k_1^0-k}{T-k}\right|C_2\frac{\sqrt{h}}{\varepsilon},
\end{aligned}$$

for some $C_2 > 0$. Then, we have

$$\left|\frac{b_{Tk} - b_{Tk_1^0}}{k_1^0 - k}\sum_{t=k_1^0+1}^{k_2^0} X_t\right| \leq C_2 \frac{1}{T-k}\left(\frac{\sqrt{h}}{\varepsilon}\right)\left|\sum_{t=k_1^0+1}^{k_2^0} X_t\right|$$
$$\leq C \times O_p\left(\frac{h}{\varepsilon}\right)$$
$$= o_p\left(\left(\frac{\sqrt{h}}{\varepsilon}\right)^2\right).$$

For the third term on the right-hand side of (6.17), we note that

$$\begin{aligned}
&|c_{Tk} - c_{Tk_1^0}| \\
&= \left|\frac{1}{T-k}\left\{(k_1^0 - k)(\mu_2 - \mu_1) + (k_2^0 - k)(\mu_3 - \mu_2)\right\} - \frac{1}{T-k_1^0}(k_2^0 - k_1^0)(\mu_3 - \mu_2)\right| \\
&= \left|\frac{k_1^0 - k}{T-k}(\mu_2 - \mu_1) + \left(\frac{k_2^0 - k}{T-k} - \frac{k_2^0 - k_1^0}{T-k_1^0}\right)(\mu_3 - \mu_2)\right| \\
&\leq \left|\frac{k_1^0 - k}{T-k}(\delta_2 - \delta_1)\frac{\sqrt{h}}{\varepsilon}\right| + \left|\frac{k_1^0 - k}{T-k}(\delta_3 - \delta_2)\frac{\sqrt{h}}{\varepsilon}\right| \\
&\leq \left|\frac{k_1^0 - k}{T-k}\right| C_3 \frac{\sqrt{h}}{\varepsilon},
\end{aligned}$$

for some $C_3 > 0$, where the first inequality holds because

$$\left(\frac{k_2^0 - k}{T-k} - \frac{k_2^0 - k_1^0}{T-k_1^0}\right) = \frac{(T-k_2^0)(k_1^0 - k)}{(T-k)(T-k_1^0)} \leq \frac{k_1^0 - k}{T-k},$$

as $\frac{T-k_2^0}{T-k_1^0} < 1$. Then, it can be seen that

$$\left|\frac{c_{Tk} - c_{Tk_1^0}}{k_1^0 - k}\sum_{t=k_2^0+1}^{T} X_t\right| \leq \frac{C_3}{T-k}\frac{\sqrt{h}}{\varepsilon}\left|\sum_{t=k_2^0+1}^{T} X_t\right|$$
$$\leq C \times O_p\left(\frac{h}{\varepsilon}\right)$$
$$= o_p\left(\left(\frac{\sqrt{h}}{\varepsilon}\right)^2\right).$$

The fourth term on the right-hand side of (6.17) divided by $(k_1^0 - k)\left(\frac{\sqrt{h}}{\varepsilon}\right)^2$ is

$$(k_1^0 - k)^{-1}\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2}(k_1^0 - k)a_{Tk}A_{Tk}^* = o_p(1),$$

because $|a_{Tk}| = O_p\left(\frac{\sqrt{h}}{\varepsilon}\right)$ and $A_{Tk}^* = O_p(T^{-1/2}) = O_p(h^{1/2})$.

For the fifth term on the right-hand side of (6.17), we observe that

$$
\begin{aligned}
&(k_2^0 - k_1^0)(b_{Tk}A_{Tk}^* - b_{Tk_1^0}A_{Tk_1^0}^*) \\
&\quad =(k_2^0 - k_1^0)[b_{Tk} - b_{Tk_1^0}]A_{Tk}^* - (k_2^0 - k_1^0)b_{Tk_1^0}[A_{Tk_1^0}^* - A_{Tk}^*]. \qquad (6.19)
\end{aligned}
$$

The first term on the right-hand side of (6.19) divided by $(k_1^0 - k)(\sqrt{h}/\varepsilon)^2$ becomes

$$
\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2} \frac{k_2^0 - k_1^0}{k_1^0 - k}[b_{Tk} - b_{Tk_1^0}]A_{Tk}^* = \left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2} O_p\left(\frac{\sqrt{h}}{\varepsilon}T^{-1/2}\right) = O_p(\varepsilon),
$$

while the second term is given by

$$
\begin{aligned}
&\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2} \frac{k_2^0 - k_1^0}{k_1^0 - k} b_{Tk_1^0}[A_{Tk_1^0}^* - A_{Tk}^*] \\
&\quad = \left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2} \frac{k_2^0 - k_1^0}{k_1^0 - k} \frac{1}{T - k_1^0}(T - k_2^0)(\mu_2 - \mu_3)[A_{Tk_1^0}^* - A_{Tk}^*] \\
&\quad = \left(\frac{\sqrt{h}}{\varepsilon}\right)^{-1} \frac{k_2^0 - k_1^0}{k_1^0 - k} \frac{T - k_2^0}{T - k_1^0}(\delta_2 - \delta_3)[A_{Tk_1^0}^* - A_{Tk}^*].
\end{aligned}
$$

Note that

$$
\begin{aligned}
\frac{k_2^0 - k_1^0}{k_1^0 - k}[A_{Tk_1^0}^* - A_{Tk}^*] &= \frac{k_2^0 - k_1^0}{(T-k)(T-k_1^0)} \sum_{t=k_1^0+1}^{T} X_t - \frac{k_2^0 - k_1^0}{(T-k)(k_1^0 - k)} \sum_{t=k+1}^{k_1^0} X_t \\
&= O_p(\sqrt{h}) - o_p\left(\frac{\sqrt{h}}{\varepsilon}\right). \qquad (6.20)
\end{aligned}
$$

Here, the first term on the last expression holds by FCLT, and the order of the second term is obtained from (6.18). Hence, the fifth term on the right-hand side of (6.17) divided by $(k_1^0 - k)(\sqrt{h}/\varepsilon)^2$ is $o_p(1)$.

The sixth term on the right-hand side of (6.17) divided by $(k_1^0 - k)(\sqrt{h}/\varepsilon)^2$ is treated similar to the fifth term.

For the seventh term on the right-hand side of (6.17), we note that $kA_{Tk}^2 = O_p(1)$ uniformly on D_{T,M^c} because k is proportional to T on D_{T,M^c}. Thus, we have

$$
(k_1^0 - k)^{-1}\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2}\left|k_1^0 A_{Tk_1^0}^2 - kA_{Tk}^2\right| \leq \frac{1}{M}\left(\frac{\sqrt{h}}{\varepsilon}\right)^{2}\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2} O_p(1) = \frac{1}{M}O_p(1),
$$

which can be small for all large values of T by choosing a large value of M.

Similarly, the eighth term on the right-hand side of (6.17) can be evaluated as

$$
(k_1^0 - k)^{-1}\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2}\left|(T - k_1^0)(A_{Tk_1^0}^{*2} - (T-k)A_{Tk}^{*2}\right| \leq \frac{1}{M}O_p(1)
$$

on D_{T,M^c}.

As all the terms on the right-hand side of (6.17) divided by $(k_1^0 - k)(\sqrt{h}/\varepsilon)^2$ converge in probability to 0, we have

$$P\left(\sup_{k\in D_{TM^c}}\left\{\frac{1}{|k-k_1^0|}\left|T(R_{1T}(k)-E[R_{1T}(k)])\right|\right\}>\eta\left(\frac{\sqrt{h}}{\varepsilon}\right)^2\right)<\epsilon.$$

The case where $k_1^0 < k$ is proved in the same manner and thus omitted.■

Proof of Proposition 1: The proof proceeds similarly to that of Corollary 1 in Bai (1997). For some $C > 0$, we have

$$\begin{aligned}
&S_T(k) - S_T(k_1^0)\\
&=S_T(k) - E[S_T(k)] - [S_T(k_1^0) - E[S_T(k_1^0)]] + E[S_T(k)] - E[S_T(k_1^0)]\\
&\quad+\left(\sum_{t=1}^T X_t^2 - E[X_t^2]\right) - \left(\sum_{t=1}^T X_t^2 - E[X_t^2]\right)\\
&\geq -2\sup_{1\leq j\leq T}\left\{\left|S_T(j) - E[S_T(j)] - \left(\sum_{t=1}^T X_t^2 - E[X_t^2]\right)\right|\right\} + E[S_T(k)] - E[S_T(k_1^0)]\\
&\geq -2\sup_{1\leq j\leq T}\left\{\left|S_T(j) - E[S_T(j)] - \left(\sum_{t=1}^T X_t^2 - E[X_t^2]\right)\right|\right\} + C\left(\frac{\sqrt{h}}{\varepsilon}\right)^2|k-k_1^0|,
\end{aligned}$$

where the last inequality holds by Lemma 3(b). As $S_T(\hat{k}) - S_T(k_1^0) \leq 0$, the above inequality implies

$$|\hat{k}-k_1^0| \leq C^{-1}\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2} 2\sup_{1\leq j\leq T}\left\{\left|S_T(j) - E[S_T(j)] - \left(\sum_{t=1}^T X_t^2 - E[X_t^2]\right)\right|\right\}.$$

Dividing both sides by T, we have, by Lemma 3(a),

$$|\hat{\tau}-\tau_1^0| \leq 2C^{-1}\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2} O_p\left(\frac{h}{\varepsilon}\right) = O_p\left(\varepsilon\right).■$$

Proof of Proposition 2: As $\hat{\tau}$ is consistent with τ_0 by Proposition 1, we can observe that, for any given value of $\epsilon > 0$, $P(\hat{k} \notin D_T) \leq \epsilon$ for all large T. Thus, we have, using Lemma 4,

$$\begin{aligned}
P\left(T|\hat{\tau}-\tau_1^0| > M\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2}\right) &\leq P(\hat{k}\notin D_T) + P\left(\hat{k}\in D_T, |\hat{k}-k| > M\left(\frac{\sqrt{h}}{\varepsilon}\right)^{-2}\right)\\
&\leq \epsilon + P\left(\min_{k\in D_{TM^c}}\{S_T(k) - S_T(k_1^0)\} \leq 0\right) \leq 2\epsilon.■
\end{aligned}$$

Proof of Theorem 2: The proof proceeds similarly to Proposition 8 in Bai (1997). Given the convergence order of $\hat{\tau}$ obtained in Proposition 2, we focus on the $O((\sqrt{h}/\varepsilon)^{-2})$ neighborhood of k_1^0. More precisely, let M be an arbitrary large positive value and k be given by, for $s \in [-M, M]$,

$$k = k_1^0 + \ell, \quad \ell = s \left(\frac{\sqrt{h}}{\varepsilon} \right)^{-2}.$$

Then, we can observe that $\hat{k} = k_1^0 + \hat{\ell}$ and

$$\begin{aligned} \hat{\ell} &= \hat{k} - k_1^0 \\ &= \arg\min_{\ell} \{ S_T(k_1^0 + \ell) - S_T(k_1^0) \}. \end{aligned} \tag{6.21}$$

Thus, we investigate the asymptotic behavior of $S_T(k_1^0 + \ell) - S_T(k_1^0)$.

First, we consider the case where $k > k_1^0$. This implies $\ell > 0$ and $s > 0$.

For $\ell > 0$, define

$$\hat{\mu}_1 = \frac{1}{k_1^0} \sum_{t=1}^{k_1^0} y_t, \quad \hat{\mu}_2 = \frac{1}{T - k_1^0} \sum_{t=k_1^0+1}^{T} y_t,$$
$$\hat{\mu}_1^* = \frac{1}{k_1^0 + \ell} \sum_{t=1}^{k_1^0+\ell} y_t, \quad \hat{\mu}_2^* = \frac{1}{T - k_1^0 - \ell} \sum_{t=k_1^0+\ell+1}^{T} y_t.$$

It is not difficult to show that

$$\hat{\mu}_1^* - \mu_1 = O_p(T^{-1/2}) = O_p(h^{1/2}), \quad \hat{\mu}_1 - \mu_1 = O_p(T^{-1/2}) = O_p(h^{1/2}),$$
$$\hat{\mu}_2 - \mu_2 - \frac{1 - \tau_2^0}{1 - \tau_1^0}(\mu_3 - \mu_2) = O_p(T^{-1/2}) = O_p(h^{1/2}), \quad \hat{\mu}_i^* - \hat{\mu}_i = O_p(\varepsilon h^{1/2}) \text{ for } i = 1, 2.$$

We decompose the sums of the squared residuals into

$$S_T(k_1^0 + \ell) = \sum_{t=1}^{k_1^0} (y_t - \hat{\mu}_1^*)^2 + \sum_{t=k_1^0+1}^{k_1^0+\ell} (y_t - \hat{\mu}_1^*)^2 + \sum_{t=k_1^0+\ell+1}^{T} (y_t - \hat{\mu}_2^*)^2, \tag{6.22}$$
$$S_T(k_1^0) = \sum_{t=1}^{k_1^0} (y_t - \hat{\mu}_1)^2 + \sum_{t=k_1^0+1}^{k_1^0+\ell} (y_t - \hat{\mu}_2)^2 + \sum_{t=k_1^0+\ell+1}^{T} (y_t - \hat{\mu}_2)^2. \tag{6.23}$$

The differences between the two first and third terms on the right-hand side of (6.22) and (6.23) are given by

$$\sum_{t=1}^{k_1^0} (y_t - \hat{\mu}_1^*)^2 - \sum_{t=1}^{k_1^0} (y_t - \hat{\mu}_1)^2 = k_1^0 (\hat{\mu}_1^* - \hat{\mu}_1)^2 = O_p(T)(O_p(\varepsilon\sqrt{h}))^2 = O_p(\varepsilon^2),$$

and

$$\sum_{t=k_1^0+\ell+1}^{T}(y_t-\hat{\mu}_2^*)^2-\sum_{t=k_1^0+\ell+1}^{T}(y_t-\hat{\mu}_2)^2=-(T-k_1^0-\ell)(\hat{\mu}_2-\hat{\mu}_2^*)^2$$
$$=O_p(T)(O_p(\varepsilon\sqrt{h}))^2=O_p(\varepsilon^2).$$

Meanwhile, the difference between the two second terms on the right-hand side of (6.22) and (6.23) becomes

$$\sum_{t=k_1^0+1}^{k_1^0+\ell}(y_t-\hat{\mu}_1^*)^2-\sum_{t=k_1^0+1}^{k_1^0+\ell}(y_t-\hat{\mu}_2)^2=2(\hat{\mu}_2-\hat{\mu}_1^*)\sum_{t=k_1^0+1}^{k_1^0+\ell}X_t+\ell\left\{(\mu_2-\hat{\mu}_1^*)^2-(\mu_2-\hat{\mu}_2)^2\right\}.$$

Note that

$$\begin{aligned}\hat{\mu}_2-\hat{\mu}_1^* &= \{(\mu_2-\mu_1)+(\mu_1-\hat{\mu}_1^*)\}-(\mu_2-\hat{\mu}_2)\\ &= \{(\mu_2-\mu_1)+(\mu_1-\hat{\mu}_1^*)\}+\frac{1-\tau_2^0}{1-\tau_1^0}(\mu_3-\mu_2)+O_p(h^{1/2})\\ &= (\mu_2-\mu_1)+\lambda_1(\mu_2-\mu_1)+O_p(h^{1/2})\\ &= (1+\lambda_1)\left(\frac{(\delta_2-\delta_1)\sqrt{h}}{\varepsilon}\right)+O_p(h^{1/2}),\end{aligned}$$

and

$$\begin{aligned}(\mu_2-\hat{\mu}_1^*)^2-(\mu_2-\hat{\mu}_2)^2 =&\{(\mu_2-\mu_1)-(\hat{\mu}_1^*-\mu_1)\}^2-(\mu_2-\hat{\mu}_2)^2\\ =&\{(\mu_2-\mu_1)-O_p(h^{1/2})\}^2-\{\lambda_1(\mu_2-\mu_1)+O_p(h^{1/2})\}^2\\ =&(\mu_2-\mu_1)^2+O_p(h)-2(\mu_2-\mu_1)O_p(h^{1/2})-(\lambda_1(\mu_2-\mu_1))^2-O_p(h/\varepsilon)\\ =&(\mu_2-\mu_1)^2(1-\lambda_1^2)+O_p(h/\varepsilon)\\ =&\left(\frac{(\delta_2-\delta_1)\sqrt{h}}{\varepsilon}\right)^2(1-\lambda_1^2)+O_p(h/\varepsilon).\end{aligned}$$

Thus,

$$\begin{aligned}&2(\hat{\mu}_2-\hat{\mu}_1^*)\sum_{t=k_1^0+1}^{k_1^0+\ell}X_t+\ell\left\{(\mu_2-\hat{\mu}_1^*)^2-(\mu_2-\hat{\mu}_2)^2\right\}\\ &=2\left\{(1+\lambda_1)\left(\frac{(\delta_2-\delta_1)\sqrt{h}}{\varepsilon}\right)+O_p(h^{1/2})\right\}\sum_{t=k_1^0+1}^{k_1^0+\ell}X_t+\ell\left\{\left(\frac{(\delta_2-\delta_1)\sqrt{h}}{\varepsilon}\right)^2(1-\lambda_1^2)+O_p(h/\varepsilon)\right\}\\ &\Rightarrow 2(1+\lambda_1)(\delta_2-\delta_1)\sigma a(1)B_1(s)+s(\delta_2-\delta_1)^2(1-\lambda_1^2), \qquad (6.24)\end{aligned}$$

where $B_1(\cdot)$ is a standard Brownian motion on $[0, \infty)$. Then, we obtain

$$\begin{aligned}
\hat{s} &= \arg\min_{s>0} \left\{2(1+\lambda_1)(\delta_2-\delta_1)\sigma a(1) B_1(s) + s(\delta_2-\delta_1)^2(1-\lambda_1^2)\right\} \\
&= \arg\min_{s>0} \left\{(1+\lambda_1)\{2\sigma a(1) B_1(s(\delta_2-\delta_1)^2) + s(\delta_2-\delta_1)^2(1-\lambda)\}\right\} \\
&= (\delta_2-\delta_1)^{-2}\sigma^2 a(1)^2 \arg\min_{u>0} \left\{2\sigma a(1) B_1(\sigma^2 a(1)^2 u) + \sigma^2 a(1)^2 u(1-\lambda)\right\} \\
&= (\delta_2-\delta_1)^{-2}\sigma^2 a(1)^2 \arg\min_{u>0} \left\{\sigma^2 a(1)^2\{2B_1(u) + u(1-\lambda)\}\right\} \\
&\overset{\mathrm{d}}{=} (\delta_2-\delta_1)^{-2}\sigma^2 a(1)^2 \arg\min_{u>0} \left\{\Gamma(u,\lambda_1)\right\}, \qquad (6.25)
\end{aligned}$$

where the last equality in distribution is obtained by letting $u = s(\delta_2-\delta_1)^2\sigma^{-2}a(1)^{-2}$.

Then, we consider the case where $k = k_1^0 - \ell$ and $\ell = s(\sqrt{h}/\varepsilon)^2 \geq 0$. As in the case of $k > k_1^0$, we observe that

$$S_T(k_1^0-\ell) = \sum_{t=1}^{k_1^0-\ell}(y_t-\hat{\mu}_1^*)^2 + \sum_{t=k_1^0-\ell+1}^{k_1^0}(y_t-\hat{\mu}_2^*)^2 + \sum_{t=k_1^0+1}^{T}(y_t-\hat{\mu}_2^*)^2, \qquad (6.26)$$

$$S_T(k_1^0) = \sum_{t=1}^{k_1^0-\ell}(y_t-\hat{\mu}_1)^2 + \sum_{t=k_1^0-\ell+1}^{k_1^0}(y_t-\hat{\mu}_1)^2 + \sum_{t=k_1^0+1}^{T}(y_t-\hat{\mu}_2)^2. \qquad (6.27)$$

The difference between the two first and third terms on the right-hand side of (6.26) and (6.27) is shown to be $o_p(1)$ in the same manner as in the case of $k > k_1^0$, whereas the difference between the two second terms converges in distribution to

$$\begin{aligned}
&\sum_{t=k_1^0-\ell+1}^{k_1^0}(y_t-\hat{\mu}_2^*)^2 - \sum_{t=k_1^0-\ell+1}^{k_1^0}(y_t-\hat{\mu}_1)^2 \\
&\quad = 2(\hat{\mu}_1-\hat{\mu}_2^*)\sum_{t=k_1^0-\ell+1}^{k_1^0} X_t + \ell((\hat{\mu}_2^*-\mu_1)^2 - (\hat{\mu}_1-\mu_1)^2) \\
&\quad \Rightarrow 2(\delta_2-\delta_1)(1+\lambda_1)\sigma a(1) B_2(s) + |s|(\delta_2-\delta_1)^2(1+\lambda_1)^2, \qquad (6.28)
\end{aligned}$$

where $B_2(\cdot)$ is a standard Brownian motion on $[0,\infty)$ independent of $B_1(\cdot)$.Thus, we have

$$\begin{aligned}
\hat{s} &= \arg\min_{s>0} \left\{2(\delta_2-\delta_1)(1+\lambda_1)\sigma a(1) B_2(s) + |s|(\delta_2-\delta_1)^2(1+\lambda_1^2)\right\} \\
&= \arg\min_{s<0} \left\{2(\delta_2-\delta_1)(1+\lambda_1)\sigma a(1) B_2(-s) + |s|(\delta_2-\delta_1)^2(1+\lambda_1^2)\right\} \\
&\overset{\mathrm{d}}{=} (\delta_2-\delta_1)^{-2}\sigma^2 a(1)^2 \arg\min_{u<0} \left\{\Gamma(u,\lambda_1)\right\}. \qquad (6.29)
\end{aligned}$$

By definition, $\hat{s} = \hat{\ell}\left(\frac{\sqrt{h}}{\varepsilon}\right)^2 = T\left(\frac{\sqrt{h}}{\varepsilon}\right)^2(\tau - \tau_1^0)$ and thus we obtain the theorem.■

References

Bai, Jushan. 1994. "Least squares estimation of a shift in linear processes." *Journal of Time Series Analysis*, 15(5): 453–472.

Bai, Jushan. 1997. "Estimating multiple breaks one at a time." *Econometric Theory*, 13(3): 315–352.

Casini, Alessandro, and Pierre Perron. 2018. "Continuous Record Asymptotics for Change-Points Models." arXiv preprint arXiv:1803.10881.

Casini, Alessandro, and Pierre Perron. 2021. "Continuous record Laplace-based inference about the break date in structural change models." *Journal of Econometrics*, 224(1): 3–21.

Chong, Terence Tai Leung. 1995. "Partial parameter consistency in a misspecified structural change model." *Economics Letters*, 49(4): 351–357.

Elliott, Graham, and Ulrich K. Mller. 2007. "Confidence sets for the date of a single break in linear time series regressions." *Journal of Econometrics*, 141(2): 1196–1218.

Hinkley, David V. 1970. "Inference about the change-point in a sequence of random variables." *Biometrika*, 57(1): 1–17.

Jiang, Liang, Xiaohu Wang, and Jun Yu. 2018. "New distribution theory for the estimation of structural break point in mean." *Journal of Econometrics*, 205(1): 156–176.

Jiang, Liang, Xiaohu Wang, and Jun Yu. 2020. "In-fill asymptotic theory for structural break point in autoregressions." *Econometric Reviews*, 40(4): 359–386.

Perron, Pierre. 1991. "A continuous time approximation to the unstable first-order autoregressive process: The case without an intercept." *Econometrica*, 59: 211–236.

Phillips, Peter C. B. 1987. "Towards a unified asymptotic theory for autoregression." *Biometrika*, 74: 535–547.

Yao, Yi-Ching. 1987. "Approximating the Distribution of the Maximum Likelihood Estimate of the Change-Point in a Sequence of Independent Random Variables." *Annals of Statistics*, 15(3): 1321–1328.

Yu, Jun. 2014. "Econometric Analysis of Continuous Time Models: A Survey of Peter Philip's Work and Some New Results." *Econometric Theory*, 30(4): 737–774.

Zhou, Qiankun, and Jun Yu. 2015. "Asymptotic theory for linear diffusions under alternative sampling schemes." *Economics Letters*, 128: 1–5.

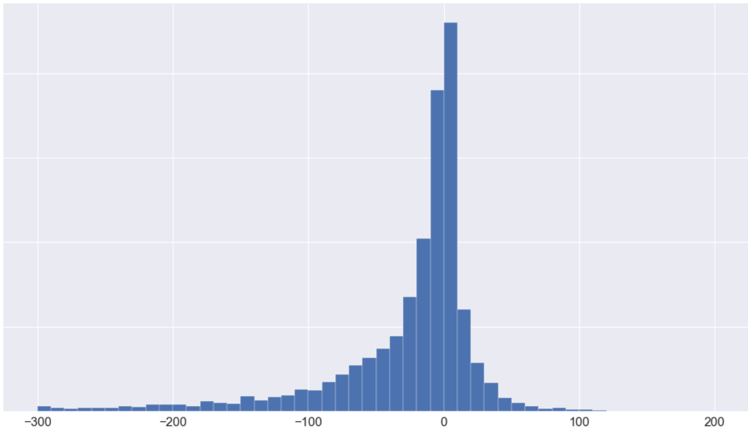

Figure 1: The long-span asymptotic distribution of Bai (1997) for $\tau_1^0 = 0.33$, $\tau_2^0 = 0.67$, $\mu = 0$, $\delta_1 = 0$, $\delta_2 = 4$, and $\delta_3 = 1$

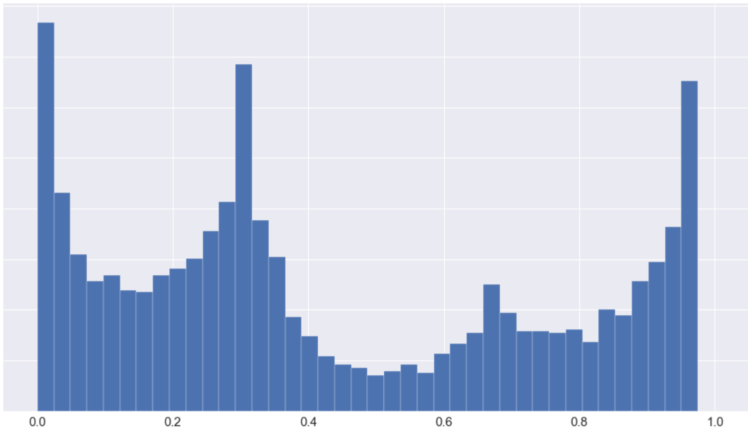

Figure 2: The finite sample distribution from DGP (2.1) for $\tau_1^0 = 0.33$, $\tau_2^0 = 0.67$, $T = 120$,$\mu = 0$, $\delta_1 = 0$, $\delta_2 = 4$, and $\delta_3 = 1$

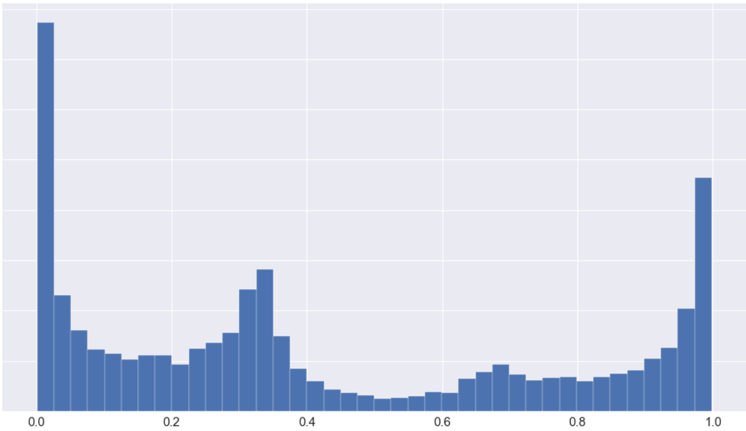

Figure 3: The in-fill asymptotic distribution with ε fixed for $\tau_1^0 = 0.33$, $\tau_2^0 = 0.67$, $\mu = 0$, $\delta_1 = 0$, $\delta_2 = 4$, $\delta_3 = 1$, and $\varepsilon = 1$

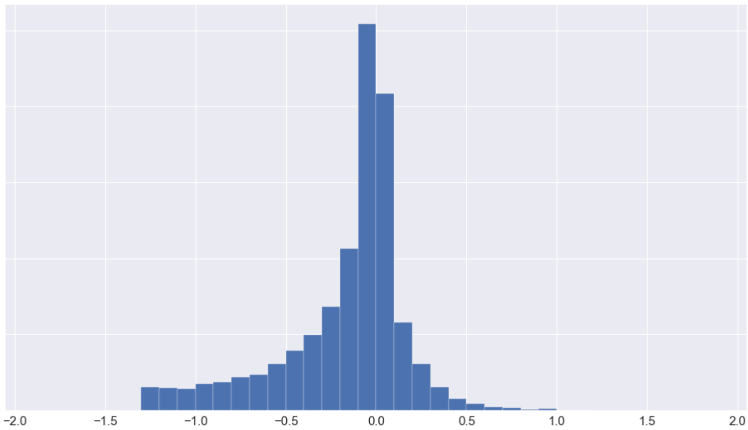

Figure 4: The in-fill asymptotic distribution with $\varepsilon \to \infty$ for $\tau_1^0 = 0.33$, $\tau_2^0 = 0.67$, $\mu = 0$, $\delta_1 = 0$, $\delta_2 = 4$, $\delta_3 = 1$, and $\varepsilon = h^{1/4}$

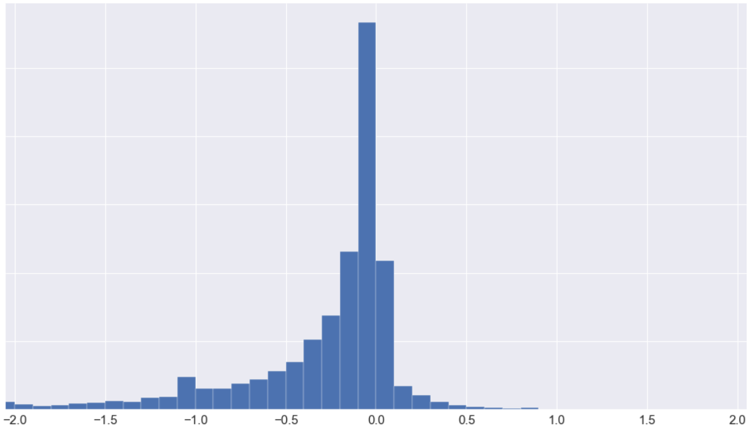

Figure 5: The finite sample distribution from GDP (3.1) for $\tau_1^0 = 0.33$, $\tau_2^0 = 0.67$, $\mu = 0$, $\delta_1 = 0$, $\delta_2 = 4$, $\delta_3 = 1$, $h = 1/T = 1/120$ and $\varepsilon = h^{1/4}$