In-fill Asymptotic Theory for Structural Break Point in Autoregression: A Unified Theory*

Liang Jiang
Singapore Management University

Jun Yu
Singapore Management University

November 10, 2016

Abstract

This paper obtains the exact distribution of the maximum likelihood estimator of the structural break point in the Ornstein–Uhlenbeck process when a continuous record is available. The exact distribution is asymmetric and tri-modal, regardless of the location of the true break point. These two properties are also found in the finite sample distribution of the least squares (LS) estimator of structural break point in autoregression (AR). The paper then develops an in-fill asymptotic theory for the LS estimator of the structural break point in AR. The in-fill asymptotic distribution is asymmetric and tri-modal and depends on the initial condition. It delivers good approximations to the finite sample distribution. Unlike the long-span asymptotic theory where the limiting distribution and sometimes even the rate of convergence depend on the underlying AR roots, the in-fill asymptotic theory is continuous in the underlying roots and, hence, offers a unified theory for making inference about the break point. Monte Carlo studies show that the in-fill asymptotic theory performs better than the existing tailor-made asymptotic theory in all cases considered.

JEL classification: C11; C46
Keywords: Structural break, AR, Exact distribution, In-fill asymptotics, Long-span asymptotics

*Corresponding author: Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University; email: yujun@smu.edu.sg. We acknowledge helpful comments from Peter C.B. Phillips, Xiaohu Wang, and Yichong Zhang.

1 Introduction

Autoregressive (AR) models with a structural break have been used extensively to describe economic time series; see for example Mankiw and Miron (1986), Mankiw, Miron and Weil (1987), Phillips, Wu, and Yu (2011) and Phillips and Yu (2011). The structural point is often linked to a significant economic event or an important economic policy. Not surprisingly, making statistical inference of the structural break point in the AR(1) model has received a great deal of attentions from both econometricians and empirical economists when they are confronted with economic time series. The asymptotic theory widely used in applications was developed by assuming that the numbers of the discrete time observations before and after the structural break point both go to infinite. The resulting long-span asymptotic distribution is discontinuous in the underlying AR(1) parameters. In particular, the asymptotic distribution and, sometimes even, the rate of convergence are different when the two underlying AR(1) parameters before and after the break point are in the range of being less than one, from that being equal to one, and further different from that being greater than one; see Chong (2001), Pang, Zhang, and Chong (2014) and Liang, et al (2014) for the development of these asymptotic distributions. Moreover, the long-span asymptotic distribution does not depend on the initial condition.

However, the finite sample distribution of the structural break point estimator is always continuous in the underlying AR parameters. That is, keeping one of the AR parameters fixed, changing the value of the other AR parameter by a small amount only leads to a small change in the finite sample distribution of the structural break point estimator. Furthermore, we expect the finite sample distribution of the structural break point estimator to depend on the initial condition. These two facts suggest that the long-span asymptotic distributions cannot always perform well in finite sample. Simulations that have been reported in the literature together with those that will be reported in Section 5 in the present paper strongly suggest that in many empirically relevant cases the long-span asymptotic distributions are inadequate.

The discontinuity in the long-span limiting distributions is also found in the AR(1) model without a structural break. This feature motivated Sims (1988) and Sims and Uhlig (1991) to use the Bayesian posterior distribution to make statistical inference about the AR parameter although Phillips (1991) showed that ignorance priors lead to the Bayesian posterior distributions which are much closer to the long-span limiting distributions. In a recent attempt, Phillips and Magdalinos (2007) developed the long-span limiting distributions for AR time series with a root which is moderately deviated from unity. They show that the rate of convergence in the new asymptotic theory provides a link between stationary and local-to-unit-root autoregressions. However, the limiting distribution itself remains discontinuous as the root passes through the unity.

Interestingly, when a continuous record of observations is available, continuous time models can provide the exact distribution of the "mean reversion" parameter, as shown in Phillips (1987a, 1987b). The exact distribution is continuous in the "mean reversion" parameter, regardless of its sign. This feature has motivated Phillips (1987a) and Perron (1991) to establish the in-fill asymptotic distribution for the AR(1) parameter in the discrete time model. It also motivates Yu (2014) and Zhou and Yu (2016) to establish the in-fill asymptotic distribution for the "mean reversion" parameter in the continuous time model. Not surprisingly, the in-fill asymptotic distribution inherits the property of continuity. Moreover, the in-fill asymptotic distribution depends explicitly on the initial condition.

In this paper, we develop an in-fill asymptotic distribution for the break point estimator in the AR(1) model with a break in the AR coefficient. The in-fill asymptotic distribution is continuous in the underlying AR parameters no matter what range of the AR parameters are in. Hence, it offers a unified framework for making statistical inference about the break point. It also depends explicitly on the initial condition. We make several contributions to the literature on structural breaks.

First, we show that when there is a continuous record of observations for the Ornstein–Uhlenbeck (OU) process with an unknown break point, we can derive the exact distribution of the maximum likelihood (ML) estimator of the break point via the Girsanov theorem. The exact distribution is continuous in the two "mean reversion" parameters, regardless of their signs and rates.

Second, we show that the exact distribution is always asymmetric about the true break point, regardless of the location of the true break point. Moreover, the distribution in general has three modes, one at the true value, two at the boundary points. The asymmetry and the trimodality have also been reported in Jiang, Wang and Yu (2016, JWY hereafter) in a model with a break in mean. However, our exact distribution remains asymmetric even when the break is in the middle of the sample. This feature is not shared by the exact distribution of JWY.

Third, motivated by the exact distributional theory, we derive the in-fill asymptotic distribution for the AR(1) model with a break, when the break point is the only unknown parameter. This AR(1) model with a break corresponds to the OU process with a break as the sampling interval shrinks. While our AR(1) model has the same model structure as those considered in the literature, we do not need to restrict the range of the AR coefficients to be less than one, or equal to one, or greater than one. That is, we allow the AR(1) model to switch from a stationary/unit root/explosive model to another stationary/unit root/explosive model. Hence, our in-fill asymptotic theory covers all the possible combinations of switches. Furthermore, our AR(1) model enables us to compare the break size and the magnitude of the initial condition with those assumed in the literature. The break size in our model has a smaller order of magnitude than those in the literature while the initial condition has a larger order than those in the literature. It is this smaller break size that allows us to develop a new and unified asymptotic theory. It is this larger initial condition that brings the prominence of the initial condition into the asymptotic distribution.

Fourth, we carry out extensive simulation studies, checking the performance of the in-fill asymptotic distribution against the long-span asymptotic distributions developed in the literature under different combinations of AR(1) coefficients. Our results show that the unified in-fill asymptotic distribution always performs better than the long-span asymptotic distributions although the later were developed to handle different kinds of regime shifts and hence were tailor-made.

The rest of the paper is organized as follows. Section 2 reviews the literature on AR models with a break. Special focus is paid on the assumptions about the AR coefficients before and after the break as well as the assumptions about the break size. Section 3 develops the exact distribution of the ML estimator of structural break point in the OU process model with a break. Section 4 develops the in-fill asymptotic theory for the LS estimator of the break point in the AR(1) model with a break. In Section 5, we provide simulation results and compare the finite sample performance of the in-fill theory with that of the long-span theory. Section 6 concludes. All proofs are contained in the Appendix.

2 A Literature Review and Motivations

The literature on estimating structural break point is too extensive to review. Among the contributions in the literature, Chong (2001), Pang, Zhang and Chong (2014) and Liang et al (2014) focused on the AR(1) model with a break. Under various assumptions on the AR(1) coefficients, the long-span asymptotic theory has been developed in these papers for the least squares (LS) estimator of the structural break point.

The model considered in these papers is

$$y_t = \begin{cases} \beta_1 y_{t-1} + \epsilon_t & \text{if } t \leq k_0 \\ \beta_2 y_{t-1} + \epsilon_t & \text{if } t > k_0 \end{cases}, \quad t = 1, 2, \ldots, T, \tag{1}$$

where T denotes the sample size, ϵ_t is a sequence of independent and identically (i.i.d.) random variables. Let k denote the break point parameter with true value k_0. The condition $1 \leq k_0 < T$ is assumed to ensure that one and only one break happens. The fractional break point parameter is defined as $\tau = k/T$ with true value $\tau_0 = k_0/T$. The break size is captured by $\beta_2 - \beta_1$. The order of the initial condition y_0 will be assumed later.

The LS estimator of k takes the form of

$$\hat{k}_{LS,T} = \arg \min_{k=1,\ldots,T-1} \left\{ S_k^2 \right\}, \tag{2}$$

where

$$S_k^2 = \sum_{t=1}^{k} \left(y_t - \hat{\beta}_1(k) y_{t-1} \right)^2 + \sum_{t=k+1}^{T} \left(y_t - \hat{\beta}_2(k) y_{t-1} \right)^2,$$

with $\hat{\beta}_1(k) = \sum_{t=1}^{k} y_t y_{t-1} / \sum_{t=1}^{k} y_{t-1}^2$ and $\hat{\beta}_2(k) = \sum_{t=k+1}^{T} y_t y_{t-1} / \sum_{t=k+1}^{T} y_{t-1}^2$ being the LS estimators of parameters β_1 and β_2 for any given k. The corresponding estimator of τ is $\hat{\tau}_{LS,T} = \hat{k}_{LS,T}/T$.

Let the break size depend on T (so either $\beta_1 = \beta_{1T}$ or $\beta_2 = \beta_{2T}$ or both). Under various settings on β_1 and β_2 (e.g. β_1, β_2 are smaller than 1, equal to 1 or greater than 1), Chong (2001), Pang, Zhang and Chong (2014) and Liang, et al (2014,) established the consistency of $\hat{\tau}_{LS,T}$ and derived its long-span asymptotic distributions under different shrinking rate of $\beta_2 - \beta_1$ as $T \to \infty$. In the following, we review the main results on the asymptotic distributions of the break point estimator in the literature.

2.1 The long-span asymptotics when $|\beta_1| < 1$ and $|\beta_2| < 1$

Chong (2001) studied the model in (1) with $|\beta_1| < 1$ and $|\beta_2| < 1$. In this case, the AR(1) model switches from a stationary root to another stationary root. Under the regularity conditions that $\epsilon_t \overset{i.i.d.}{\sim} (0, \sigma^2)$, $E(\epsilon_t^4) < \infty$, $E(y_0) = 0$, and $E(y_0^2) < \infty$, assuming $\beta_{2T} - \beta_1 \to 0$ with $\sqrt{T} \left| \beta_{2T} - \beta_1 \right| \to \infty$ as $T \to \infty$, he derived an asymptotic distribution of $\hat{\tau}_{LS,T}$ as

$$\frac{T(\beta_{2T} - \beta_1)^2}{1 - \beta_1^2} (\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg \max_{u \in (-\infty, \infty)} \left\{ W(u) - \frac{1}{2} |u| \right\},$$

where $W(u)$ is a two-sided Brownian motion, defined as $W(u) = W_1(-u)$ if $u \leq 0$ and $W(u) = W_2(u)$ if $u > 0$, with W_1 and W_2 being two independent Brownian motions. The probability density function (pdf) and the cumulative distribution function for this limiting distribution were obtained in Yao (1987).

2.2 The long-span asymptotics when $|\beta_1| < 1$ and $\beta_2 = 1$

Chong (2001) studied the model in (1) with $|\beta_1| < 1$ and $\beta_2 = 1$. In this case, the AR(1) model switches from a stationary root to a unit root. Under the same regularity conditions as in Section 2.1, assuming $1 - \beta_{1T} \to 0$ with $T(1 - \beta_{1T}) \to \infty$ as $T \to \infty$, he derived an asymptotic distribution of $\hat{\tau}_{LS,T}$ as

$$T(1 - \beta_{1T})(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg \max_{u \in (-\infty, \infty)} \left\{ \frac{W_a^*(u)}{R_1} - \frac{1}{2} |u| \right\},$$

where $W_a^*(u) = W_1(-u)$ if $u \leq 0$ and

$$W_a^*(u) = -W_2(u) - \int_0^u \frac{W_2(s)}{R_1} dW_2(s) - \int_0^u \left(\frac{W_2(s)}{2R_1} + 1 \right) W_2(s) ds,$$

if $u > 0$ with $W_1(\cdot)$ and $W_2(\cdot)$ being two independent Brownian motions and $R_1 = \int_0^\infty \exp(-s) dW_1(s)$,

2.3 The long-span asymptotics when $\beta_1 = 1$ and $|\beta_2| < 1$

Chong (2001) studied the model in (1) with $\beta_1 = 1$ and $|\beta_2| < 1$. In this case the AR(1) model switches from a unit root to a stationary root. Under the same regularity conditions as in Section 2.1, assuming $\sqrt{T}(1 - \beta_{2T}) \to 0$ with $T^{3/4}(1 - \beta_{2T}) \to \infty$ as $T \to \infty$, he derived an asymptotic distribution of $\hat{\tau}_{LS,T}$ as

$$T^2(\beta_{2T} - 1)^2(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg \max_{u \in (-\infty, \infty)} \left\{ \frac{W(u)}{W_3(\tau_0)} - \frac{1}{2} |u| \right\},$$

where $W(u)$ is a two-sided Brownian motion and W_3 is an independent standard Brownian motion.

2.4 The long-span asymptotics when $|\beta_1| < 1$ and $\beta_{2T} = 1 - c/T$

Pang, et al (2014) studied the model in (1) with $|\beta_1| < 1$ and $\beta_{2T} = 1 - c/T$ where c being a fixed constant. In this case an AR(1) model switches from a stationary root to a local-to-unit-root. Under the regularity conditions that $y_0 = o_p(\sqrt{T})$, assuming $|\beta_{2T} - \beta_{1T}| \to 0$ with $T(\beta_{2T} - \beta_{1T}) \to \infty$, they derived an asymptotic distribution of $\hat{\tau}_{LS,T}$ as

$$T(\beta_{2T} - \beta_1)(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg \max_{u \in (-\infty, \infty)} \left\{ \frac{W_b^*(u)}{R_1} - \frac{1}{2} |u| \right\},$$

where $W_b^*(u) = W_1(-u)$ if $u \leq 0$ and

$$W_b^*(u) = -I(W_2, c, \tau_0, u) - \int_0^u \frac{I(W_2, c, \tau_0, s)}{R_1} dI(W_2, c, \tau_0, s) - \int_0^u \left(\frac{I(W_2, c, \tau_0, s)}{2R_1} + 1 \right) I(W_2, c, \tau_0, s) ds,$$

if $u > 0$ with

$$I(W_2, c, \tau_0, s) = W_2(\tau_0 + s) - W_2(\tau_0) - c \int_{\tau_0}^{\tau_0+s} e^{-c(\tau_0+s-r)} \left(W_2(r) - W_2(\tau_0)\right) ds,$$

and W_1 and W_2 being two independent Brownian motions and $R_1 = \int_0^\infty \exp(-s) dW_1(s)$.

2.5 The long-span asymptotics when $\beta_{1T} = 1 - c/T$ and $|\beta_2| < 1$

Pang, et al (2014) also studied the model in (1) with $\beta_{1T} = 1 - c/T$ and $\beta_2 = 1$ where c being a fixed constant. In this case an AR(1) model switches from a local-to-unit-root to a stationary root. Under the regularity conditions as in Section 2.4, assuming $\sqrt{T}(\beta_{2T} - \beta_{1T}) \to 0$ with $T^{3/4}(\beta_{2T} - \beta_{1T}) \to \infty$, they derived $\hat{\tau}_{LS,T}$ an asymptotic distribution as

$$T^2(\beta_2 - \beta_{1T})^2(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg \max_{u \in (-\infty, \infty)} \left\{ \frac{W(u)}{\exp\left(c(1-\tau_0)\right) G(W_1, c, \tau_0)} - \frac{1}{2}|u| \right\},$$

where $W(u)$ is a two-sided Brownian motion defined in Section 2.1, and

$$G(W_1, c, \tau_0) = \exp\left(-c(1-\tau_0)\right) W_1(\tau_0) - c \int_0^{\tau_0} \exp\left(-c(1-s)\right) W_1(s) ds.$$

2.6 The long-span asymptotics when $\beta_{1T} = 1 - c/k_T$ and $\beta_2 = 1$ with $k_T/T \to 0$

Liang, et al (2014) studied the model in (1) with $\beta_{1T} = 1 - c/k_T$ and $\beta_2 = 1$ with c being a fixed positive constant. In this case an AR(1) model switches from a mildly stationary one to a unit root. Under the regularity conditions that $y_0 = o_p(\sqrt{k_T})$, assuming $k_T \to \infty$ and $k_T/T \to 0$ as $T \to \infty$,[1] they derived an asymptotic distribution of $\hat{\tau}_{LS,T}$ as

$$\frac{cT}{k_T}(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg \max_{u \in (-\infty, \infty)} \left\{ \frac{W_c^*(u)}{R_c} - \frac{1}{2}|u| \right\},$$

where $W_c^*(u) = W_1(-u)$ if $u \leq 0$ and

[1] It may be easier to understand $k_T = T^\alpha$, with $\alpha \in (0, 1)$.

Table 1: The table shows the long-span asymptotic distributions of $\hat{\tau}_{LS,T}$ under different model settings on the AR(1) coefficients before and after the break point.

β_1	β_2	$\|\beta_2-\beta_1\|$	y_0	rate	long-span asymptotic
$\|\beta_1\|<1$	$\|\beta_2\|<1$	$(T^{-0.5},T^{-\varepsilon})$	$O_p(1)$	$\frac{T(\beta_2-\beta_1)^2}{1-\beta_1^2}$	$\arg\max_{u\in(-\infty,\infty)}\left\{W(u)-\frac{1}{2}\|u\|\right\}$
$\|\beta_1\|<1$	1	$(T^{-1},T^{-\varepsilon})$	$O_p(1)$	$T(1-\beta_1)$	$\arg\max_{u\in(-\infty,\infty)}\left\{\frac{W_a^*(u)}{R_1}-\frac{1}{2}\|u\|\right\}$
1	$\|\beta_2\|<1$	$\left(T^{-0.75},T^{-0.5}\right)$	$O_p(1)$	$T^2(\beta_2-1)^2$	$\arg\max_{u\in(-\infty,\infty)}\left\{\frac{W(u)}{W_3(\tau_0)}-\frac{1}{2}\|u\|\right\}$
$\|\beta_1\|<1$	$1-c/T$	$(T^{-1},T^{-\varepsilon})$	$o_p(\sqrt{T})$	$T(\beta_{2T}-\beta_1)$	$\arg\max_{u\in(-\infty,\infty)}\left\{\frac{W_b^*(u)}{R_1}-\frac{1}{2}\|u\|\right\}$
$1-c/T$	$\|\beta_2\|<1$	$\left(T^{-0.75},T^{-0.5}\right)$	$o_p(\sqrt{T})$	$T^2(\beta_2-\beta_{1T})^2$	$\arg\max_{u\in(-\infty,\infty)}\left\{\frac{e^{-c(1-\tau_0)}W(u)}{G(W_1,c,\tau_0)}-\frac{1}{2}\|u\|\right\}$
$1-c/k_T$	1	$(T^{-1},T^{-\varepsilon})$	$o_p(T^{\alpha/2})$	$\frac{cT}{k_T}$	$\arg\max_{u\in(-\infty,\infty)}\left\{\frac{W_c^*(u)}{R_c}-\frac{1}{2}\|u\|\right\}$
1	$1-c/k_T$	$\left(T^{-0.75},T^{-0.5}\right)$	$o_p(T^{\alpha/2})$	$\frac{c^2T^2}{k_T^2}$	$\arg\max_{u\in(-\infty,\infty)}\left\{\frac{W(u)}{W_1(\tau_0)}-\frac{1}{2}\|u\|\right\}$

$$W_c^*(u) = -W_2(u) - \int_0^u \frac{W_2(s)}{R_c}dW_2(s) - \int_0^u \left(\frac{W_2(s)}{2R_c}+1\right)W_2(s)ds,$$

if $u>0$ with W_1 and W_2 being two independent Brownian motions and $R_c=\sqrt{c}\int_0^\infty \exp(-cs)dW_1(s)$.

2.7 The long-span asymptotics when $\beta_1 = 1$ and $\beta_{2T} = 1 - c/k_T$ with $k_T/T \to 0$

Liang, et al (2014) also studied the model in (1) with $\beta_1 = 1$ and $\beta_{2T} = 1 - c/k_T$, with c being a fixed positive constant. In this case an AR(1) model switches from a unit root to a mildly stationary root. Under the regularity conditions as in section 2.7, assuming $k_T \to \infty$ and $\sqrt{T}/k_T \to 0$ with $T^{3/4}/k_T \to \infty$ as $T \to \infty$, they derived an asymptotic distribution of $\hat{\tau}_{LS,T}$ as

$$\frac{c^2T^2}{k_T^2}(\hat{\tau}_{LS,T}-\tau_0) \xrightarrow{d} \arg\max_{u\in(-\infty,\infty)}\left\{\frac{W(u)}{W_1(\tau_0)}-\frac{1}{2}|u|\right\},$$

where $W(u)$ is a two-sided Brownian motion.

We summarize all the long-span asymptotic theory derived in the literature in Table 1 where we report the range of the two AR coefficients, the order of the break size, the order of the initial condition, the rate of convergence of the LS break point estimator, and the long-span asymptotic distribution. Both the break size and the initial condition are expressed in the power order to facilitate the comparison and discussion, where ε is an arbitrarily small postive number.

Several observations can be made from Table 1. First, the long-span asymptotic distribution is discontinuous in the underlying AR(1) parameters. The asymptotic distributions and, sometimes even, the rates of convergence are different when the AR(1) parameters are in different ranges. This feature causes a great deal of difficulties in making statistical inference about the break point in practice because users typically do not know *ex ante* whether the underlying AR(1) model has a root that is stationary/unity/explosive before and after the break point. Moreover, even if users know the range of the underlying AR(1) parameters before and after the break point, it is not easy to decide which long-span asymptotic distribution to use. For example, if we know the AR(1) parameter changes from 0.9 to 1, should we use the model that switches from a stationary root to a unit root, or from a mildly stationary root to a unit root? For another example, if we know the underlying AR(1) parameter changes from 0.9 to 0.95, should we use the model that switches from a stationary root to another stationary root, or from a stationary root to a local-to-unit-root? This choice is important as a different choice leads to a different long-span asymptotic distribution although the finite sample distribution is the same.

Second, none of the long-span asymptotic distributions depends on the initial condition. The initial condition is assumed to be either $O_p(1)$ or $o_p(\sqrt{T})$ or $o_p(T^{\alpha/2})$ for $\alpha \in (0,1)$. These initial conditions are imposed so that they disappeared asymptotically. It is always smaller than $O_p(\sqrt{T})$.

Third, while the order of the break size varies across different models, it is always larger than $O(T^{-1})$. The order of the break size shown in Table 1 is critical to deriving the consistency and corresponding limiting distribution of $\hat{\tau}_{LS,T}$ in each case.

Fourth, the interval to find the argmax in all cases is always $(-\infty, \infty)$. It is computationally expensive to obtain the limiting distribution except for the AR model that switches from a stationary root to another stationary root, for which the closed-form expression for the pdf was given by Yao (1987). To obtain the long-run asymptotic distributions in all other cases, we need to numerically obtain the argmax over an interval that is sufficiently wide. Since the grid must be fine enough to well approximate the true argmax, the number of grid points is inevitably very large, leading to a high computational cost.

Despite that these long-span asymptotic distributions are tailor-developed, catering for different persistency in AR, they do not perform well in many empirically relevant cases. To see this problem, we plot the density of the limiting distribution of $\hat{\tau}_{LS,T}$ obtained from 200 observations with $\tau_0 = 0.3$, when the AR coefficient switches from a stationary root to another stationary root in Figure 1, when the AR coefficient switches from a stationary root to the unit root in Figure 2, when the AR coefficient switches from a local-unit-root to a stationary root in Figure 3. In the same graphs we plot the

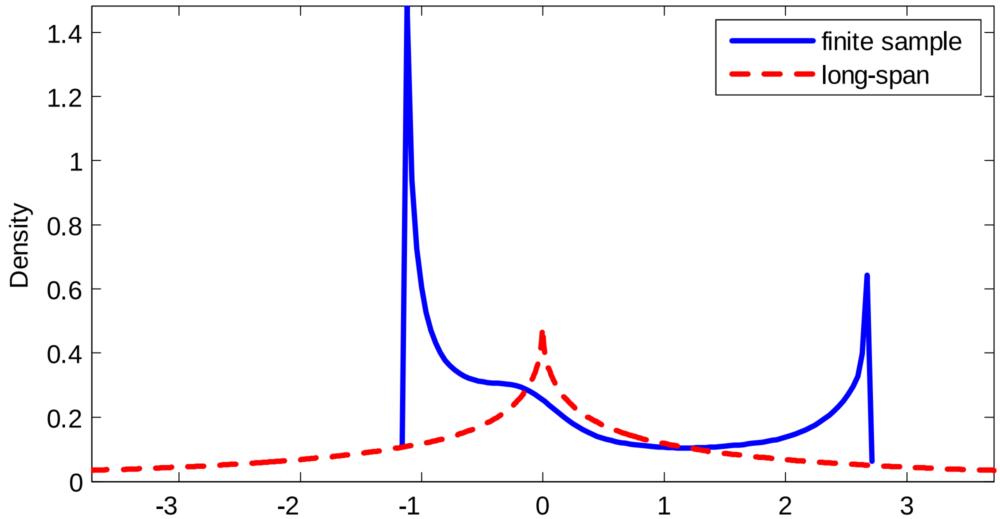

Figure 1: The pdf of the finite sample distribution of $\frac{T(\beta_2-\beta_1)^2}{1-\beta_1^2}(\hat{\tau}_{LS,T}-\tau_0)$ when $T = 200$, $\beta_1 = 0.5, \beta_2 = 0.38$, $\sigma = 1$ and $\tau_0 = 0.3$ in Model (1) and the pdf of $\arg\max_{u\in(-\infty,\infty)} \left\{W(u) - \frac{1}{2}|u|\right\}$.

corresponding finite sample distributions of $\hat{\tau}_{LS,T}$.[2] The finite sample distributions are obtained from simulated data with 100,000 replications. It is clear that the long-span asymptotic distributions are distinctively different from the finite sample distributions in all cases. First, the support of the finite sample distributions is bounded, whereas the support of the long-span asymptotic distributions is infinite. Second, the finite sample distributions are asymmetric in all cases, whereas the long-span asymptotic distributions are symmetric in some cases. Third, the finite sample distributions display trimodality while the long-span asymptotic distributions has a unique mode. The big discrepancy between the two densities suggests that the long-span asymptotic theory is inadequate in many practically relevant cases and that there is a need to develop the exact distribution theory or an alternative asymptotic theory to better approximate the finite sample distribution for the break point.

[2] The complete comparisions with different combination of β_1 and β_2, covering all the cases in Table 1, are in Section 5.

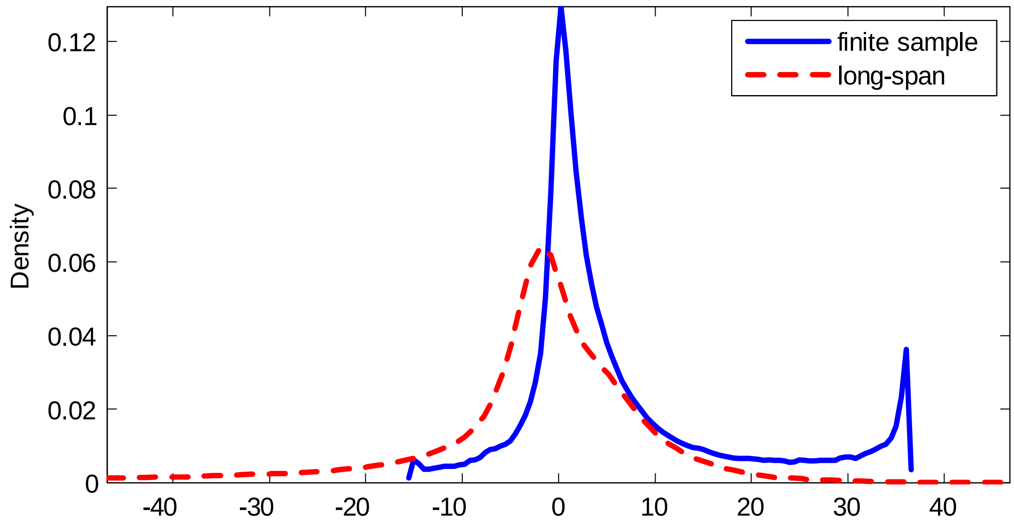

Figure 2: The pdf of the finite sample distribution of $T(\beta_2 - \beta_1)(\hat{\tau}_{LS,T} - \tau_0)$ when $T = 200$, $\beta_1 = 0.73, \beta_2 = 1$, $\sigma = 1$ and $\tau_0 = 0.3$ in Model (1) and the pdf of $\arg\max_{u \in (-\infty,\infty)} \left\{ \frac{W_a^*(u)}{R_1} - \frac{1}{2}|u| \right\}$.

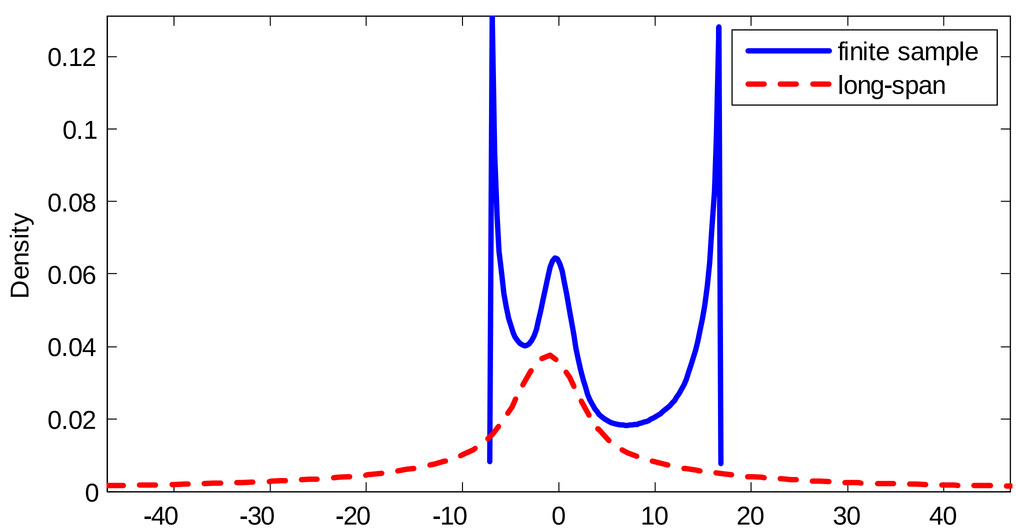

Figure 3: The pdf of the finite sample distribution of $T(\beta_2 - \beta_1)(\hat{\tau}_{LS,T} - \tau_0)$ when $T = 200$, $\beta_1 = 0.995, \beta_2 = 0.97$, $\sigma = 1$ and $\tau_0 = 0.3$ in Model (1) and the pdf of $\arg\max_{u \in (-\infty,\infty)} \left\{ \frac{W(u)}{\exp(c(1-\tau_0))G(W_1,c,\tau_0)} - \frac{1}{2}|u| \right\}$

3 A Continuous Time Model and Exact Distribution

In this section we consider the following Ornstein-Uhlenbeck (OU) process over a finite time interval $[0,1]$,[3]

$$dy(t) = -\left(\kappa + \delta 1_{[t>\tau_0]}\right) y(t)dt + \sigma dB(t), \tag{3}$$

where $t \in [0,1]$, $y(0)$ is the initial condition, $1_{[t>\tau_0]}$ is an indicator function, κ, δ and τ_0 are constants with τ_0 being the break point and δ being the break size, σ is another constant capturing the noise level, and $B(t)$ denotes a standard Brownian motion. Clearly δ/σ is the signal-to-noise ratio. We impose the assumption $\tau_0 \in [\alpha, \beta]$ with $0 < \alpha < \beta < 1$ so that a break indeed occurs.

We assume that a continuous record of observations, $\{y(t)\}$ for $t \in [0,1]$, is available and that all parameters are known except for τ_0. Once a continuous record is available, one can make more complicated assumption about the diffusion function, such as $\sigma = \sigma(y(t))$ without changing the analysis developed below. This is because the diffusion function can be estimated by the quadratic variation $[y]_t$ without estimation error for all t.

3.1 $\widetilde{J}_{\kappa,\delta,\tau}(r)$ process

The $y(t)$ process defined by Model (3) is closely related to the following Gaussian process,

$$\widetilde{J}_{\kappa,\delta,\tau}(r) = \begin{cases} \int_0^r e^{-(r-s)\kappa} dB(s) & \text{if } r \le \tau \\ \int_0^\tau e^{-(r-s)(\kappa+\delta)+(\tau-s)\delta} dB(s) + \int_\tau^r e^{-(r-s)(\kappa+\delta)} dB(s) & \text{if } r > \tau, \end{cases}, \tag{4}$$

which plays a central role in our theory.

The following lemma gives the distributional properties of $\widetilde{J}_{\kappa,\delta,\tau}(r)$.

Lemma 3.1 *Let $\widetilde{J}_{\kappa,\delta,\tau}(r)$ be a process generated by (4), then*

(a) for fixed $r > 0$, $\widetilde{J}_{\kappa,\delta,\tau}(r)$ has the distribution

$$\widetilde{J}_{\kappa,\delta,\tau}(r) \overset{d}{=} N(0, \sigma_J^2(r)),$$

with

$$\sigma_J^2(r) = \begin{cases} \frac{1-e^{-2r\kappa}}{2\kappa} & \text{if } r \le \tau \\ \frac{\delta e^{-2(r-\tau)(\kappa+\delta)}}{2\kappa(\kappa+\delta)} - \frac{e^{-2\kappa-2(r-\tau)\delta}}{2\kappa} + \frac{1}{2(\kappa+\delta)} & \text{if } r > \tau. \end{cases}$$

[3] A different length of time interval, such as $[0, N]$, may be assumed without qualitatively changing the results derived in the present paper.

where $\overset{d}{=}$ *denotes equivalence in distribution.*

(b) $\widetilde{J}_{\kappa,\delta,\tau}(r)$ *is a process generated by the stochastic differential equation (SDE)*

$$d\widetilde{J}_{\kappa,\delta,\tau}(r) = -(\kappa + \delta 1_{[r>\tau]})\widetilde{J}_{\kappa,\delta,\tau}(r)dr + dB(r), \tag{5}$$

with initial condition $\widetilde{J}_{\kappa,\delta,\tau}(0) = 0$.

Remark 3.1 $\widetilde{J}_{\kappa,\delta,\tau}(r)$ *is related to the following OU process widely used in the literature (for example, see Phillips, 1987b):*

$$dJ_c(r) = cJ_c dr + dB(r),$$

with the initial condition $J_c(0) = 0$. *When* $r \leq \tau$, $\widetilde{J}_{\kappa,\delta,\tau} = J_{-\kappa}(r)$. *When* $r > \tau$, $\widetilde{J}_{\kappa,\delta,\tau} = \exp(-(r-\tau)(\kappa+\delta))J_{-\kappa}(\tau) + J_{-(\kappa+\delta)}(r) - J_{-(\kappa+\delta)}(\tau)$, *which follows from the proof of Lemma 3.1 and the fact* $J_c(r) = \int_0^r e^{(r-s)c} dB(s)$.

Remark 3.2 *It is clear that an alternative representation for* $\widetilde{J}_{\kappa,\delta,\tau}(r)$ *in (4) is*

$$\widetilde{J}_{\kappa,\delta,\tau}(r) = \begin{cases} \int_0^r e^{-(r-s)\kappa} dB(s) & \text{if } r \leq \tau \\ e^{-(r-\tau)(\kappa+\delta)}\widetilde{J}_{\kappa,\delta,\tau}(\tau) + \int_\tau^r e^{-(r-s)(\kappa+\delta)} dB(s) & \text{if } r > \tau \end{cases}. \tag{6}$$

The following lemma establishes the connection between $y(t)$ in Model (3) with $\widetilde{J}_{\kappa,\delta,\tau_0}(t)$.

Lemma 3.2 *If* $y(t)$ *is a process generated by (3), we have*

$$y(t) = \sigma \widetilde{J}_{\kappa,\delta,\tau_0}(t) + \exp\left(-t\kappa - (t-\tau_0)\delta 1_{[t>\tau_0]}\right) y(0),$$

where $\widetilde{J}_{\kappa,\delta,\tau}(t)$ *is defined in (4) over the finite time interval* $[0,1]$.

The following lemma gives us a useful relationship for the development of our distribution theory.

Lemma 3.3 *If* $y(t)$ *is a process generated by (3), we have*

$$2\sigma \int_0^r y(s)dB(s) = y^2(r) - y^2(0) + 2\int_0^r (\kappa + \delta 1_{[s>\tau_0]})y^2(s)ds - r\sigma^2.$$

Remark 3.3 *By directly applying the result in Lemma 3.2, we can write Lemma 3.3 in terms of* $\widetilde{J}_{\kappa,\delta,\tau_0}(t)$.

3.2 The exact distribution

For any $\tau \in (0,1)$, we can obtain the exact log-likelihood of Model (3) via the Girsanov Theorem as

$$\log \mathcal{L}(\tau) = \log \frac{dP_\tau}{dP_B} = \frac{1}{\sigma^2}\left\{ \int_0^1 \left(-(\kappa + \delta 1_{[t>\tau]})\right) y(t)dy(t) - \frac{1}{2}\int_0^1 \left(\kappa + \delta 1_{[t>\tau]}\right)^2 y^2(t)dt \right\},$$

where P_τ is the probability measure corresponding to Model (3) with τ_0 replaced by τ for any $\tau \in (0,1)$ and P_B is the probability measure corresponding to $B(t)$. This leads to the ML estimator of τ_0 as

$$\hat{\tau}_{ML} = \arg \max_{\tau \in (0,1)} \log \mathcal{L}(\tau). \tag{7}$$

and Theorem 3.1 reports the exact distribution of $\hat{\tau}_{ML}$.

Theorem 3.1 *Consider Model (3) with a continuous record being available. For the ML estimator $\hat{\tau}_{ML}$ defined in (7), we have the exact distribution as*

$$\hat{\tau}_{ML} \overset{d}{=} \arg \max_{\tau \in (0,1)} \left\{ \left(\frac{\delta}{\sigma}\right) \int_{\tau_0}^{\tau} y(t)dB(t) - \frac{1}{2}\left| \int_{\tau_0}^{\tau} \left(\frac{\delta}{\sigma}\right)^2 y^2(t)dt \right| \right\}. \tag{8}$$

where $y(t)$ is the process defined in (3) and we use the notation $\int_{\tau_0}^{\tau}(\cdot)dB(s)$ to denote $\left[\int_0^{\tau}(\cdot)dB(s) - \int_0^{\tau_0}(\cdot)dB(s)\right]$.

Remark 3.4 *Since $y(t)$ is continuous in κ, the exact distribution given in (8) is continuous in κ. Moreover, it is continuous in δ.*

Remark 3.5 *One may write the exact distribution as the argmax of two pieces as follow:*

$$\hat{\tau}_{ML} \overset{d}{=} \arg \max_{\tau \in (0,1)} \begin{cases} -\int_{\tau}^{\tau_0} \frac{y(r)}{\delta\sigma} dB(r) - \frac{1}{2}\int_{\tau}^{\tau_0} \left(\frac{y(r)}{\sigma}\right)^2 dr & \text{for } \tau \leq \tau_0 \\ \int_{\tau_0}^{\tau} \frac{y(r)}{\delta\sigma} dB(r) - \frac{1}{2}\int_{\tau_0}^{\tau} \left(\frac{y(r)}{\sigma}\right)^2 dr & \text{for } \tau > \tau_0 \end{cases}.$$

Since neither $y(r)$ nor $y^2(r)$ is symmetric about τ_0 over the interval $(0,1)$ even when $\tau_0 = 50\%$, the distribution of $\hat{\tau}_{ML}$ is asymmetric for all τ_0, including $\tau_0 = 50\%$.

Theorem 3.1 gives the exact distribution of $\hat{\tau}_{ML}$ when a continuous record over a finite time span is available. The following Lemma 3.2 gives an alternative expression for the exact distribution using $\widetilde{J}_{\kappa,\delta,\tau}(t)$ defined in (4).

Corollary 3.1 *Consider Model (3) with a continuous record being available. For the ML estimator $\hat{\tau}_{ML}$ defined in (7), we have the exact distribution as*

$$\hat{\tau}_{ML} \stackrel{d}{=} \arg\max_{\tau\in(0,1)} \left\{ \int_{\tau_0}^{\tau} \frac{1}{\delta} \left(\widetilde{J}_{\kappa,\delta,\tau_0}(s) + e^{-s\kappa-(s-\tau_0)\delta 1_{[s>\tau_0]}} \frac{y(0)}{\sigma} \right) dB(s) \right. \tag{9}$$
$$\left. -\frac{1}{2} \left| \int_{\tau_0}^{\tau} \left(\widetilde{J}_{\kappa,\delta,\tau_0}(s) + e^{-s\kappa-(s-\tau_0)\delta 1_{[s>\tau_0]}} \frac{y(0)}{\sigma} \right)^2 ds \right| \right\},$$

where $B(s)$ is the standard Brownian motion corresponding to $\widetilde{J}_{\kappa,\delta,\tau_0}(s)$ and we use the notation $\int_{\tau_0}^{\tau}(\cdot)dB(s)$ to denote $\left[\int_0^{\tau}(\cdot)dB(s) - \int_0^{\tau_0}(\cdot)dB(s)\right]$.

Remark 3.6 *An advantage of using this alternative expression is that it depends on the Gaussian process $\widetilde{J}_{\kappa,\delta,\tau_0}$. explicitly. Another advantage is that the distribution depends explicitly on the initial condition $y(0)$, or more precisely on $y(0)/\sigma$.*

The exact distribution in (8) and (9) corresponds to the extreme value of stochastic integrals over a finite interval. Unfortunately, we cannot obtain the pdf or cdf in closed-form. As a result, we obtain the pdf by simulated data with 100,000 replications as in JWY (2016).

Figures 4-8 plot the densities of $\widehat{\tau}_{ML} - \tau_0$ given in Equation (8) with different values of κ and δ, when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and σ is 1. These cover several interesting cases of switches, from a stationary OU process to another stationary OU process, from a stationary OU process to a nearly nonstationary OU process, from a nearly nonstationary OU process to a stationary OU process, from a stationary OU process to a nonstationary OU process, and from a nonstationary OU process to a stationary OU process.

There are several interesting observations from these plots. First, the density is always asymmetric, even when $\tau_0 = 50\%$. This feature is different from that in the estimation of break point in the mean derived in JWY (2016), where the density of $\widehat{\tau}_{ML} - \tau_0$ is symmetric about zero when $\tau_0 = 50\%$. Second, there are trimodality in the finite sample distribution. The true value is one of the three modes while the two boundary points are the other two modes. The highest mode may not even be the true value τ_0. This feature is shared by the density derived in JWY. Third, when $\tau_0 = 0.3$, the distribution of $\widehat{\tau}_{ML} - \tau_0$ is not a mirror image of that when $\tau_0 = 0.7$. This feature is different from that of JWY.

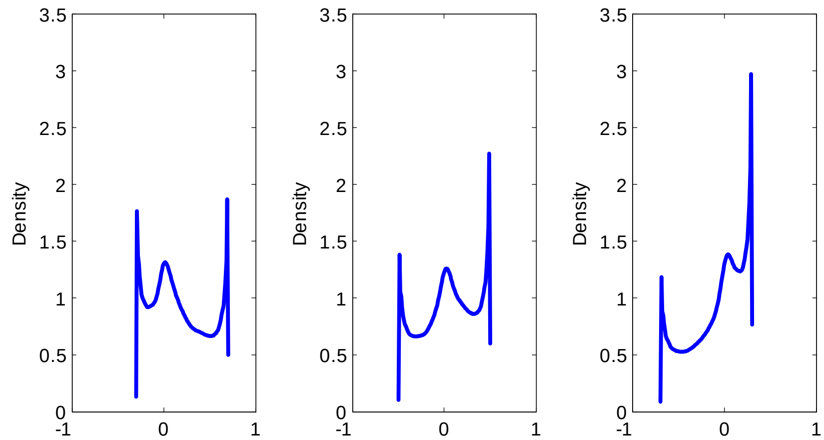

Figure 4: The density of $\widehat{\tau}_{ML} - \tau_0$ given in Equation (8) when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\kappa = 138, \delta = -20$, $\sigma = 1$.

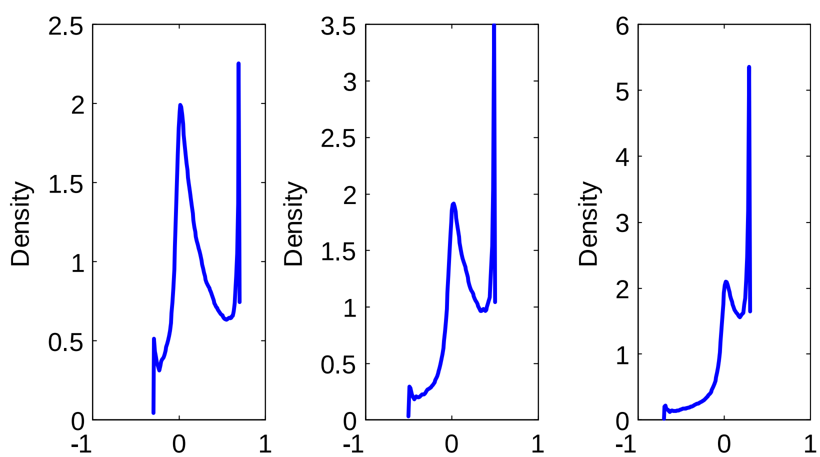

Figure 5: The density of $\widehat{\tau}_{ML} - \tau_0$ given in Equation (8) when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\kappa = 10, \delta = -9$, $\sigma = 1$.

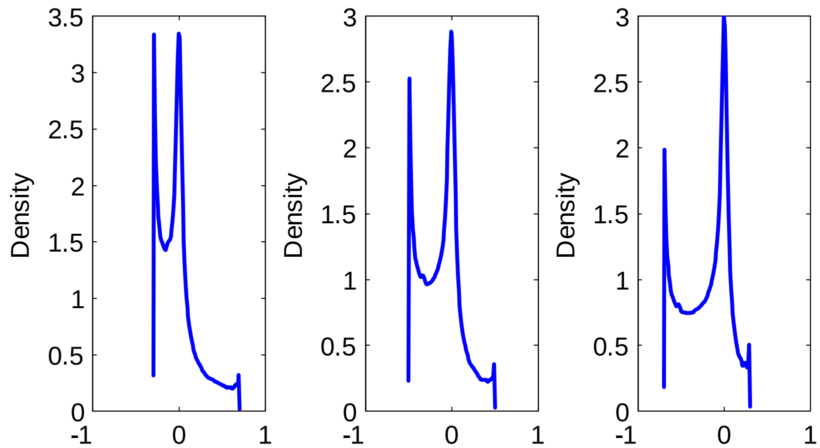

Figure 6: The density of $\widehat{\tau}_{ML} - \tau_0$ given in Equation (8) when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\kappa = 1, \delta = 5, \sigma = 1$.

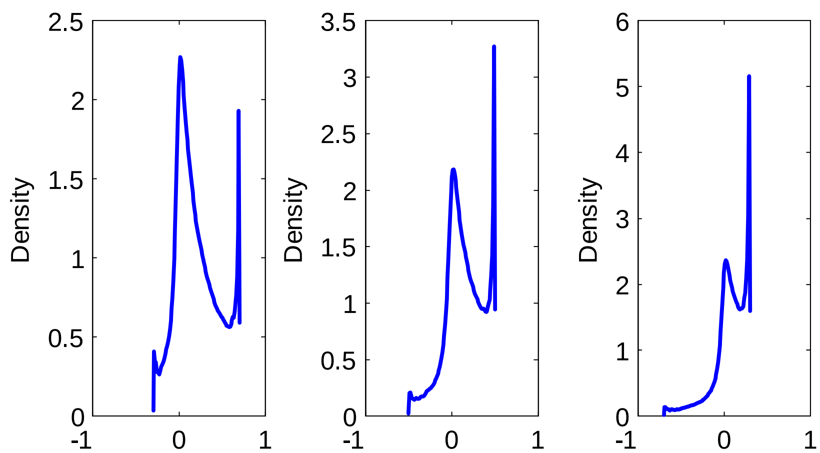

Figure 7: The density of $\widehat{\tau}_{ML} - \tau_0$ given in Equation (8) when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\kappa = 10, \delta = -10, \sigma = 1$.

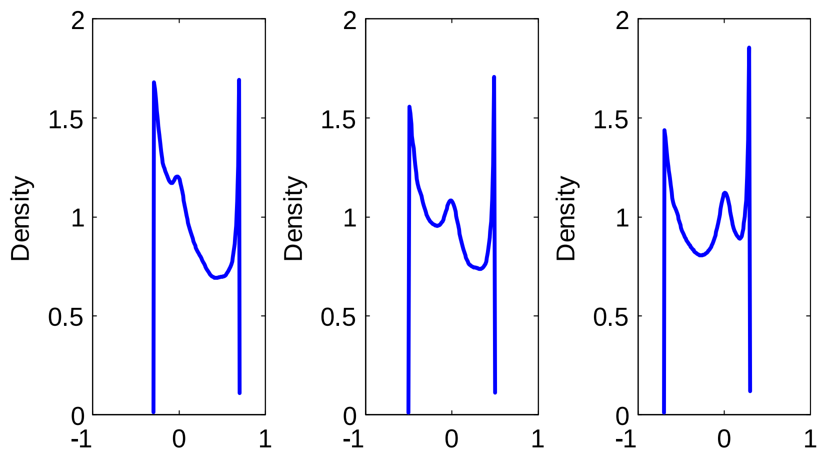

Figure 8: The density of $\widehat{\tau}_{ML}-\tau_0$ given in Equation (8) when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\kappa = 0, \delta = 1$, $\sigma = 1, y(0) = 0$.

4 A Discrete Time Model and In-fill Asymptotic Distribution

Applying the Euler discretization scheme to Model (3) at the equi-spaced intervals over the time span $[0, 1]$, we get the following discrete time model:

$$y_{th} - y_{(t-1)h} = -\left(\kappa + \delta 1_{[t>k_0]}\right) y_{(t-1)h}h + \sigma\sqrt{h}\varepsilon_{th},\ \varepsilon_{th} \sim N(0,1),\ y_0 = O_p(1), \tag{10}$$

where h is the sampling interval, $t = 1, ..., T$, with $T = 1/h$. For simplicity, we assume τ_0/h to be an integer, denoted by k_0. As $h \to 0$, Model (10) converges to Model (3) and $T \to \infty$. However, due to the difference in the order of the error term, Model (10) is not directly comparable to the models in Chong (2001), Pang, et al (2014) and Jiang, et al (2014). To facilitate such a comparison, we divide both sides of Model (10) by $\sqrt{h}$ and denote $Y_t = y_{th}/\sqrt{h}$, $e_t = \varepsilon_{th}$. Then, we have, for $t = 1, ..., T$,

$$Y_t = \left(1 - (\kappa + \delta 1_{[t>k_0]})h\right) Y_{t-1} + \sigma e_t,\ e_t \sim N(0,1),\ Y_0 = y_0/\sqrt{h} = O_p\left(1/\sqrt{h}\right). \tag{11}$$

To obtain the invariance principle, we remove the assumption of Gaussian errors in the development of the in-fill asymptotic theory and get

$$Y_t = \left(1 - (\kappa + \delta 1_{[t>k_0]})h\right) Y_{t-1} + \sigma e_t,\ e_t \overset{iid}{\sim} (0,1)\ ,\ Y_0 = O_p\left(\sqrt{T}\right). \tag{12}$$

Comparing Model (12) to the models in Chong (2001), Pang, Zhang and Chong (2014) and Jiang, et al (2014), we see two main differences. First, the break size is δh in our discrete time model, which is of the order $O(T^{-1})$. This break size is smaller than those considered in the literature; see Table 1 in Section 2. It is also smaller than that in Elliott and Müller (2007) where the break size is assumed to be $O(T^{-1/2})$. Second, the initial condition in Model (12) is $O_p(T^{-1/2})$ in our discrete time model. This initial condition is larger than all the initial conditions assumed in the literature; also see Table 1 in Section 2.

Following Phillips (1987b), we construct

$$X_T(r) = T^{-\frac{1}{2}}\sigma^{-1}\sum_{t=1}^{\lfloor Tr\rfloor}\sigma e_t = T^{-\frac{1}{2}}\sigma^{-1}\sum_{t=1}^{j-1}\sigma e_t, \quad (j-1)/T \leq r < j/T \quad (j=1,\ldots,T),$$

where $\lfloor\cdot\rfloor$ is the integer-valued function. We start from the asymptotic theory for the sample moments generated by (12). As given in the following lemma, $\widetilde{J}_{\kappa,\delta,\tau}(r)$ plays a central role in the asymptotic theory.

Lemma 4.1 *If $\{Y_t\}$ is a time series generated by (12), then, as $T\to\infty$:*
(a) $T^{-\frac{1}{2}}Y_{\lfloor Tr\rfloor} \Rightarrow \sigma\widetilde{J}_{\kappa,\delta,\tau_0}(r) + e^{-r\kappa-(r-\tau_0)\delta 1_{[r>\tau_0]}}y_0$;
(b) $T^{-\frac{3}{2}}\sum_{t=1}^{\lfloor Tr\rfloor}Y_t \Rightarrow \int_0^r\left[\sigma\widetilde{J}_{\kappa,\delta,\tau_0}(s) + e^{-s\kappa-(s-\tau_0)\delta 1_{[s>\tau_0]}}y_0\right]ds$;
(c) $T^{-2}\sum_{t=1}^{\lfloor Tr\rfloor}Y_t^2 \Rightarrow \int_0^r\left[\sigma\widetilde{J}_{\kappa,\delta,\tau_0}(s) + e^{-s\kappa-(s-\tau_0)\delta 1_{[s>\tau_0]}}y_0\right]^2 ds$;
(d) $T^{-1}\sum_{t=1}^{\lfloor Tr\rfloor}Y_{t-1}e_t \Rightarrow \int_0^r\left[\sigma\widetilde{J}_{\kappa,\delta,\tau_0}(s) + e^{-s\kappa-(s-\tau_0)\delta 1_{[s>\tau_0]}}y_0\right]dB(s)$;
with $\Rightarrow$ denoting the weak convergence of probability measures.

We now develop the asymptotic theorem of the LS estimator of $\tau_0 = k_0/T$ in Model (12) under the in-fill asymptotic scheme where $h\to 0$ with a fixed time span $Th=1$.

When κ and δ are known, the LS estimator of k in Model (12) is defined as

$$\hat{k}_{LS,T} = \arg\min_{k=1,\ldots,T-1} S(k) \tag{13}$$

with

$$\begin{aligned} S(k) &= \sum_{t=1}^{k}\left(Y_t-(1-\kappa h)Y_{t-1}\right)^2 + \sum_{t=k+1}^{T}\left(Y_t-(1-(\kappa+\delta)h)Y_{t-1}\right)^2 \\ &= \sum_{t=1}^{k}\left(Y_t-(1-\kappa h)Y_{t-1}\right)^2 + \sum_{t=k+1}^{T}\left(Y_t-(1-\kappa h)Y_{t-1}+\delta hY_{t-1}\right)^2 \\ &= \sum_{t=1}^{T}\left(Y_t-(1-\kappa h)Y_{t-1}\right)^2 + 2\sum_{t=k+1}^{T}\left(Y_t-(1-\kappa h)Y_{t-1}\right)\delta hY_{t-1} + \sum_{t=k+1}^{T}\left(\delta hY_{t-1}\right)^2 \end{aligned}$$

Since $\sum_{t=1}^{T}\left(Y_t-(1-\kappa h)\,Y_{t-1}\right)^2$ does not depend on k, we have

$$\begin{aligned}
\hat{k}_{LS,T} &= \arg\min_{k=1,\ldots,T-1}\left\{2\sum_{t=k+1}^{T}\left(Y_t-(1-\kappa h)\,Y_{t-1}\right)\delta hY_{t-1}+\sum_{t=k+1}^{T}\left(\delta hY_{t-1}\right)^2\right\} \\
&= \arg\min_{k=1,\ldots,T-1}\left\{-2\sum_{t=1}^{k}\left(Y_t-(1-\kappa h)\,Y_{t-1}\right)\delta hY_{t-1}-\sum_{t=1}^{k}\left(\delta hY_{t-1}\right)^2\right\} \\
&= \arg\max_{k=1,\ldots,T-1}\left\{\sum_{t=1}^{k}\frac{1}{\sigma^2}\left(Y_t-(1-\kappa h)\,Y_{t-1}\right)\delta hY_{t-1}+\frac{1}{2}\sum_{t=1}^{k}\left(\frac{\delta}{\sigma}hY_{t-1}\right)^2\right\}. \quad (14)
\end{aligned}$$

Theorem 4.1 *Consider Model (12) with known κ and δ. Denote the LS estimator $\hat{\tau}_{LS,T}=\hat{k}_{LS,T}/T$ with $\hat{k}_{LS,T}$ defined in (14). Then, when $h\to 0$, we have the in-fill asymptotic distribution as*

$$\begin{aligned}
\hat{\tau}_{ML} \xrightarrow{d} \arg\max_{\tau\in(0,1)} &\left\{\int_{\tau_0}^{\tau}\frac{1}{\delta}\left[\widetilde{J}_{\kappa,\delta,\tau_0}(s)+\exp(-s\kappa-(s-\tau_0)\delta 1_{[s>\tau_0]})\frac{y(0)}{\sigma}\right]dB(s)\right. \\
&\left.-\frac{1}{2}\left|\int_{\tau_0}^{\tau}\left[\widetilde{J}_{\kappa,\delta,\tau_0}(s)+\exp(-s\kappa-(s-\tau_0)\delta 1_{[s>\tau_0]})\frac{y(0)}{\sigma}\right]^2 ds\right|\right\}.
\end{aligned}$$

where $B(s)$ is the standard Brownian motion corresponding to $\widetilde{J}_{\kappa,\delta,\tau}(s)$ and we use the notation $\int_{\tau_0}^{\tau}(\cdot)dB(s)$ to denote $\left[\int_0^{\tau}(\cdot)dB(s)-\int_0^{\tau_0}(\cdot)dB(s)\right]$.

Remark 4.1 *The in-fill asymptotic distribution is the same as the exact distribution derived in the continuous time model when a continuous record is available. This is expected as the in-fill limit of our discrete time AR model with a break converges to the continuous time model with a break. Not surprisingly, the in-fill asymptotic distribution has trimodality and asymmetric for all values of τ_0, even when $\tau_0=50\%$.*

5 Monte Carlo Results

In this section, we design seven Monte Carlo experiments to compare the performance of our in-fill asymptotic distributions with their corresponding long-span asymptotic distributions developed in the literature. In each experiment, we draw densities of the long-span asymptotic distribution and our in-fill asymptotic distribution together and compare them with their corresponding finite sample distribution. The seven experiments are selected to ensure that all the available long-span asymptotic distributions are covered.

In each experiment, data are generated from Model (12) with $\tau_0 = 0.3, 0.5, 0.7$, $\sigma = 1$, $e_t \overset{iid}{\sim} N(0,1)$, $T = 200$ (ie $h = 1/200$) and different combination of κ and δ. In the cases where the first regime is unit root, we set $y_0 = 0$. Otherwise, we randomly draw y_0 from $N(0, \sigma^2/2\kappa)$. All the pdfs are obtained by simulated data with 100,000 replications. When we calculate the in-fill asymptotic distribution and the long-span asymptotic distribution, the stochastic integrals are approximated over very small grid size (0.0001). Let β_1 and β_1 be the corresponding AR(1) coefficients before and after the break.

In the first experiment, we first set $\kappa = 138$ and $\delta = 55$, which implies $\beta_1 = 0.5$ and $\beta_2 = 0.38$. Then we set $\kappa = 138$ and $\delta = -20$, which implies $\beta_1 = 0.5$ and $\beta_2 = 0.55$ and leads to a small break size. For both cases in this experiment, we assume the AR(1) model switches from a stationary root to another stationary root. In this experiment the long-span asymptotic distribution is given in Section 2.1. The three densities are plotted in Figures 9-10.

In the second experiment, we first set $\kappa = 138$ and $\delta = -138$, implying $\beta_1 = 0.5$ and $\beta_2 = 1$.Then we set $\kappa = 61$ and $\delta = -61$, implying $\beta_1 = 0.73$ and $\beta_2 = 1$. Finally, we set $\kappa = 10$ and $\delta = -10$, implying $\beta_1 = 0.95$ and $\beta_2 = 1$. For all cases in this experiment, we assume the AR(1) model switches from a stationary root to a unit root, but the break size gets smaller. In this experiment the long-span asymptotic distribution is given in Section 2.2. The three densities are plotted in Figures 11-13.

In the third experiment, we set $\kappa = 0$ and $\delta = 1$, implying $\beta_1 = 1$ and $\beta_2 = 0.995$. We assume the AR(1) model switches from a unit root to a stationary root, and hence, the long-span asymptotic distribution is given in Section 2.3. The three densities are plotted in Figures 14.

In the fourth experiment, we first set $\kappa = 138$ and $\delta = -137$, implying $\beta_1 = 0.5$ and $\beta_2 = 0.995$. Then we set $\kappa = 10$ and $\delta = -9$, implying $\beta_1 = 0.95$ and $\beta_2 = 0.995$. The second case has a smaller break size than the first case. For both cases in this experiment, we assume the AR(1) model switches from a stationary root to a local-to-unit-root, and hence, the long-span asymptotic distribution is given in Section 2.4 where we set $c = 1$ in the long-span asymptotic distribution. The three densities are plotted in Figures 15-16.

In the fifth experiment, we set $\kappa = 1$ and $\delta = 5$, implying $\beta_1 = 0.995$ and $\beta_2 = 0.97$. In this experiment, we assume the AR(1) model switches from a local-to-unit-root to a stationary root, and hence, the long-span asymptotic distribution is given in Section 2.5 where we set $c = 1$ in the long-span asymptotic distribution. The three densities are plotted in Figure 17.

In the sixth experiment, we set $\kappa = 10$ and $\delta = -10$, implying $\beta_1 = 0.95$ and $\beta_2 = 1$.In this experiment, we assume the AR(1) model switches from a mildly stationary root to a unit root, and hence, the long-span asymptotic distribution is given in Section 2.6 where we set $c = 1$ and $k_T = 20$ in the long-span asymptotic distribution. The three densities are plotted in Figure 18.

In the seventh experiment, we first set $\kappa = 0$ and $\delta = 7$, implying $\beta_1 = 1$ and $\beta_2 = 0.96$. Then we set $\kappa = 0$ and $\delta = 1$, implying $\beta_1 = 1$ and $\beta_2 = 0.995$. For both cases in this experiment, we assume the AR(1) model switches a unit root to a mildly stationary root. Hence, the long-span asymptotic distribution is given in Section 2.7 where we set $k_T = 30$ in the long-span asymptotic distribution, so that $c = 1.2$ in the former case and $c = 0.15$ in the latter case. The three densities are plotted in Figures 19-20.

Several features are apparent in these figures. First, the finite sample distribution is asymmetric about 0 even when $\tau_0 = 50\%$. This feature is different from that in the estimation of break point in the mean discussed in JWY (2016), where the finite sample distribution is symmetric about zero when $\tau_0 = 50\%$. Second, the finite sample distribution has trimodality. The origin is one of the three modes and the two boundary points are the other two. Third and most importantly, the in-fill asymptotic distribution given in Theorem 4.1 has trimodality and is asymmetric about zero, just like the finite sample distribution. It always provides better approximations to the finite sample distribution than the long-span limiting distribution, despite that the sample size is reasonably large ($T = 200$). When the signal-to-noise ratio is small, such as in case 3 in Experiment 2 where $\beta_1 = 0.95$ and $\beta_2 = 1$, the long-span limiting distribution is so far away from the finite sample distribution that any meaningful statistical inference should not be based on the long-span limiting distribution. Arguably, the switch from $\beta_1 = 0.95$ to $\beta_2 = 1$ is empirically interesting and relevant. When the signal-to-noise ratio gets larger, the trimodality becomes less pronounced and the degree of asymmetry reduces. However, the in-fill asymptotic distribution continues to provide better approximations to the finite sample distribution than the long-span asymptotic distribution.

6 Conclusions

This paper is concerned about the large sample approximation to the exact distribution in the estimation of structural break point in autoregressive models. Based on the Girsanov theorem, we obtain the exact distribution of the ML estimator of structural break point in the OU process when a continuous record is available. We find that the

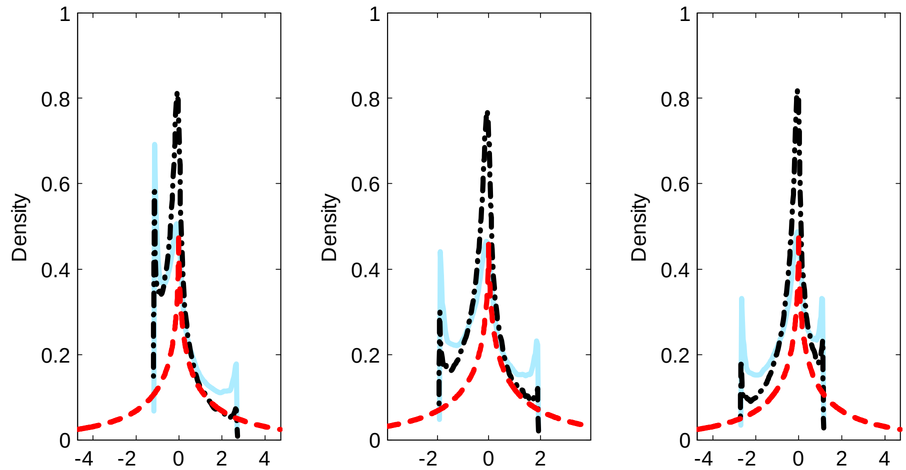

Figure 9: The pdf of $\frac{T(\beta_2-\beta_1)^2}{1-\beta_1^2}(\hat{\tau}_{LS,T}-\tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\beta_1 = 0.5$, $\beta_2 = 0.38$. The blue solid line is the finite sample distribution when $T = 200$; the black broken line is the density given in Theorem 4.1; and the red dotted line is the long-span limiting distribution in given Section 2.1.

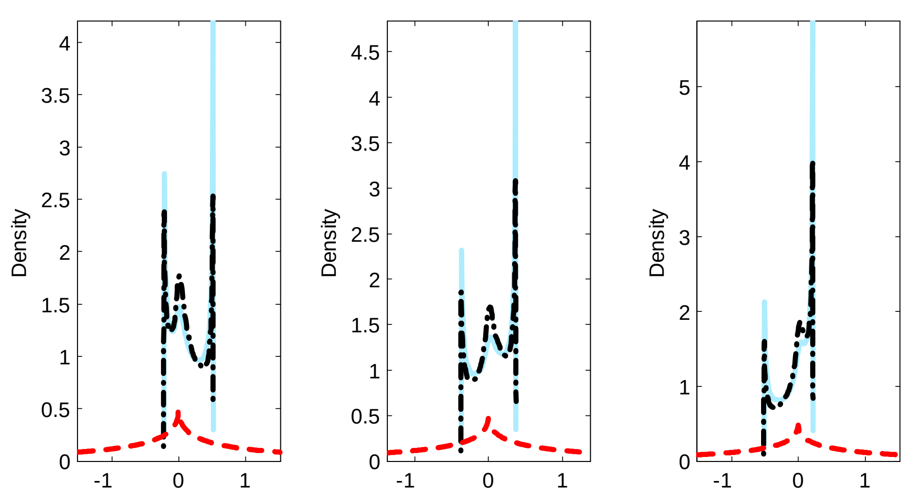

Figure 10: The pdf of $\frac{T(\beta_2-\beta_1)^2}{1-\beta_1^2}(\hat{\tau}_{LS,T}-\tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\beta_1 = 0.5$, $\beta_2 = 0.55$. The blue solid line is the finite sample distribution when $T = 200$; the black broken line is the density given in Theorem 4.1; and the red dotted line is the long-span limiting distribution in given Section 2.1.

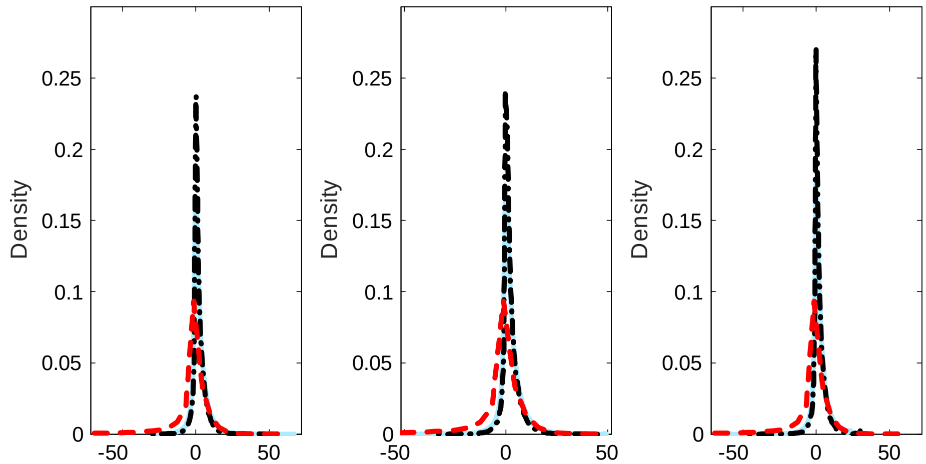

Figure 11: The pdf of $T(1-\beta_1)(\hat{\tau}_{LS,T}-\tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\beta_1 = 0.5$, $\beta_2 = 1$. The blue solid line is the finite sample distribution when $T = 200$; the black broken line is the density given in Theorem 4.1; and the red dotted line is the long-span limiting distribution in given Section 2.2.

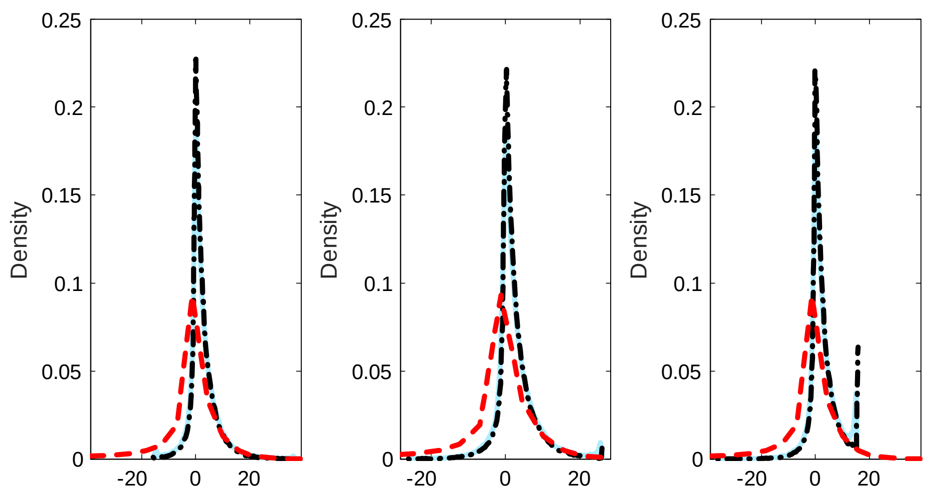

Figure 12: The pdf of $T(1-\beta_1)(\hat{\tau}_{LS,T}-\tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\beta_1 = 0.73$, $\beta_2 = 1$. The blue solid line is the finite sample distribution when $T = 200$; the black broken line is the density given in Theorem 4.1; and the red dotted line is the long-span limiting distribution in given Section 2.2.

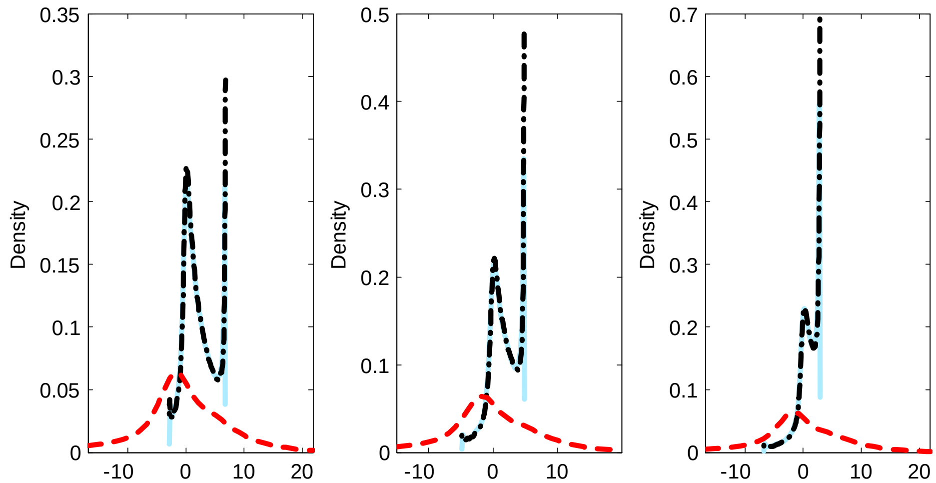

Figure 13: The pdf of $T(1-\beta_1)(\hat{\tau}_{LS,T}-\tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\beta_1 = 0.95$, $\beta_2 = 1$. The blue solid line is the finite sample distribution when $T = 200$; the black broken line is the density given in Theorem 4.1; and the red dotted line is the long-span limiting distribution in given Section 2.2.

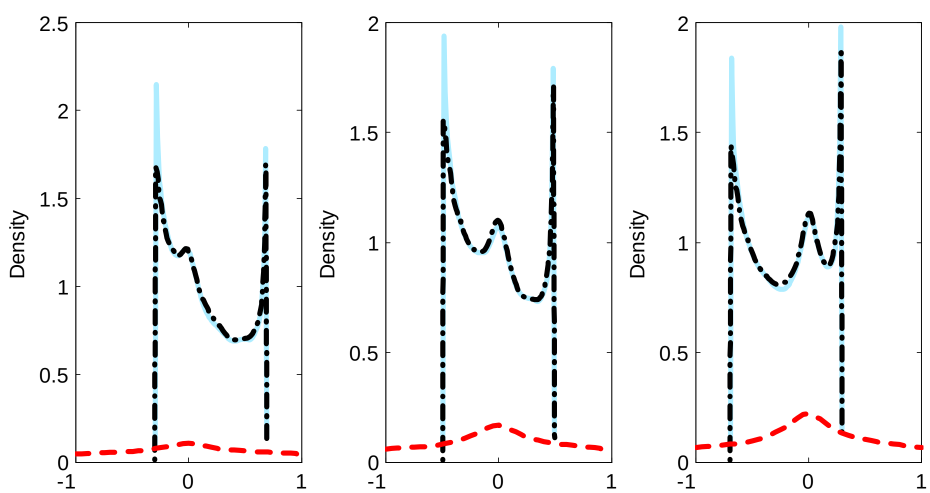

Figure 14: The pdf of $T^2(\beta_2 - 1)^2(\hat{\tau}_{LS,T} - \tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\beta_1 = 1$, $\beta_2 = 0.995$. The blue solid line is the finite sample distribution when $T = 200$; the black broken line is the density given in Theorem 4.1; and the red dotted line is the long-span limiting distribution in given Section 2.3.

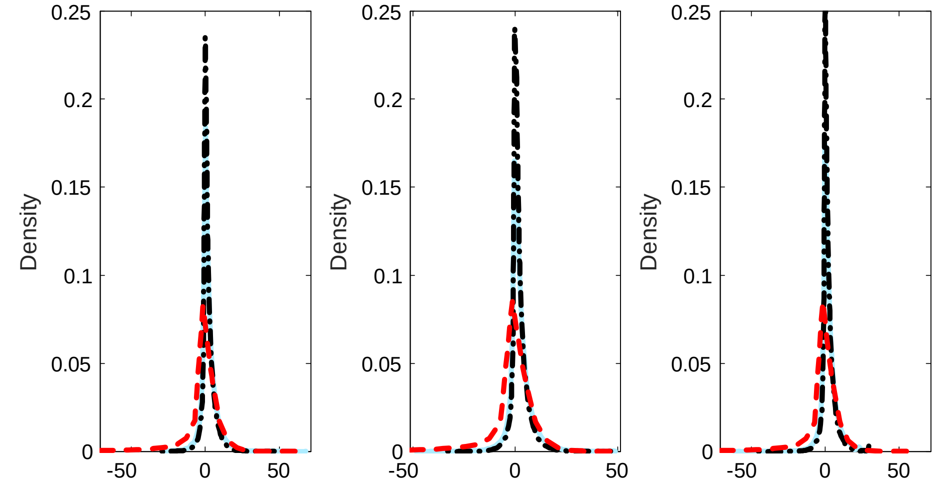

Figure 15: The pdf of $T(\beta_2 - \beta_1)(\hat{\tau}_{LS,T} - \tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\beta_1 = 0.5$, $\beta_2 = 0.995$. The blue solid line is the finite sample distribution when $T = 200$; the black broken line is the density given in Theorem 4.1; and the red dotted line is the long-span limiting distribution in given Section 2.4.

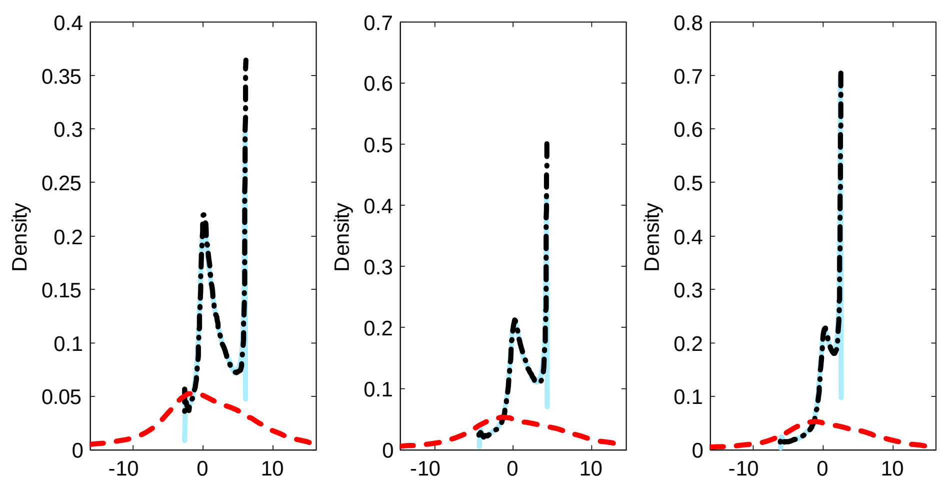

Figure 16: The pdf of $T(\beta_2 - \beta_1)(\hat{\tau}_{LS,T} - \tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\beta_1 = 0.95$, $\beta_2 = 0.995$. The blue solid line is the finite sample distribution when $T = 200$; the black broken line is the density given in Theorem 4.1; and the red dotted line is the long-span limiting distribution in given Section 2.4.

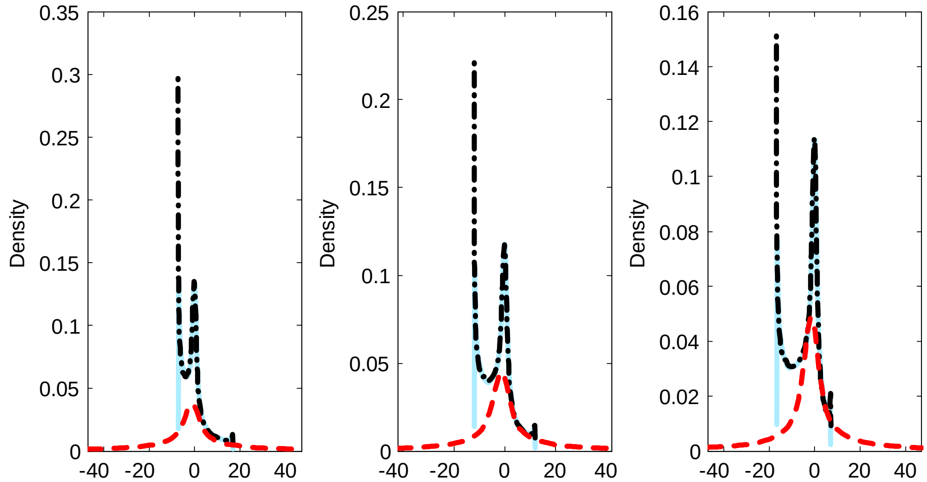

Figure 17: The pdf of $T^2(\beta_2 - \beta_1)^2(\hat{\tau}_{LS,T} - \tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\beta_1 = 0.995$, $\beta_2 = 0.97$. The blue solid line is the finite sample distribution when $T = 200$; the black broken line is the density given in Theorem 4.1; and the red dotted line is the long-span limiting distribution in given Section 2.5.

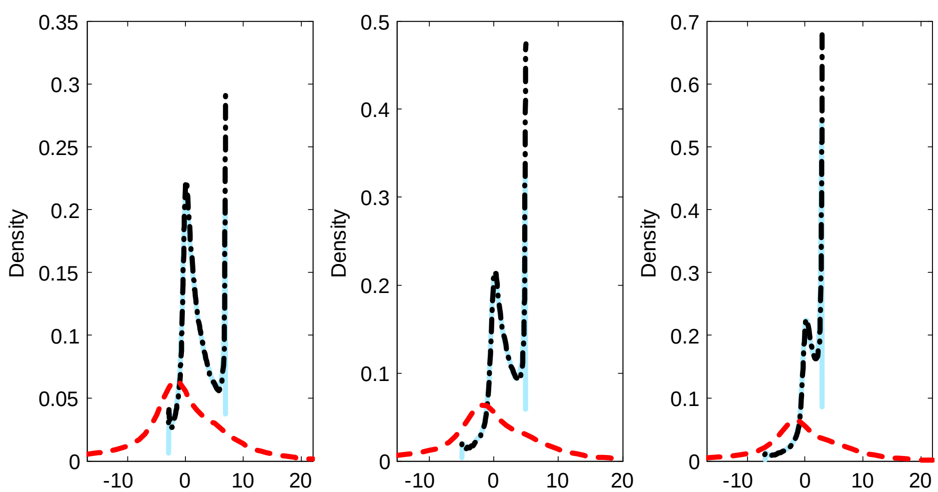

Figure 18: The pdf of $\frac{cT}{k_T}(\hat{\tau}_{LS,T} - \tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\beta_1 = 0.95$, $\beta_2 = 1, c = 1, k_T = 20$. The blue solid line is the finite sample distribution when $T = 200$; the black broken line is the density given in Theorem 4.1; and the red dotted line is the long-span limiting distribution in given Section 2.6.

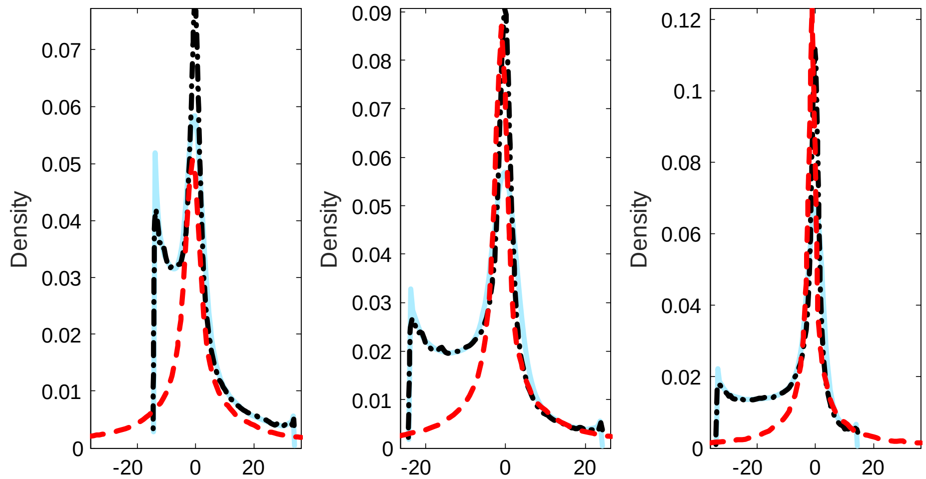

Figure 19: The pdf of $\frac{c^2T^2}{k_T^2}(\hat{\tau}_{LS,T} - \tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\beta_1 = 1$, $\beta_2 = 0.96, c = 1.2, k_T = 30$. The blue solid line is the finite sample distribution when $T = 200$; the black broken line is the density given in Theorem 4.1; and the red dotted line is the long-span limiting distribution in given Section 2.7.

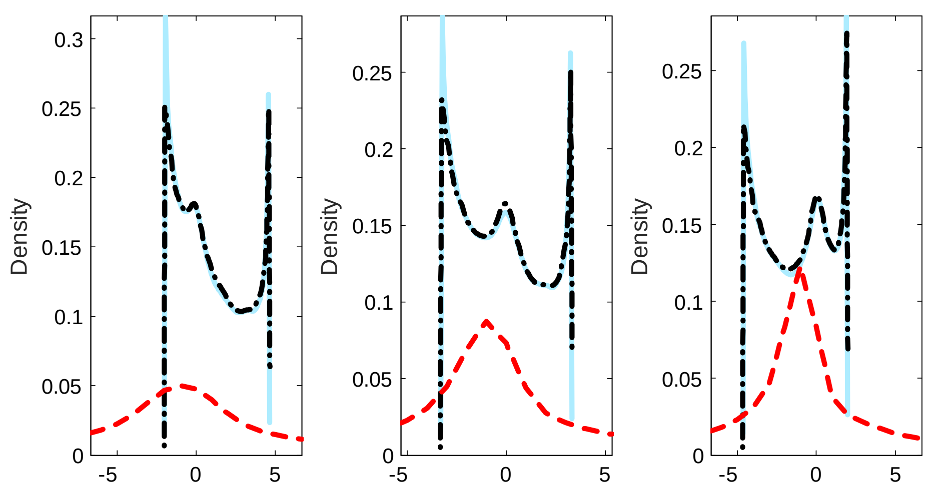

Figure 20: The pdf of $\frac{c^2T^2}{k_T^2}(\hat{\tau}_{LS,T} - \tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\beta_1 = 1$, $\beta_2 = 0.995, c = 0.15, k_T = 30$. The blue solid line is the finite sample distribution when $T = 200$; the black broken line is the density given in Theorem 4.1; and the red dotted line is the long-span limiting distribution in given Section 2.7.

exact distribution is asymmetric and has trimodality. These two properties are also found in the finite sample distribution of the LS estimator of structural break point in the AR model.

Unfortunately, the literature on the estimation of structural break point in AR(1) models has always focused on developing asymptotic theory with the time spans before and after the break being assumed to go to infinity, which has been shown in this paper to provide poor approximations to the finite sample distribution in many empirically relevant cases. Moreover, the long-span asymptotics developed in the literature are different whether the underlying AR(1) coefficient is less than, equal to or greater than one. This discontinuity in the long-span asymptotic distributions makes the limiting distributions developed in the literature difficult to use in practice.

This paper provides a unified limiting theory for estimating the break point in AR(1) models when nusiance parameters κ and δ are known. It considers a continuous time approximation to the discrete time AR model and develops an in-fill asymptotic theory for the LS estimator of structural break point. The developed in-fill asymptotic theory is continuous in the underlying roots and, hence, offers a unified theory for making inference about the break point. We also show that this distribution has trimodality and is asymmetric, and approximates the finite sample distribution better than the long-span limiting distribution developed in the literature in all cases.

The same method can be used to provide a unified limiting theory for estimating the break point in AR(1) models when all the nuisance parameters are unknown. As shown in LWY (2016) for the model with a break in mean, the presence of unknown nuisance parameters has implications for the in-fill asymptotic theory. The same kind of implications will be applicable in AR(1) models with a break. The results will be reported in another paper.

Appendix

Proof of Lemma 3.1: (a) It is obvious that $E\left(\widetilde{J}_{\kappa,\delta,\tau}(r)\right) = 0$.

For any fixed $r > 0$, when $r \leq \tau$, it is straightforward that $\sigma_J^2(r) = \frac{1}{2}(1 - \exp(-2\kappa r))/\kappa$.

When $r > \tau$, from (4), we have

$$
\begin{aligned}
\widetilde{J}_{\kappa,\delta,\tau}(r) &= \int_0^\tau \exp(-(\kappa+\delta)(r-s)+\delta(\tau-s))dB(s) + \int_\tau^r \exp(-(\kappa+\delta)(r-s))dB(s) \\
&= \int_0^\tau \exp(-(\kappa+\delta)r+\delta\tau+\kappa s)dB(s) + \int_\tau^r \exp(-(\kappa+\delta)(r-s))dB(s) \\
&= \int_0^\tau \exp(-(\kappa+\delta)r+(\kappa+\delta)\tau-\kappa\tau+\kappa s)dB(s) + \int_\tau^r \exp(-(\kappa+\delta)(r-s))dB(s) \\
&= \int_0^\tau \exp(-(\kappa+\delta)(r-\tau)-\kappa(\tau-s))dB(s) + \int_\tau^r \exp(-(\kappa+\delta)(r-s))dB(s) \\
&= \exp(-(\kappa+\delta)(r-\tau))\int_0^\tau \exp(-\kappa(\tau-s))dB(s) + \int_\tau^r \exp(-(\kappa+\delta)(r-s))dB(s).
\end{aligned}
$$

$$
\begin{aligned}
\sigma_J^2(r) &= \exp\left(-2(\kappa+\delta)(r-\tau)\right)\left(\frac{1}{2\kappa}(1-\exp(-2\kappa\tau))\right) + \int_\tau^r \exp\left(-2(\kappa+\delta)(r-s)\right)ds. \\
&= \frac{1}{2\kappa}\left\{\exp\left(-2(\kappa+\delta)(r-\tau)\right) - \exp\left(-2(\kappa+\delta)r+2\delta\tau\right)\right\} \\
&\quad + \frac{1}{2(\kappa+\delta)}\left(1-\exp(-2(\kappa+\delta)(r-\tau))\right) \\
&= \left(\frac{1}{2\kappa}-\frac{1}{2(\kappa+\delta)}\right)\exp(-2(\kappa+\delta)(r-\tau)) - \frac{1}{2\kappa}\exp\left(-2(\kappa+\delta)r+2\delta\tau\right) + \frac{1}{2(\kappa+\delta)} \\
&= \frac{\delta}{2\kappa(\kappa+\delta)}\exp(-2(\kappa+\delta)(r-\tau)) - \frac{1}{2\kappa}\exp\left(-2(\kappa+\delta)r+2\delta\tau\right) + \frac{1}{2(\kappa+\delta)},
\end{aligned}
$$

which gives the result in Lemma 3.1 (a).

(b) When $r \le \tau$,

$$d\widetilde{J}_{\kappa,\delta,\tau}(r) = -\kappa\widetilde{J}_{\kappa,\delta,\tau}(r) + dB(r),$$

with $\tilde{J}(0) = 0$.

Following Phillips (1987b), we have

$$\widetilde{J}_{\kappa,\delta,\tau}(r) = \int_0^r \exp(-(r-s)\kappa)dB(s).$$

When $r > \tau$,

$$d\widetilde{J}_{\kappa,\delta,\tau}(r) = -(\kappa+\delta)\widetilde{J}_{\kappa,\delta,\tau}(r) + dB(r),$$

with $\widetilde{J}_{\kappa,\delta,\tau}(\tau) = \int_0^\tau \exp(-(\tau-s)\kappa)dB(s)$.

Define the integration factor $v(s) = \exp((\kappa+\delta)s)$. By definition, we have

$$dv(s) = (\kappa+\delta)v(s)ds.$$

When $s > \tau$,we have

$$\begin{aligned} d\left(v(s)\widetilde{J}_{\kappa,\delta,\tau}(s)\right) &= v(s)d\widetilde{J}_{\kappa,\delta,\tau}(s) + \widetilde{J}_{\kappa,\delta,\tau}(s)dv(s) \\ &= v(s)\left(-(\kappa+\delta)\widetilde{J}_{\kappa,\delta,\tau}(s)ds + dB(s)\right) + \widetilde{J}_{\kappa,\delta,\tau}(s)(\kappa+\delta)v(s)ds \\ &= v(s)dB(s) \\ &= \exp((\kappa+\delta)s)dB(s), \end{aligned}$$

which leads to

$$\int_\tau^r d\left(v(s)\widetilde{J}_{\kappa,\delta,\tau}(s)\right) = \int_\tau^r \exp((\kappa+\delta)s)dB(s).$$

Therefore, we have

$$\begin{aligned} v(r)\widetilde{J}_{\kappa,\delta,\tau}(r) &= v(\tau)\widetilde{J}_{\kappa,\delta,\tau}(\tau) + \int_\tau^r \exp((\kappa+\delta)s)dB(s) \\ &= \exp((\kappa+\delta)\tau)\int_0^\tau \exp(-\kappa(\tau-s))dB(s) + \int_\tau^r \exp((\kappa+\delta)s)dB(s). \end{aligned} \tag{15}$$

So

$$\begin{aligned} \widetilde{J}_{\kappa,\delta,\tau}(r) &= \exp(-(\kappa+\delta)(r-\tau))\int_0^\tau \exp(-\kappa(\tau-s))dB(s) + \int_\tau^r \exp(-(\kappa+\delta)(r-s))dB(s) \\ &= \int_0^\tau \exp(-(r-s)(\kappa+\delta)+(\tau-s)\delta)dB(s) + \int_\tau^r \exp(-(r-s)(\kappa+\delta))dB(s), \end{aligned}$$

where the first equality is obtained by dividing equation (15) by $v(r)$. The proof of Lemma 3.1 (b) is completed.

Proof of Lemma 3.2: When $t \leq \tau_0$,

$$dy(t) = -\kappa y(t)dt + \sigma dB(t),$$

with the initial condition $y(0)$. Following Phillips (1987), we know that

$$y(t) = \sigma\tilde{J}_{\kappa,\delta,\tau_0}(t) + \exp(-\kappa t)y(0). \tag{16}$$

When $t > \tau_0$,

$$dy(t) = -(\kappa+\delta)y(t)dt + \sigma dB(t)$$

with $y(\tau_0) = \sigma\widetilde{J}_{\kappa,\delta,\tau_0}(\tau_0) + \exp(-\kappa\tau_0)y(0)$.

By methods similar to those in the proof of Lemma 3.1 (b), we have

$$\begin{aligned}
y(t) &= \exp(-(\kappa+\delta)(t-\tau_0))y(\tau_0) + \sigma \int_{\tau_0}^{t} \exp(-(\kappa+\delta)(t-s))dB(s) \\
&= \exp(-(\kappa+\delta)(t-\tau_0)) \left(\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(\tau_0) + \exp(-\kappa\tau_0)y(0)\right) + \sigma \int_{\tau_0}^{t} \exp(-(\kappa+\delta)(t-s))dB(s) \\
&= \sigma \left(\exp(-(\kappa+\delta)(t-\tau_0))\widetilde{J}_{\kappa,\delta,\tau_0}(\tau_0) + \int_{\tau_0}^{t} \exp(-(\kappa+\delta)(t-s))dB(s) \right) \\
&\quad + \exp(-(\kappa+\delta)(t-\tau_0) - \kappa\tau_0)y(0) \\
&= \sigma \widetilde{J}_{\kappa,\delta,\tau_0}(t) + \exp(-t\kappa - (t-\tau_0)\delta)y(0). \qquad (17)
\end{aligned}$$

where the last equation is from (6).

Alternatively, we can write (16) and (17) as

$$y(t) = \sigma \widetilde{J}_{\kappa,\delta,\tau_0}(t) + \exp\left(-t\kappa - (t-\tau_0)\delta 1_{[t>\tau_0]}\right) y(0),$$

which gives the result in Lemma 3.2.

Proof of Lemma 3.3: First write $g(y(s)) = y^2(s)$. Since g function is twice continuously differentiable, by Ito's Lemma, we have

$$dg(y(s)) = \frac{dg}{dy}(y(s))dy(s) + \frac{1}{2}\frac{d^2 g}{dy^2}(y(s))(dy(s))^2,$$

which implies

$$\begin{aligned}
dy^2(s) &= 2y(s)dy(s) + (dy(s))^2 \\
&= 2y(s)\left[-(\kappa + \delta 1_{[s>\tau_0]})y(s)ds + \sigma dB(s)\right] + \sigma^2 ds \\
&= -2(\kappa + \delta 1_{[s>\tau_0]})y^2(s)ds + 2\sigma y(s)dB(s) + \sigma^2 ds
\end{aligned}$$

from which we obtain the result in Lemma 3.3 by taking the integral over $[0, r]$ at both sides

$$2\sigma \int_0^r y(s)dB(s) = y^2(r) - y^2(0) + 2\int_0^r (\kappa + \delta 1_{[s>\tau_0]})y^2(s)ds - r\sigma^2.$$

Proof of Theorem 3.1: Note that

$$
\begin{aligned}
\hat{\tau}_{ML} &= \arg\max_{\tau\in(0,1)} \{\log \mathcal{L}(\tau)\} = \arg\max_{\tau\in(0,1)} \log\left(\frac{dP_\tau}{dB_t}\right) \\
&= \arg\max_{\tau\in(0,1)} \left[\log\left(\frac{dP_\tau}{dB_t}\right) - \log\left(\frac{dP_{\tau_0}}{dB_t}\right)\right] \\
&= \arg\max_{\tau\in(0,1)} \log\left(\frac{dP_\tau}{dP_{\tau_0}}\right),
\end{aligned}
$$

where $\log\left(\frac{dP_\tau}{dP_{\tau_0}}\right)$ is the log-likelihood ratio with the expression

$$
\begin{aligned}
\log\left(\frac{dP_\tau}{dP_{\tau_0}}\right) &= \frac{1}{\sigma^2}\left\{\int_0^1 \left(-(\kappa+\delta 1_{[t>\tau]}) + (\kappa+\delta 1_{[t>\tau_0]})\right) y(t)dy(t)\right. \\
&\quad \left. -\frac{1}{2}\int_0^1 \left((\kappa+\delta 1_{[t>\tau]})^2 - (\kappa+\delta 1_{[t>\tau_0]})^2\right) y^2(t)dt\right\} \\
&= \int_0^1 \frac{\delta}{\sigma}(1_{[t>\tau_0]} - 1_{[t>\tau]})y(t)dB(t) - \frac{1}{2}\int_0^1 \left(\frac{\delta}{\sigma}\right)^2 (1_{[t>\tau_0]} - 1_{[t>\tau]})^2 y^2(t)dt
\end{aligned}
$$

When $\tau \le \tau_0$, we have

$$
\begin{aligned}
\log\left(\frac{dP_\tau}{dP_{\tau_0}}\right) &= -\frac{\delta}{\sigma}\int_0^1 1_{[\tau<t\le\tau_0]}y(t)dB(t) - \frac{1}{2}\left(\frac{\delta}{\sigma}\right)^2 \int_0^1 1_{[\tau<t\le\tau_0]}y^2(t)dt \\
&= -\frac{\delta}{\sigma}\int_\tau^{\tau_0} y(t)dB(t) - \frac{1}{2}\left(\frac{\delta}{\sigma}\right)^2 \int_\tau^{\tau_0} y^2(t)dt \\
&= -\frac{\delta}{\sigma}\left[\int_0^{\tau_0} y(t)dB(t) - \int_0^{\tau} y(t)dB(t)\right] - \frac{1}{2}\left(\frac{\delta}{\sigma}\right)^2 \int_\tau^{\tau_0} y^2(t)dt \\
&= \frac{\delta}{\sigma}\left[\int_0^{\tau} y(t)dB(t) - \int_0^{\tau_0} y(t)dB(t)\right] - \frac{1}{2}\left(\frac{\delta}{\sigma}\right)^2 \int_\tau^{\tau_0} y^2(t)dt
\end{aligned}
$$

When $\tau > \tau_0$, we have

$$
\begin{aligned}
\log\left(\frac{dP_\tau}{dP_{\tau_0}}\right) &= \frac{\delta}{\sigma}\int_0^1 1_{[\tau_0<t\le\tau]}y(t)dB(t) - \frac{1}{2}\left(\frac{\delta}{\sigma}\right)^2 \int_0^1 1_{[\tau_0<t\le\tau]}y^2(t)dt \\
&= \frac{\delta}{\sigma}\int_{\tau_0}^{\tau} y(t)dB(t) - \frac{1}{2}\left(\frac{\delta}{\sigma}\right)^2 \int_{\tau_0}^{\tau} y^2(t)dt \\
&= \frac{\delta}{\sigma}\left[\int_0^{\tau} y(t)dB(t) - \int_0^{\tau_0} y(t)dB(t)\right] - \frac{1}{2}\left(\frac{\delta}{\sigma}\right)^2 \int_{\tau_0}^{\tau} y^2(t)dt
\end{aligned}
$$

Therefore, the exact log-likelihood ratio can be written as

$$
\log\left(\frac{dP_\tau}{dP_{\tau_0}}\right) = \frac{\delta}{\sigma}\left[\int_0^{\tau} y(t)dB(t) - \int_0^{\tau_0} y(t)dB(t)\right] - \frac{1}{2}\left|\int_{\tau_0}^{\tau} \left(\frac{\delta}{\sigma}\right)^2 y^2(t)dt\right|.
$$

This implies that the ML estimator of break point is

$$\hat{\tau}_{ML} \stackrel{d}{=} \arg\max_{\tau\in(0,1)} \left\{ \frac{\delta}{\sigma} \left[\int_0^{\tau} y(t)dB(t) - \int_0^{\tau_0} y(t)dB(t) \right] - \frac{1}{2} \left| \int_{\tau_0}^{\tau} \left(\frac{\delta}{\sigma}\right)^2 y^2(t)dt \right| \right\}$$

$$= \arg\max_{\tau\in(0,1)} \left\{ \frac{1}{\delta} \left[\int_0^{\tau} \frac{y(t)}{\sigma} dB(t) - \int_0^{\tau_0} \frac{y(t)}{\sigma} dB(t) \right] - \frac{1}{2} \left| \int_{\tau_0}^{\tau} \left(\frac{y(t)}{\sigma}\right)^2 dt \right| \right\},$$

which gives the result in Theorem 3.1.

Proof of Lemma 4.1 (a): Following Phillips (1987b), first note that

$$Y_t = \exp(-(\kappa + \delta 1_{[t>k_0]})/T)Y_{t-1} + \sigma e_t + O_p(T^{-2}), \tag{18}$$

where $O_p(T^{-2})$ term is the approximation error.

When $t \leq k_0$, from (18) we have

$$\begin{aligned} Y_t &= \exp(-\kappa/T)Y_{t-1} + \sigma e_t + O_p(T^{-2}) \\ &= \sigma \sum_{j=1}^{t} \exp(-(t-j)\kappa/T)e_j + \exp(-t\kappa/T)Y_0 + O_p(T^{-1}), \end{aligned}$$

and thus when $r \leq \tau_0$, we have

$$\begin{aligned} T^{-\frac{1}{2}} Y_{\lfloor Tr \rfloor} &= \sigma \sum_{j=1}^{\lfloor Tr \rfloor} \exp(-(\lfloor Tr \rfloor - j)\kappa/T) \int_{(j-1)/T}^{j/T} dX_T(s) + \exp(-\lfloor Tr \rfloor \kappa/T)\frac{Y_0}{\sqrt{T}} + O_p(T^{-\frac{3}{2}}) \\ &= \sigma \sum_{j=1}^{\lfloor Tr \rfloor} \int_{(j-1)/T}^{j/T} \exp(-(r-s)\kappa)dX_T(s) + \exp(-r\kappa)\frac{Y_0}{\sqrt{T}} + O_p(T^{-\frac{3}{2}}) \\ &= \sigma \int_0^r \exp(-(r-s)\kappa)dX_T(s) + \exp(-r\kappa)\frac{Y_0}{\sqrt{T}} + O_p(T^{-\frac{3}{2}}) \\ &\Rightarrow \sigma \int_0^r \exp(-(r-s)\kappa)dB(s) + \exp(-r\kappa)y_0 \\ &= \sigma \widetilde{J}_{\kappa,\delta,\tau_0}(r) + \exp(-r\kappa)y_0, \end{aligned} \tag{19}$$

where $\sigma \int_0^r \exp(-(r-s)\kappa)dX_T(s) \Rightarrow \sigma \int_0^r \exp(-(r-s)\kappa)dB(s)$ is from Phillips (1987b) and the last equation comes from the definition of $\widetilde{J}_{\kappa,\delta,\tau}(r)$ in (4).

Similarly, when $t > k_0$, we have

$$\begin{aligned} Y_t &= \exp(-(\kappa+\delta)/T)Y_{t-1} + \sigma e_t + O_p(T^{-2}) \\ &= \sigma \sum_{j=k_0+1}^{t} \exp(-(t-j)(\kappa+\delta)/T)e_j + \exp(-(t-k_0)(\kappa+\delta)/T)Y_{k_0} + O_p(T^{-1}), \end{aligned}$$

and thus when $r > \tau_0$, we have

$$
\begin{aligned}
T^{-\frac{1}{2}}Y_{\lfloor Tr\rfloor} &= \sigma \sum_{j=\lfloor T\tau_0\rfloor+1}^{\lfloor Tr\rfloor} \exp(-(\lfloor Tr\rfloor - j)(\kappa+\delta)/T)\int_{(j-1)/T}^{j/T} dX_T(s) \\
&\quad + \exp(-(\lfloor Tr\rfloor - \lfloor T\tau_0\rfloor)(\kappa+\delta)/T)\frac{Y_{\lfloor T\tau_0\rfloor}}{\sqrt{T}} + O_p(T^{-\frac{3}{2}}) \\
&= \sigma \sum_{j=\lfloor T\tau_0\rfloor+1}^{\lfloor Tr\rfloor} \int_{(j-1)/T}^{j/T} \exp(-(r-s)(\kappa+\delta))dX_T(s) \\
&\quad + \exp(-(r-\tau_0)(\kappa+\delta))\frac{Y_{\lfloor T\tau_0\rfloor}}{\sqrt{T}} + O_p(T^{-\frac{3}{2}}) \\
&= \sigma \int_{\tau_0}^{r} \exp(-(r-s)(\kappa+\delta))dX_T(s) + \exp(-(r-\tau_0)(\kappa+\delta))\frac{Y_{\lfloor T\tau_0\rfloor}}{\sqrt{T}} + O_p(T^{-\frac{3}{2}}) \\
&\Rightarrow \sigma \int_{\tau_0}^{r} \exp(-(r-s)(\kappa+\delta))dB(s) + \exp(-(r-\tau_0)(\kappa+\delta))\left(\sigma\tilde{J}_{\kappa,\delta,\tau_0}(\tau_0) + \exp(-\tau_0\kappa)y_0\right) \\
&= \sigma\left\{\exp(-(r-\tau_0)(\kappa+\delta))\tilde{J}_{\kappa,\delta,\tau_0}(\tau_0) + \int_{\tau_0}^{r}\exp(-(r-s)(\kappa+\delta))dB(s)\right\} \\
&\quad + \exp(-(r-\tau_0)(\kappa+\delta) - \tau_0\kappa)y_0, \\
&= \sigma\tilde{J}_{\kappa,\delta,\tau_0}(r) + \exp(-(r-\tau_0)\kappa - (r-\tau_0)\delta - \tau_0\kappa)y_0 \\
&= \sigma\tilde{J}_{\kappa,\delta,\tau_0}(r) + \exp(-r\kappa - (r-\tau_0)\delta)y_0. \qquad (20)
\end{aligned}
$$

where the the second last equation is from (6).

The combination of (19) and (20) leads to the result in Lemma 4.1 (a).

The proofs of Lemma 4.1 (b) and (c) are entirely similar to the proof of Lemma 4.1 (a). We skip them for simplicity.

The proof of Lemma 4.1 (d): We first square (12) to obtain

$$
\begin{aligned}
Y_t^2 &= \left[(1-(\kappa+\delta 1_{[t>k_0]})h)Y_{t-1} + \sigma e_t\right]^2 \\
&= (1-(\kappa+\delta 1_{[t>k_0]})h)^2 Y_{t-1}^2 + 2\sigma(1-(\kappa+\delta 1_{[t>k_0]})h)Y_{t-1}e_t + \sigma^2 e_t^2 \\
&= (1-2(\kappa+\delta 1_{[t>k_0]})h)Y_{t-1}^2 + 2\sigma(1-(\kappa+\delta 1_{[t>k_0]})h)Y_{t-1}e_t + \sigma^2 e_t^2 + O_p(T^{-2}) \\
&= \left[1 - \frac{2(\kappa+\delta 1_{[t>k_0]})}{T}\right] Y_{t-1}^2 + 2\sigma Y_{t-1}e_t + \sigma^2 e_t^2 - 2\sigma(\kappa+\delta 1_{[t>k_0]})\frac{Y_{t-1}e_t}{T} + O_p(T^{-2}),
\end{aligned}
$$

and we have

$$
Y_t^2 - Y_{t-1}^2 = -\frac{2(\kappa+\delta 1_{[t>k_0]})}{T}Y_{t-1}^2 + 2\sigma Y_{t-1}e_t + \sigma^2 e_t^2 - 2\sigma(\kappa+\delta 1_{[t>k_0]})\frac{Y_{t-1}e_t}{T} + O_p(T^{-2}),
$$

where by first summing over t from $t=1$ to $t=\lfloor Tr\rfloor$ and then dividing T at both sides, we obtain

$$\begin{aligned}\frac{Y_{\lfloor Tr\rfloor}^2}{T}-\frac{Y_0^2}{T}&=-\frac{2}{T^2}\sum_{t=1}^{\lfloor Tr\rfloor}(\kappa+\delta 1_{[t>k_0]})Y_{t-1}^2+\frac{2\sigma}{T}\sum_{t=1}^{\lfloor Tr\rfloor}Y_{t-1}e_t+\frac{1}{T}\sum_{t=1}^{\lfloor Tr\rfloor}\sigma^2e_t^2\\&\quad-\frac{2\sigma}{T^2}\sum_{t=1}^{\lfloor Tr\rfloor}(\kappa+\delta 1_{[t>k_0]})Y_{t-1}e_t+O_p(T^{-2})\\&=-\frac{2}{T^2}\sum_{t=1}^{\lfloor Tr\rfloor}(\kappa+\delta 1_{[t>k_0]})Y_{t-1}^2+\frac{2\sigma}{T}\sum_{t=1}^{\lfloor Tr\rfloor}Y_{t-1}e_t+\frac{1}{T}\sum_{t=1}^{\lfloor Tr\rfloor}\sigma^2e_t^2+O_p(T^{-\frac{1}{2}}).\end{aligned}\tag{21}$$

To obtain (21), we have claimed $\frac{2\sigma}{T^2}\sum_{t=1}^{\lfloor Tr\rfloor}(\kappa+\delta 1_{[t>k_0]})Y_{t-1}e_t=O_p(T^{-\frac{1}{2}})$ since

$$\begin{aligned}\left|\frac{2\sigma}{T^2}\sum_{t=1}^{\lfloor Tr\rfloor}(\kappa+\delta 1_{[t>k_0]})Y_{t-1}e_t\right|&\leq\frac{2\sigma C}{T^2}\sum_{t=1}^{\lfloor Tr\rfloor}|Y_{t-1}e_t|\\&\leq\frac{2\sigma C}{\sqrt{T}}\max_{t=1,\dots,\lfloor Tr\rfloor}\left|T^{-\frac{1}{2}}Y_{t-1}\right|\frac{1}{T}\sum_{t=1}^{\lfloor Tr\rfloor}|e_t|\\&=O_p(T^{-\frac{1}{2}}),\end{aligned}$$

where C is a positive constant, the last equation comes from Lemma 4.1 (a) for $T^{-\frac{1}{2}}Y_{t-1}$ and the law of large number for the i.i.d sequence $\{|e_t|\}$.

So when $r\leq\tau_0$, from (21) we have

$$\frac{Y_{\lfloor Tr\rfloor}^2}{T}-\frac{Y_0^2}{T}=-\frac{2\kappa}{T^2}\sum_{t=1}^{\lfloor Tr\rfloor}Y_{t-1}^2+\frac{2\sigma}{T}\sum_{t=1}^{\lfloor Tr\rfloor}Y_{t-1}e_t+\frac{1}{T}\sum_{t=1}^{\lfloor Tr\rfloor}\sigma^2e_t^2+O_p(T^{-\frac{1}{2}}),$$

and after some rearrangement, we have

$$\begin{aligned}\frac{2\sigma}{T}\sum_{t=1}^{\lfloor Tr\rfloor}Y_{t-1}e_t&=\frac{Y_{\lfloor Tr\rfloor}^2}{T}-\frac{Y_0^2}{T}+\frac{2\kappa}{T^2}\sum_{t=1}^{\lfloor Tr\rfloor}Y_{t-1}^2-\frac{1}{T}\sum_{t=1}^{\lfloor Tr\rfloor}\sigma^2e_t^2+O_p(T^{-\frac{1}{2}})\\&\Rightarrow\left[\sigma\widetilde{J}_{\kappa,\delta,\tau_0}(r)+\exp(-r\kappa)y_0\right]^2-y_0^2+2\kappa\int_0^r\left[\sigma\widetilde{J}_{\kappa,\delta,\tau_0}(s)+\exp(-s\kappa)y_0\right]^2ds-\sigma^2r\\&=2\sigma\int_0^r\left[\sigma\widetilde{J}_{\kappa,\delta,\tau_0}(s)+\exp(-s\kappa)y_0\right]dB(s),\end{aligned}\tag{22}$$

where the convergence result comes from Lemma 4.1 (a), (c) and the law of large number for the i.i.d sequence $\{e_t^2\}$, and the last equation comes from Lemma 3.2 and 3.3.

Similarly, when $r > \tau_0$, we have

$$\begin{aligned}
\frac{2\sigma}{T}\sum_{t=1}^{\lfloor Tr \rfloor} Y_{t-1}e_t &= \frac{Y_{\lfloor Tr \rfloor}^2}{T} - \frac{Y_0^2}{T} + \frac{2}{T^2}\sum_{t=1}^{\lfloor Tr \rfloor}(\kappa + \delta 1_{[t>k_0]})Y_{t-1}^2 - \frac{1}{T}\sum_{t=1}^{\lfloor Tr \rfloor}\sigma^2 e_t^2 + O_p(T^{-\frac{1}{2}}) \\
&\Rightarrow \left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(r) + \exp(-r\kappa - (r-\tau_0)\delta)y_0\right]^2 - y_0^2 \\
&\quad + 2\int_0^r (\kappa + \delta 1_{[s>\tau_0]})\left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa - (s-\tau_0)\delta 1_{[s>\tau_0]})y_0\right]^2 ds - \sigma^2 r \\
&= 2\sigma\int_0^r \left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa - (s-\tau_0)\delta 1_{[s>\tau_0]})y_0\right] dB(s), \qquad (23)
\end{aligned}$$

where the last equation comes from Lemma 3.2 and 3.3.

The combination of (22) and (23) leads to

$$T^{-1}\sum_{t=1}^{\lfloor Tr \rfloor} Y_{t-1}e_t \Rightarrow \int_0^r \left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa - (s-\tau_0)\delta 1_{[s>\tau_0]})y(0)\right] dB(s),$$

the result in Lemma 4.1 (d).

Proof of Theorem 4.1: From (14), we have

$$\hat{k}_{LS,T} = \arg\max_{k=1,\ldots,T-1}\left\{\sum_{t=1}^{k}\frac{1}{\sigma^2}\left(Y_t - (1-\kappa h)\,Y_{t-1}\right)\delta h Y_{t-1} + \frac{1}{2}\sum_{t=1}^{k}\left(\frac{\delta}{\sigma}hY_{t-1}\right)^2\right\}.$$

Let $\Gamma(k) = \sum_{t=1}^{k}\frac{1}{\sigma^2}\left(Y_t - (1-\kappa h)\,Y_{t-1}\right)\delta h Y_{t-1} + \frac{1}{2}\sum_{t=1}^{k}\left(\frac{\delta}{\sigma}hY_{t-1}\right)^2$. Then, the LS estimator $\hat{k}_{LS,T}$ defined in (13) can be expressed as

$$\hat{k}_{LS,T} = \arg\max_{k=1,\ldots,T-1}\left\{\Gamma(k)\right\} = \arg\max_{k=1,\ldots,T-1}\left\{\Gamma(k) - \Gamma(k_0)\right\}.$$

When $k \leq k_0$, $h \to 0$, we have

$$\begin{aligned}
\Gamma(k) - \Gamma(k_0) &= -\sum_{t=k+1}^{k_0} \frac{1}{\sigma^2} \left(Y_t - (1-\kappa h) Y_{t-1}\right) \delta h Y_{t-1} - \frac{1}{2} \sum_{t=k+1}^{k_0} \left(\frac{\delta}{\sigma} h Y_{t-1}\right)^2 \\
&= -\frac{\delta}{\sigma} \frac{1}{T} \sum_{t=k+1}^{k_0} Y_{t-1} e_t - \frac{1}{2} \left(\frac{\delta}{\sigma}\right)^2 \frac{1}{T^2} \sum_{t=k+1}^{k_0} Y_{t-1}^2 \\
&\Rightarrow \left(-\frac{\delta}{\sigma}\right) \int_{\tau}^{\tau_0} \left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa) y_0\right] dB(s) \\
&\quad -\frac{1}{2} \left(\frac{\delta}{\sigma}\right)^2 \int_{\tau}^{\tau_0} \left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa) y_0\right]^2 dr \\
&= \left(-\frac{\delta}{\sigma}\right) \left\{ \int_0^{\tau_0} \left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa) y_0\right] dB(s) \right. \\
&\quad \left. - \int_0^{\tau} \left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa) y_0\right] dB(s) \right\} \\
&\quad -\frac{1}{2} \left(\frac{\delta}{\sigma}\right)^2 \int_{\tau}^{\tau_0} \left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa) y_0\right]^2 dr \\
&= \left(\frac{\delta}{\sigma}\right) \left\{ \int_0^{\tau} \left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa) y_0\right] dB(s) \right. \\
&\quad \left. - \int_0^{\tau_0} \left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa) y_0\right] dB(s) \right\} \\
&\quad -\frac{1}{2} \left(\frac{\delta}{\sigma}\right)^2 \left| \int_{\tau_0}^{\tau} \left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa) y_0\right]^2 dr \right|,
\end{aligned}$$

where $B(\cdot)$ is a standard Brownian motion, $\widetilde{J}_{\kappa,\delta,\tau_0}(s)$ is a OU process defined in (4) and the convergence result comes from Lemma 4.1.

When $k > k_0$, $h \to 0$, we have

$$\begin{aligned}
\Gamma(k) - \Gamma(k_0) &= \sum_{t=k_0+1}^{k} \frac{1}{\sigma^2}\left(Y_t - (1-\kappa h) Y_{t-1}\right)\delta h Y_{t-1} + \frac{1}{2}\sum_{t=k_0+1}^{k}\left(\frac{\delta}{\sigma} h Y_{t-1}\right)^2 \\
&= \frac{1}{T}\sum_{t=k_0+1}^{k} \frac{1}{\sigma^2}\left(-\delta h Y_{(t-1)} + \sigma e_t\right)\delta Y_{t-1} + \frac{1}{2}\frac{1}{T^2}\sum_{t=k_0+1}^{k}\left(\frac{\delta}{\sigma}\right)^2 Y_{(t-1)}^2 \\
&= \frac{\delta}{\sigma}\frac{1}{T}\sum_{t=k_0+1}^{k} Y_{(t-1)} e_t - \frac{1}{2}\left(\frac{\delta}{\sigma}\right)^2 \frac{1}{T^2}\sum_{t=k_0+1}^{k} Y_{(t-1)}^2 \\
&\Rightarrow \left(\frac{\delta}{\sigma}\right)\int_{\tau_0}^{\tau}\left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa - (s-\tau_0)\delta 1_{[s>\tau_0]})y_0\right] dB(s) \\
&\quad -\frac{1}{2}\left(\frac{\delta}{\sigma}\right)^2 \int_{\tau_0}^{\tau}\left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa - (s-\tau_0)\delta 1_{[s>\tau_0]})y_0\right]^2 dr \\
&= \left(\frac{\delta}{\sigma}\right)\left\{\int_{0}^{\tau}\left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa - (s-\tau_0)\delta 1_{[s>\tau_0]})y_0\right] dB(s)\right. \\
&\quad \left. - \int_{0}^{\tau_0}\left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa - (s-\tau_0)\delta 1_{[s>\tau_0]})y_0\right] dB(s)\right\} \\
&\quad -\frac{1}{2}\left(\frac{\delta}{\sigma}\right)^2 \left|\int_{\tau_0}^{\tau}\left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa - (s-\tau_0)\delta 1_{[s>\tau_0]})y(0)\right]^2 ds\right|,
\end{aligned}$$

where the second equation comes from the fact that $Y_t = (1-(\kappa+\delta)h)\, Y_{t-1} + \sigma e_t$ for $t > k_0$ and and the convergence result comes from Lemma 4.1.

Applying the continuous mapping theorem to the argmax function leads to

$$\begin{aligned}
\hat{\tau}_{LS,T} &\xrightarrow{d} \arg\max_{\tau\in(0,1)}\left\{\int_{\tau_0}^{\tau}\left(\frac{\delta}{\sigma}\right)\left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa - (s-\tau_0)\delta 1_{[s>\tau_0]})y(0)\right] dB(s)\right. \\
&\quad \left. -\frac{1}{2}\left|\int_{\tau_0}^{\tau}\left(\frac{\delta}{\sigma}\right)^2\left[\sigma \widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa - (s-\tau_0)\delta 1_{[s>\tau_0]})y(0)\right]^2 ds\right|\right\} \\
&= \arg\max_{\tau\in(0,1)}\left\{\int_{\tau_0}^{\tau}\frac{1}{\delta}\left[\widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa - (s-\tau_0)\delta 1_{[s>\tau_0]})\frac{y(0)}{\sigma}\right] dB(s)\right. \\
&\quad \left. -\frac{1}{2}\left|\int_{\tau_0}^{\tau}\left[\widetilde{J}_{\kappa,\delta,\tau_0}(s) + \exp(-s\kappa - (s-\tau_0)\delta 1_{[s>\tau_0]})\frac{y(0)}{\sigma}\right]^2 ds\right|\right\}.
\end{aligned}$$

where we use the notation $\int_{\tau_0}^{\tau}(\cdot)dB(s)$ to denote $\left[\int_0^{\tau}(\cdot)dB(s) - \int_0^{\tau_0}(\cdot)dB(s)\right]$ .This gives the result in Theorem 4.1 immediately. For a rigorous treatment of the continuous mapping theorem for the argmax function, see Kim and Pollard (1990).

References

[1] Chong, T.T.L., 2001. Structural change in AR (1) models. Econometric Theory, 17(1), pp.87-155.

[2] Elliott, G . and Müller, U.K., 2007. Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics* 141(2), pp.1196 -1218.

[3] Jiang, L., Wang, X. H., and Yu, J., 2016. New Distribution Theory for the Estimation of Structural Break Point in Mean. *Working paper.*

[4] Liang, Y. L., Pang, T., Zhang, D., and Chong, T.T.L., 2014. Structural change in mildly integrated AR(1) models. *Working Paper.*

[5] Mankiw, N. G., Miron, J. A., 1986. The changing behavior of the term structure of interest rates. *Quarterly Journal of Economics* 101(2), 221–228.

[6] Mankiw, N. G., Miron, J. A. and Weil, D N., 1987. The adjustment of expectations to a change in regime: A study of the founding of the Federal Reserve. *American Economic Review* 77(3), 358–374.

[7] Pang, T., Zhang, D. and Chong, T.T.L., 2014. Asymptotic Inferences for an AR (1) Model with a Change Point: Stationary and Nearly Non-Stationary Cases. Journal of Time Series Analysis, 35(2), 133-150.

[8] Perron, P., 1991. A continuous time approximation to the unstable first order autoregressive processes: The case without an intercept. *Econometrica* 59, 211-236.

[9] Phillips, P.C.B., 1987a. Time series regression with a unit root. *Econometrica* 55, 277-301.

[10] Phillips, P.C.B., 1987b. Toward a unified asymptotic theory for autoregression. *Biometrika* 74, 533-547.

[11] Phillips, P.C.B., 1991. To criticize the critics: An objective Bayesian analysis of stochastic trends. Journal of Applied Econometrics, 6(4), 333-364.

[12] Phillips, P.C.B. and T. Magdalinos, 2007. Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136, 115-130.

[13] Phillips, P.C.B., Y. Wu and Yu, J., 2011. Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review* 52, 201-226.

[14] Phillips, P.C.B. and Yu, J., 2009. A Two-Stage Realized Volatility Approach to Estimation of Diffusion Processes with Discrete Data. *Journal of Econometrics* 150, 139-150.

[15] Phillips, P.C.B. and Yu, J., 2011. Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics* 2, 455-491.

[16] Sims, C.A., 1988. Bayesian skepticism on unit root econometrics. *Journal of Economic dynamics and Control*, 12(2), 463-474.

[17] Sims, C.A. and Uhlig, H., 1991. Understanding unit rooters: a helicopter tour. *Econometrica*, 59, 1591-1599.

[18] Yao, Y. C., 1987. Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. The Annals of Statistics 15, 1321-1328.

[19] Yu, J., 2014. Econometric Analysis of Continuous Time Models: A Survey of Peter Phillips' Work and Some New Results. *Econometric Theory* 30, 737-774.

[20] Zhou, Q., and Yu, J., 2015. Asymptotic Theory for Linear Diffusions under Alternative Sampling Schemes. *Economic Letters* 128, 1-5.