5-2017

In-fill asymptotic theory for structural break point in autoregression: A unified theory

Liang JIANG
Singapore Management University, ljiang@smu.edu.sg

Xiaohu WANG
Chinese University of Hong Kong

Jun YU
Singapore Management University, yujun@smu.edu.sg

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

Liang Jiang
Singapore Management University

Xiaohu Wang
The Chinese University of Hong Kong

Jun Yu
Singapore Management University

May 17, 2017

Abstract

This paper obtains the exact distribution of the maximum likelihood estimator of structural break point in the Ornstein–Uhlenbeck process when a continuous record is available. The exact distribution is asymmetric, tri-modal, dependent on the initial condition. These three properties are also found in the finite sample distribution of the least squares (LS) estimator of structural break point in autoregressive (AR) models. Motivated by these observations, the paper then develops an in-fill asymptotic theory for the LS estimator of structural break point in the AR(1) coefficient. The in-fill asymptotic distribution is also asymmetric, tri-modal, dependent on the initial condition, and delivers excellent approximations to the finite sample distribution. Unlike the long-span asymptotic theory, which depends on the underlying AR root and hence is tailor-made but is only available in a rather limited number of cases, the in-fill asymptotic theory is continuous in the underlying roots. Monte Carlo studies show that the in-fill asymptotic theory performs better than the long-span asymptotic theory for cases where the long-span theory is available and performs very well for cases where no long-span theory is available.

JEL classification: C11; C46
Keywords: Asymmetry; Bias; Exact distribution; Long-span asymptotics; In-fill asymptotics; Trimodality.

*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, Yichong Zhang, and seminar participants at Berkeley, Princeton, Singapore Management University and University of California, San Diego.

1 Introduction

Autoregressive (AR) models with a structural break in the AR(1) coefficient 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 break point is often linked to a significant economic event or an important economic policy. Not surprisingly, making statistical inference about the structural break point has received a great deal of attention from both econometricians and empirical economists when they are confronted with economic time series.

Existing asymptotic theory assumes that the time spans, before and after the structural break point, both go to infinity; see Chong (2001), Pang, Zhang, and Chong (2014) and Liang et al. (2017) for the development of these asymptotic distributions. Unfortunately, the resulting long-span asymptotic theory makes statistical inference about the structural break point very complicated for a number of reasons.

First, depending on the values of the AR(1) coefficients before and after the break point, the process in each regime can have a stationary, or a mildly stationary, or a local-to-unit, or a unit, or a mildly explosive, or an explosive root. The asymptotic theory developed in the literature was tailor-made to accommodate different combinations of two roots, but so far only covers a very small number of cases. In many empirically interesting examples, including that considered in Phillips, Wu, and Yu (2011) and Phillips and Yu (2011), no asymptotic theory is available.

Second, to aggravate the matter, the derived asymptotic distribution often does not perform well in finite sample. It is discontinuous in the underlying AR(1) parameters. In particular, the long-span asymptotic distribution and, sometimes even, the rate of convergence depend on how one classifies the two AR roots, although no guidance is given about the classification.[1] Moreover, the long-span asymptotic distribution does not depend on the initial condition. However, the finite sample distribution of 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 break point estimator. Furthermore, the finite sample distribution of break point estimator depends on the initial condition. These two facts explain why the long-span asymptotic theory can perform poorly in finite sample. Evidence from the simulations reported later strongly suggests that in many empirically relevant cases the long-span asymptotic

[1] For example, if the AR(1) coefficient is 0.9, should it be classified as a stationary, or a mildly stationary, or a local-to-unit root? Different classification leads to different asymptotic distribution.

theory is inadequate.

The discontinuity in the long-span limiting distributions is also found in the AR(1) model without break. In a recent attempt, Phillips and Magdalinos (2007) developed the long-span limiting distributions when the root is moderately deviated from unity. They show that the rate of convergence in their 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.[2]

Interestingly, when a continuous record of observations is available, continuous time models can provide the exact distribution of persistency parameter, as shown in Phillips (1987a, 1987b). The exact distribution is continuous in the persistency parameter, regardless of its sign and value. This feature motivated Phillips (1987a) and Perron (1991) to establish the in-fill asymptotic distribution for the AR(1) parameter in discrete time models. It also motivates Yu (2014) and Zhou and Yu (2016) to establish the in-fill asymptotic distribution for the persistency parameter in continuous time models. Not surprisingly, these in-fill asymptotic distributions are continuous in the underlying parameters and dependent on the initial condition.

In this paper, we develop an in-fill asymptotic distribution of break point estimator in time series models with a break in the AR(1) coefficient. The in-fill asymptotic distribution is continuous in the two underlying AR parameters. Hence, it offers a unified framework for making statistical inference about the break point. Moreover, it 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 maximum likelihood (ML) estimator of break point via the Girsanov theorem. The exact distribution is applicable to all values for two persistency parameters. It is continuous in two persistency parameters, regardless of their signs and values, and is dependent on the initial condition.

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,

[2]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.

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 propose an AR model with a break in the AR coefficient and derive the in-fill asymptotic distribution for the break point. Our model converges to the OU process with a break as the sampling interval shrinks. To develop our in-fill theory, we do not need to restrict any of the AR coefficients to be less than one, or equal to one, or greater than one. Furthermore, our model enables us to compare the magnitude of the break size and 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 extend our limit theory to a more general time series model where the AR(1) coefficient has a break but the error term is weakly dependent. The assumption of an independent error term has been imposed in the literature to develop the long-span asymptotic theory. Since the assumption can be too strong for empirical work, it is important to relax the assumption.

Finally, we carry out extensive simulation studies, checking the performance of the in-fill asymptotic distribution against the long-span counterpart developed in the literature for cases where the long-span theory is available. Our results show that our unified in-fill asymptotic distribution always performs better than the long-span counterpart although the later was tailor-made to accommodate different kinds of regime shift. We also investigate the performance of the in-fill asymptotic distribution for cases where the long-span theory is not available. Our results show that our in-fill asymptotic distribution continues to perform well.

There are several drawbacks in our in-fill asymptotic theory, however. First, under the in-fill asymptotic scheme, our estimator of break point is inconsistent. However, our estimator is the same as that under the long-span scheme. Hence, our in-fill scheme can be understood as a vehicle of obtaining a better approximation than the long-span scheme. Second, the asymptotic distribution is not pivotal. Third, the distribution is non-standard and the density function is not available analytically. Hence, simulations are needed to obtain critical values, as in most of the long-span asymptotic distributions. Fourth, while the features of asymmetry and trimodality in the in-fill distribution are shared by the finite sample distribution, they make the construction of confidence

intervals more difficult.

The rest of the paper is organized as follows. Section 2 reviews the literature on AR(1) models with a break. Special focus is paid to the assumptions about the two AR(1) coefficients as well as to the assumptions about the break size. Section 3 develops the exact distribution of the ML estimator of break point in the OU process 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. Section 5 develops the in-fill asymptotic theory for the LS estimator of the break point in a general time series model. In Section 6, we provide simulation results and check the finite sample performance of the in-fill theory. Section 7 concludes. Appendix A gives a detailed literature review and Appendix B collects all the proofs of the theoretical result.

2 A Literature Review and Motivations

The literature on the structural break model is too extensive to review. Among the contributions in the literature, Chong (2001), Pang, Zhang, and Chong (2014) and Liang et al. (2017) focused on the AR(1) model with a break in the root. Under different 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 break point.

The model considered in these papers is

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

where T denotes the sample size, ε_t is a sequence of independent and identically distributed (i.i.d.) random variables. Let k denote the break point parameter with the true value k_0. The condition $1 \leq k_0 < T$ is assumed to ensure that one break happens. The fractional break point parameter is defined as $\tau = k/T$ with the true value $\tau_0 = k_0/T$. Clearly $\tau_0 \in (0, 1)$. 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

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

Table 1: The long-span asymptotic distributions of $\hat{\tau}_{LS,T} - \tau_0$ under different settings of the AR roots before and after the break. $W(u)$ is a two-sided Brownian motion, whose definition and the meanings of other notations are introduced in Appendix A.

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

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 estimates of β_1 and β_2 for any fixed k. The corresponding estimator of τ is $\hat{\tau}_{LS,T} = \hat{k}_{LS,T}/T$.

As it is well-known in the literature, there are seven possible cases for the root of an AR model, and the asymptotic properties of the AR model crucially depend on which case its root is in. Let $c > 0$ be a positive constant, $\alpha \in (0, 1)$, and β denote the AR root. When β is a constant and with modulus smaller than one (i.e. $|\beta| < 1$) the AR model is a stationary process. When $\beta = 1 - \frac{c}{T^\alpha}$, it becomes a mildly stationary process. When $\beta = 1 - \frac{c}{T}$, it is a left-side local-to-unity process. When $\beta = 1$, it is a random walk. When $\beta = 1 + \frac{c}{T}$, it is a right-side local-to-unity process. When $\beta = 1 + \frac{c}{T^\alpha}$, it is a mildly explosive process. When $\beta > 1$ is a constant, it is an explosive process. Under different settings of the AR roots before and after the break (β_1 and β_2, respectively), Chong (2001), Pang, Zhang, and Chong (2014) and Liang et al. (2017) established the consistency of $\hat{\tau}_{LS,T}$ and derived the long-span asymptotic distributions of $\hat{\tau}_{LS,T} - \tau_0$ as $T \to \infty$. In Table 1 we give a brief summary of the developed long-span asymptotic distributions and the rate of convergence together with the assumptions on AR roots, the order of break size and the initial value. 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 positive number. A detailed review of the long-span asymptotics is in Appendix A.

Several observations can be made from Table 1 which motivates the paper. First, except for the seven cases reported in Table 1, the long-span asymptotic theory remains

unknown for many cases that are interesting from practical viewpoints. For example, the AR process changes from a random walk to a mildly explosive process, a case widely studied in the bubble testing literature. Other interesting cases include (1) $|\beta_1| < 1$ and $\beta_2 = 1 \pm c/T^\alpha$; (2) $|\beta_1| < 1$ and $\beta_2 > 1$; (3) $\beta_1 = 1 \pm c/T^\alpha$ and $\beta_2 < 1$; (4) $\beta_1 = 1 \pm c_1/T^\alpha$ and $\beta_2 < 1 \pm c_2/T$; (5) $\beta_1 = 1 \pm c_1/T^{\alpha_1}$ and $\beta_2 = 1 \pm c_2/T^{\alpha_2}$; (6) $\beta_1 = 1 \pm c_1/T^{\alpha_1}$ and $\beta_2 > 1$; (7) $\beta_1 = 1 \pm c_1/T$ and $\beta_2 = 1 \pm c_2/T$; (8) $\beta_1 = 1 \pm c_1/T$ and $\beta_2 = 1 \pm c_2/T^\alpha$; (9) $\beta_1 = 1 \pm c_1/T$ and $\beta_2 = 1$; (10) $\beta_1 = 1 \pm c_1/T$ and $\beta_2 > 1$; (11) $\beta_1 = 1$ and $\beta_2 = 1 \pm c_1/T$; (12) $\beta_1 = 1$ and $\beta_2 > 1$; (13) $\beta_1 > 1$ and $|\beta_2| < 1$; (14) $\beta_1 > 1$ and $\beta_2 < 1 \pm c/T$; (15) $\beta_1 > 1$ and $\beta_2 < 1 \pm c/T^\alpha$; (16) $\beta_1 > 1$ and $\beta_2 = 1$; (17) $\beta_1 > 1$ and $\beta_2 > 1$; (18) $\beta_1 = 1 + c/T^\alpha$ and $\beta_2 = 1$, here c, c_1 and c_2 are positive constants, and $\alpha, \alpha_1, \alpha_2 \in (0, 1)$.

Second, Table 1 shows that the long-span asymptotic theory is discontinuous in β_1 and β_2 when one of them passes the unity. Both the expression of limiting distribution and the rate of convergence crucially depend on the distance and the direction of the AR roots away from unity. On the other hand, the finite sample distribution is always continuous in the underlying AR roots. This feature of discontinuity causes a great deal of difficulties in making statistical inference about the break point in practice. This is because users typically do not know *ex ante* the values of β_1 and β_2. Consequently, they do not have any clue about how far and in which direction β_1 and β_2 are away from unity. Furthermore, even if the values of the AR roots on both sides of the break are known *ex ante*, it is still unclear which asymptotic distribution reported in Table 1 should to used. For example, if it is known for sure that the AR root changes its value from 0.5 to 0.9, should we use the large sample theory reported in the second row of Table 1 where the AR(1) model changes from a stationary process to another stationary process, or should we use the large sample theory reported in the fifth row of Table 1 where the AR(1) model changes from a stationary process to a local-to-unity process?

Third, all the long-span asymptotic distributions reported in Table 1 are invariant to the value of initial condition y_0. However, it is well-known in the nonstationary time series literature that the finite sample distribution of the LS estimate of AR root can be very sensitive to the value of y_0; see, for example, Evans and Savin (1981) and Perron (1991) for local-to-unity models, and Wang and Yu (2016) for mildly explosive processes. Hence, it is reasonable to expect that the finite sample distribution of $\hat{k}_{LS,T}$ as defined in (1) should also depend on the value of y_0, especially for the case when the AR root on either side of the break is close to or mildly greater than one. The simulation results that will be reported in Section 6 confirm this expectation.

Fourth, the property of finite sample bias in the estimation of break point has not been discussed in the literature with the only exception in LWY (2016). Given that bias exists in the estimation of AR(1) coefficients, we expect the bias to exist in the estimation of break point. In fact, there are two sources for the bias. The first one lies in the asymmetry of the two time spans. As long as $\tau_0 \neq 1/2$, the time spans and hence the numbers of observations are not equivalent in the two regimes. The second source lies in the fact that the variance of the AR process changes after the break happens. However, as shown by the red dashed line in Figure 1, the long-span asymptotic distribution reported in the second row of Table 1 is symmetric about zero, suggesting no bias in $\hat{\tau}_{LS,T}$. The long-span asymptotic scheme requires the two time spans diverge to infinity, and hence the asymmetry in the sample information in the two regimes disappears in the limit.

Finally, except for the asymptotic distribution in the second row of Table 1 where the density function was derived analytically in Yao (1987), the density function of any other distribution in Table 1 does not have a closed-form expression. Simulation methods are required to obtain the densities and quantiles. Unfortunately, the interval to find the argmax is always $(-\infty, \infty)$ in these distributions, rendering simulation methods computationally expensive. This is because, to well approximate the true argmax, one must numerically calculate the argmax over an sufficiently wide interval and choose a very fine grid, leading to a very large number of grid points and a high computational cost.

Besides the five observations discussed above, it is also worthwhile to point out that the developed long-span asymptotic distributions may not perform well in finite sample in many empirically relevant cases. For example, consider the case where the AR root switches from a stationary root to another. The blue line in Figure 1 plots the finite sample density of $\hat{\tau}_{LS,T}$, centered at the true value and normalized by the convergence rate, i.e., $\frac{T(\beta_2-\beta_1)^2}{1-\beta_1^2}(\hat{\tau}_{LS,T} - \tau_0)$, when $\tau_0 = 1/2$, $T = 200$, the AR root changes from $\beta_1 = 0.5$ to $\beta_2 = 0.61$. The finite sample distribution is obtained from simulated data with 100,000 replications. The red broken line in Figure 1 plots the density of the long-span asymptotic distribution. In two aspects the finite sample distribution is notably different from the long-span asymptotic distribution ($\arg\max_{u\in(-\infty,\infty)} \left\{W(u) - \frac{|u|}{2}\right\}$).[3] First, the finite sample distribution is asymmetric, indicating an upward bias in the estimate of the break point, whereas the long-span asymptotic distribution is symmetric. Second, the finite sample distribution has three modes with one at the origin and others at the

[3]See Yao (1987) and Bai (1994) for further properties about $\arg\max_{u\in(-\infty,\infty)} \left\{W(u) - \frac{|u|}{2}\right\}$.

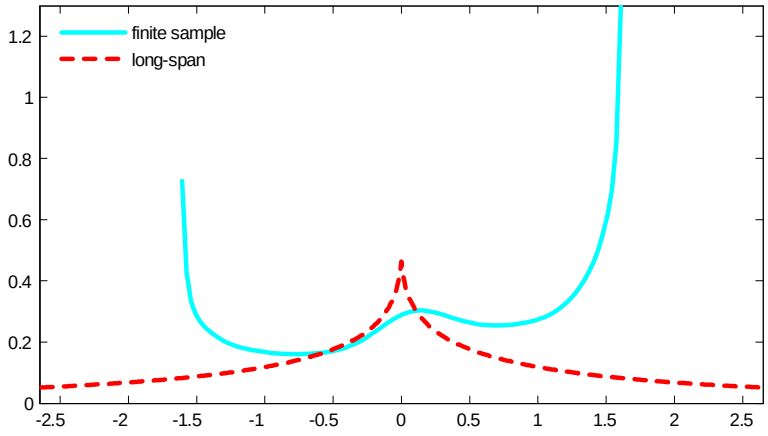

Figure 1: The pdf of the finite sample distribution of $\frac{T(\beta_2-\beta_1)^2}{1-\beta_1^2}(\hat{\tau}_{LS,T}-\tau_0)$ when $T=200$, $\beta_1=0.5, \beta_2=0.61$, $\sigma=1$ and $\tau_0=0.5$ in Model (1) and the pdf of $\arg\max_{u\in(-\infty,\infty)}\left\{W(u)-\frac{1}{2}|u|\right\}$.

two boundary points of the support, whereas the long-span asymptotic distribution has a unique mode. The trimodality has important implications for statistical inference. For example, the confidence interval may contain two or three disjoined intervals. The asymmetry and trimodality in finite sample distribution can also be found in Figure 7(c) of Chong (2001).

The five concerns about the long-span asymptotic distributions reported in Table 1 and the large discrepancy between the long-span asymptotic distribution and the finite sample distribution motivate us to introduce an alternative asymptotic theory to approximate the finite sample distribution of break point.

3 A Continuous Time Model

In this section we study a continuous time OU process with a break in the drift function:

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

where $t \in [0,1]$,[4] $1_{[t>\tau_0]}$ is an indicator function, κ, δ and τ_0 are constants with $\tau_0 \in (0,1)$ being the break point and δ being the break size, the constant σ measures the noise level, and $B(t)$ denotes a standard Brownian motion. The initial condition is assumed

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

to be $x(0) = O_p(1)$. The time span is τ_0 in the first regime while it is $1 - \tau_0$ in the second regime.

We assume that a continuous record of observations, $\{x(t)\}$ for $t \in [0, 1]$, is available and that all parameters except τ_0 are known. With a continuous record, a more complicated assumption about the diffusion function, such as $\sigma = \sigma(x(t))$, will not cause any change in the analysis developed below. This is because the diffusion function can be estimated by the quadratic variation without estimation error.

There are four reasons for studying a continuous time model. First, it provides a natural choice to study the effect of the difference in the two time spans. As is well-known in the continuous time literature, properties of estimators of persistency parameter depend crucially on the time span; see, for example, Tang and Chen (2009) and Yu (2012). As a result, we expect properties of estimators of break point depend crucially on the difference in the time spans. Second, as it becomes clear later, the exact distribution of the ML estimator of the break point $\hat{\tau}_{ML}$ defined in (4) is a continuous function of both persistency parameters. This property sheds light on how we will address the discontinuity problem of the long-span asymptotic distributions reported in Table 1. Third, explicit effect of the initial condition can be found in the exact distribution of $\hat{\tau}_{ML}$. Finally, the continuous time model provides a benchmark for us to set up a discrete time AR model with a break in AR roots under which the in-fill asymptotic scheme is considered.

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

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

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

$$\hat{\tau}_{ML} = \arg \max_{\tau \in (0,1)} \log \mathcal{L}(\tau). \tag{4}$$

It is difficult to find the pdf and the cdf of $\hat{\tau}_{ML}$ by analytical methods or numerical methods. To facilitate the approximation of the density function via simulations and to better examine properties of the density, Theorem 3.1 gives an equivalent representation of $\hat{\tau}_{ML}$.

Theorem 3.1 *Consider Model (3) with a continuous record being available. The ML*

estimator $\hat{\tau}_{ML}$ defined in (4) has the exact distribution as

$$\hat{\tau}_{ML} \overset{d}{=} \arg\max_{\tau\in(0,1)} \delta\left\{\left[\widetilde{J}_{\tau_0}(\tau)\right]^2 - \tau + (2\kappa+\delta)\int_0^{\tau}\left[\widetilde{J}_{\tau_0}(r)\right]^2 dr\right\}, \tag{5}$$

where $\widetilde{J}_{\tau_0}(r)$, for $r \in [0,1]$, is a Gaussian process defined by

$$d\widetilde{J}_{\tau_0}(r) = -\left(\kappa + \delta 1_{[r>\tau_0]}\right)\widetilde{J}_{\tau_0}(r)dt + dB(r), \tag{6}$$

with the initial condition $\widetilde{J}_{\tau_0}(0) = x(0)/\sigma$, and $B(r)$ is a standard Brownian motion which is the same as in (3).

Remark 3.1 *The exact distribution given in (5) depends on κ and δ which describe the drift function of the OU process in (3). As $\widetilde{J}_{\tau_0}(r)$ is a continuous function of κ and δ, the exact distribution given in (5) should also be continuous in κ and δ. While it would be useful to have an analytical proof of continuity of the exact distribution in κ and δ, without knowing the pdf of $\hat{\tau}_{ML}$ in closed-form, such a proof is not easy to obtain. Moreover, the exact distribution explicitly depends on $x(0)/\sigma$ through the process $\widetilde{J}_{\tau_0}(r)$.*

Remark 3.2 *From the exact distribution (5), an alternative expression can be derived:*

$$\begin{aligned}
&\hat{\tau}_{ML} - \tau_0 \\
&\overset{d}{=} \arg\max_{u\in(-\tau_0,1-\tau_0)} \begin{cases} \delta\left\{\left[\widetilde{J}_{\tau_0}(\tau_0+u)\right]^2 - u - (2\kappa+\delta)\int_{\tau_0+u}^{\tau_0}\left[\widetilde{J}_{\tau_0}(r)\right]^2 dr\right\} & \text{for } u \le 0 \\ \delta\left\{\left[\widetilde{J}_{\tau_0}(\tau_0+u)\right]^2 - u + (2\kappa+\delta)\int_{\tau_0}^{\tau_0+u}\left[\widetilde{J}_{\tau_0}(r)\right]^2 dr\right\} & \text{for } u > 0 \end{cases}
\end{aligned} \tag{7}$$

It is easier to understand why $\hat{\tau}_{ML}$ is asymmetrically distributed around the true value τ_0 and the bias in $\hat{\tau}_{ML}$ from (7). One reason is that the interval $(-\tau_0, 1-\tau_0)$ is not symmetric about zero as long as $\tau_0 \neq 1/2$. This asymmetry comes from the fact that the two time spans are different in the model. The second reason is that the two objective functions in the argmax are different in (7). The asymmetry in the objective functions comes from the asymmetry of $\widetilde{J}_{\tau_0}(r)$ before and after the break. The second reason suggests that the bias in $\hat{\tau}_{ML}$ is still expected even when $\tau_0 = 1/2$.

Remark 3.3 *To understand why $\hat{\tau}_{ML}$ has three modes, denote*

$$Z_1(u) = \delta\left\{\left[\widetilde{J}_{\tau_0}(\tau_0+u)\right]^2 - u - (2\kappa+\delta)\int_{\tau_0+u}^{\tau_0}\left[\widetilde{J}_{\tau_0}(r)\right]^2 dr\right\} \text{ for } u \le 0,$$

$$Z_2(u) = \delta\left\{\left[\widetilde{J}_{\tau_0}(\tau_0+u)\right]^2 - u + (2\kappa+\delta)\int_{\tau_0}^{\tau_0+u}\left[\widetilde{J}_{\tau_0}(r)\right]^2 dr\right\} \text{ for } u > 0.$$

It is easy to show that $\widetilde{J}_{\tau_0}(\tau_0+u) \sim N\left(0, \frac{1-e^{-2\kappa(\tau_0+u)}}{2\kappa}\right)$ *for* $u \in (-\tau_0, 0]$ *and that* $\widetilde{J}_{\tau_0}(r) \sim N\left(0, \frac{1-e^{-2\kappa r}}{2\kappa}\right)$ *for* $r \in [\tau_0+u, \tau_0]$ *with* $u \in (-\tau_0, 0]$. *Hence,*

$$\begin{aligned} E\left(Z_1(u)\right) &= \delta\left\{\frac{1-e^{-2\kappa(\tau_0+u)}}{2\kappa} - u - (2\kappa+\delta)\int_{\tau_0+u}^{\tau_0}\frac{1-e^{-2\kappa r}}{2\kappa}dr\right\} \\ &= \delta\left\{\frac{1+u\delta}{2\kappa} - \frac{e^{-2\kappa\tau_0}}{2\kappa} - \frac{\delta e^{-2\kappa\tau_0}\left(1-e^{-2\kappa u}\right)}{(2\kappa)^2}\right\}. \end{aligned}$$

Taking the derivative of $E\left(Z_1(u)\right)$ *with respect to* u, *we have*

$$\frac{\partial E\left(Z_1(u)\right)}{\partial u} = \frac{\delta^2}{2\kappa}\left(1-e^{-2\kappa(\tau_0+u)}\right) > 0 \quad for \quad u \in (-\tau_0, 0],$$

suggesting that on average $Z_1(u)$ *has the unique maximum at the origin. Similarly,* $E\left(Z_2(u)\right)$ *has a supremum at the origin. This property is similar to that of* $E\left(W(u) - |u|/2\right)$, *as explained in JWY (2016). That the expectation of the objection function in (7) is maximized at the origin explains why the origin is a mode in* $\hat{\tau}_{ML}$. *If the interval to find the argmax is* $(-\infty, \infty)$, *we would not expect any other mode in* $\hat{\tau}_{ML}$, *as in the long-span asymptotic distributions. However, the interval for the argmax in (7) is bounded with two boundary points,* $-\tau_0$ *and* $1-\tau_0$. *In an argument similar to that in JWY (2016), there are two modes at the boundary points in the distribution of* $\hat{\tau}_{ML}$.

In Figure 2, we plot the density of $\widehat{\tau}_{ML} - \tau_0$ given in (7) with $\kappa = 138$, $\delta = -20$, $\sigma = 1$, $\tau_0 = 0.3, 0.5, 0.7$, respectively. The blue solid line corresponds to the density when $x(0) = 0.2$, and the black broken line corresponds to the density when $x(0) = 1$. The densities are obtained from 100,000 replications.

The simulation results in Figure 2 support the remarks made above. First, the density is sensitive to $x(0)/\sigma$. Second, all the densities are asymmetric, indicating that $\hat{\tau}_{ML}$ is a biased estimator even when $\tau_0 = 1/2$. Moreover, as τ_0 varies, both the level and the direction of asymmetry of density may change. Third, trimodality is found in the density for all cases with $0 (= \widehat{\tau}_{ML} - \tau_0)$ being one mode and the two boundary points being the other two.

4 A Discrete Time Model and In-fill Asymptotic Distribution

Motivated by the findings in the continuous time model, in this section we propose a discrete time model that is closely related to the continuous OU process (3). The

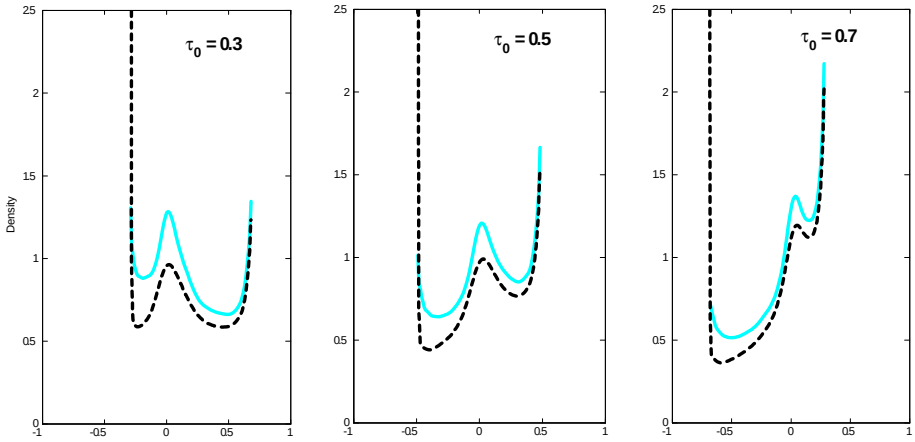

Figure 2: Densities of $\widehat{\tau}_{ML} - \tau_0$ given in Equation (7) when $\kappa = 138, \delta = -20$, $\sigma = 1$ and $\tau_0 = 0.3, 0.5, 0.7$, respectively. Solid lines are densities for $x(0) = 0.2$; broken lines are densities for $x(0) = 1$.

discrete time model has the form of

$$x_t = \left(\beta_1 1_{[t\leq k_0]} + \beta_2 1_{[t>k_0]}\right) x_{t-1} + \sqrt{h}\varepsilon_t, \;\; \varepsilon_t \overset{i.i.d.}{\sim} (0, \sigma^2), \;\; x_0 = O_p(1) \tag{8}$$

where $\beta_1 = \exp\{-\kappa/T\}$ and $\beta_2 = \exp\{-(\kappa+\delta)/T\}$ are the AR roots before and after the break, k_0 denotes the break point, $t = 1, ..., T$ with T being the sample size, and $h = 1/T$.[5] The fractional break point is defined as $\tau_0 = k_0/T$.

If $\tau_0/h = T\tau_0 = k_0$ is an integer, the exact discretization of Model (3) over the interval $[0, 1]$ with the sampling interval h is given by

$$x_{th} = \left(\beta_1 1_{[th\leq\tau_0]} + \beta_2 1_{[th>\tau_0]}\right) x_{(t-1)h} + \sqrt{\frac{1 - e^{-2\left(\kappa+\delta 1_{[th>\tau_0]}\right)h}}{2\left(\kappa + \delta 1_{[th>\tau_0]}\right)}}\varepsilon_t, \;\; \varepsilon_t \overset{i.i.d.}{\sim} N(0, \sigma^2), \tag{9}$$

where $t = 1, ..., T$ and $x_0 = x(0) = O_p(1)$. The proposed discrete time model in (8) is nearly the same as the exact discretization given in (9) with two small differences. First, in Model (8) we relax the normality assumption on the errors. This generalization is important as in many empirical applications, the normality assumption is too strong. Second, the variances of the errors are different. However, since

$$\frac{1 - \exp\{-2\kappa h\}}{2\kappa} = h + O\left(h^2\right) \quad \text{and} \quad \frac{1 - \exp\{-2(\kappa+\delta)h\}}{2(\kappa+\delta)} = h + O\left(h^2\right),$$

if $h \to 0$, the two sets of the variance are asymptotically the same.

[5] An implicit assumption we make here is that $1/h$ is an integer.

The LS estimator of the break point in Model (8) takes the form of

$$\hat{k}_{LS} = \arg \min_{k=1,\cdots,T-1} \sum_{t=1}^{k} \left(x_t - \hat{\beta}_1 (k) x_{t-1} \right)^2 + \sum_{t=k+1}^{T} \left(x_t - \hat{\beta}_2 (k) x_{t-1} \right)^2 \qquad (10)$$

where $\hat{\beta}_1 (k) = \sum_{t=1}^{k} x_t x_{t-1} / \sum_{t=1}^{k} x_{t-1}^2$ and $\hat{\beta}_2 (k) = \sum_{t=k+1}^{T} x_t x_{t-1} / \sum_{t=k+1}^{T} x_{t-1}^2$ are LS estimates of β_1 and β_2, respectively. The LS estimator of the fractional break point is defined as

$$\hat{\tau}_{LS} = \hat{k}_{LS} / T \ . \qquad (11)$$

The connection between the proposed discrete time model (8) and the exact discrete time model (9) and hence the continuous OU process (3) motivates us to study the in-fill asymptotic theory. In particular, if we allow $h \to 0$ (which increases the sample size T), the discrete observations form a continuous record in the limit and the proposed discrete time model (8) converges to the continuous OU process (3). Therefore, it is expected that, the in-fill asymptotic distribution will converge to the exact distribution developed under the assumption of a continuous record.

Before reporting the in-fill asymptotic distribution of $\hat{\tau}_{LS}$, it is worth comparing the proposed discrete time model (8) with the the discrete time models considered in the literature. While the order of errors is $O_p\left(\sqrt{h}\right)$ in our model, it is $O_p(1)$ in the models considered in the literature. To facilitate such a comparison, we divide both sides of Model (8) by $\sqrt{h}$ and denote $y_t = x_t / \sqrt{h}$. Then, we have, for $t = 1, ..., T$,

$$y_t = \left(\beta_1 1_{[t \leq k_0]} + \beta_2 1_{[t > k_0]} \right) y_{t-1} + \varepsilon_t, \ \varepsilon_t \overset{i.i.d.}{\sim} (0, \sigma^2) \ , \ y_0 = x_0 / \sqrt{h} = O_p\left(T^{1/2}\right) . \qquad (12)$$

Model (12) is almost the same as the model in (1) except for three important differences. First, the initial condition of y_t in (12) diverges at the rate of $T^{1/2}$ as $T \to \infty$, whereas the initial condition in Model (1) is set to be $o_p\left(T^{1/2}\right)$ as shown in Table 1. This difference explains why the in-fill asymptotic distribution of $\hat{\tau}_{LS}$ explicitly depends on the initial value x_0.

Second, in Model (12), $\beta_1 = \exp\{-\kappa/T\} \to 1$ and $\beta_2 = \exp\{-(\kappa + \delta)/T\} \to 1$ as $T \to \infty$. Whereas, for model in (1), β_1 and β_2 are allowed to be further away from one. It looks as if the in-fill asymptotic theory for Model (12) only works for the case where the AR roots in both regimes are in a small vicinity of unity. However, our simulation results show that the in-fill theory works well even when β_1 and\or β_2 are distant from unity in finite sample.

The third difference lies in the order of break size. The break size is $\beta_2 - \beta_1 = O\left(T^{-1}\right)$ in Model (12) while it is $O\left(T^{-\alpha}\right)$ with $0 < \alpha < 1$ in Model (1); see Table 1. Clearly

under the in-fill scheme we assume a smaller break size. Interestingly, in the context of time series regression with a break in the slope coefficient, the break size is usually set to $O\left(T^{-\alpha}\right)$ with $0 < \alpha < 1/2$; see, for example, Bai (1994, 1997). Elliott and Müller (2007) argued that such a break size may be empirically too large. They introduced a regression model with the break size reducing to zero at the rate of $O\left(T^{-1/2}\right)$. JWY (2016) provided evidence that, when the break size is $O\left(T^{-1/2}\right)$, the asymptotic distribution is closer to the finite sample distribution. The present paper extends the argument of Elliott and Müller to the AR models. The smaller break size is important to produce asymmetry and trimodality in our asymptotic distribution and to explain why our asymptotic distribution performs better than the asymptotic distributions summarized in Table 1.

Theorem 4.1 *Consider the discrete time model in (8). When $T \to \infty$ with a fixed τ_0, the in-fill asymptotic distribution of the estimator $\hat{\tau}_{LS} = \hat{k}_{LS}/T$ with $\hat{k}_{LS}$ defined in (10) is*

$$\hat{\tau}_{LS} \Rightarrow \arg\max_{\tau \in (0,1)} \frac{\left\{ \left[\widetilde{J}_{\tau_0}(\tau) \right]^2 - \left[\widetilde{J}_{\tau_0}(0) \right]^2 - \tau \right\}^2}{\int_0^\tau \left[\widetilde{J}_{\tau_0}(r) \right]^2 dr} + \frac{\left\{ \left[\widetilde{J}_{\tau_0}(1) \right]^2 - \left[\widetilde{J}_{\tau_0}(\tau) \right]^2 - [1-\tau] \right\}^2}{\int_\tau^1 \left[\widetilde{J}_{\tau_0}(r) \right]^2 dr} \tag{13}$$

where $\widetilde{J}_{\tau_0}(r)$, for $r \in [0,1]$, is the Gaussian process defined in (6) with the initial condition $\widetilde{J}_{\tau_0}(0) = x_0/\sigma$, and $\Rightarrow$ denotes weak convergence.

Remark 4.1 *When deriving the exact distribution for Model (3), we assumed that two persistency parameters are known. In Model (8), both β_1 and β_2 are assumed unknown and are estimated. That explains why the in-fill asymptotic distribution in (13) is different from the exact distribution in (5). If β_1 and β_2 in (8) are known, then the corresponding in-fill asymptotic distribution will be the same as the exact distribution in (5).*

Remark 4.2 *Through the Gaussian process $\widetilde{J}_{\tau_0}(r)$, the in-fill asymptotic distribution given in (13) explicitly depends on the initial condition x_0/σ. Moreover, it also depends on the persistency parameters κ and δ. Since $\widetilde{J}_{\tau_0}(r)$ is continuous in κ and δ, the in-fill asymptotic distribution in (13) should also be continuous in κ and δ.*

Remark 4.3 *Let $\tau = \tau_0 + u$. An equivalent representation of the in-fill asymptotic*

distribution is:

$$\hat{\tau}_{LS} - \tau_0$$
$$\Rightarrow \quad \arg \max_{u \in (-\tau_0, 1-\tau_0)} \frac{\left\{ \left[\widetilde{J}_{\tau_0}(\tau) \right]^2 - \left[\widetilde{J}_{\tau_0}(0) \right]^2 - \tau \right\}^2}{\int_0^\tau \left[\widetilde{J}_{\tau_0}(r) \right]^2 dr} + \frac{\left\{ \left[\widetilde{J}_{\tau_0}(1) \right]^2 - \left[\widetilde{J}_{\tau_0}(\tau) \right]^2 - [1-\tau] \right\}^2}{\int_\tau^1 \left[\widetilde{J}_{\tau_0}(r) \right]^2 dr}.$$

As in the exact distribution, both the asymmetry in $(-\tau_0, 1-\tau_0)$ and the asymmetry in $\widetilde{J}_{\tau_0}(r)$ at different sides of τ_0 contribute to the asymmetry of the in-fill asymptotic distribution. Hence, the in-fill distribution is asymmetric for all τ_0 even when $\tau_0 = 1/2$, suggesting that $\hat{\tau}_{LS}$ is generally biased.

Remark 4.4 *Although it is much harder to obtain the expectation of the objective function in this case, we still expect trimodality in $\hat{\tau}_{LS} - \tau_0$ for the same reason as before, namely, the origin is the unique maximum of the expectation of the objective function and the maximization is done over a finite interval $(-\tau_0, 1-\tau_0)$, not the infinite interval $(-\infty, \infty)$. The conjecture of asymmetry and trimodality will be confirmed in simulations, which also show that the in-fill asymptotic distribution performs very well in approximating the finite sample distributions.*

5 A General Model

In this section we extend the in-fill asymptotic theory to a general discrete-time model with weakly dependent errors:

$$x_t = \left(\beta_1 1_{[t \le k_0]} + \beta_2 1_{[t > k_0]} \right) x_{t-1} + u_t, \quad x_0 = O_p(1), \tag{14}$$

where $\beta_1 = \exp\{-\kappa/T\}$, $\beta_2 = \exp\{-(\kappa+\delta)/T\}$, T is the sample size, and

$$u_t = \sum_{j=0}^{\infty} c_j e_{t-j} \quad \text{with} \quad e_t \overset{i.i.d.}{\sim} \left(0, \sigma^2 h\right) \text{ and } h = 1/T.$$

It is assumed that $c_0 = 1$ and $\sum_{j=0}^{\infty} j\,|c_j| < \infty$. Define $\gamma(j) \equiv E(u_t u_{t-j})$ for $j = 0, \pm 1, \pm 2, \ldots$, and $C(1) \equiv \sum_{j=0}^{\infty} c_j$. Note that the long-run variance of u_t goes to zero as $h \to 0$**:**

$$\lambda^2 \equiv \sum_{j=-\infty}^{\infty} \gamma(j) = [C(1)]^2 \sigma^2 h = O\left(T^{-1}\right) \to 0, \text{ as } T \to \infty.$$

It is also clear that Model (14) reduces to Model (8) if $c_j = 0$ for $j \geq 1$. For this model, no long-span asymptotic theory is available in the literature regardless of the value of β_1 and β_2.

To estimate the break point, the LS estimator defined in (10) is used. Note that $\hat{k}_{LS}$ is also the LS estimator of the break point for the process $y_t = x_t/\sqrt{h}$ which evolves as

$$y_t = \left(\beta_1 1_{[t \leq k_0]} + \beta_2 1_{[t > k_0]}\right) y_{t-1} + u_t^*, \quad \text{with} \quad y_0 = x_0/\sqrt{h} = O_p\left(T^{1/2}\right), \tag{15}$$

where

$$u_t^* = \frac{u_t}{\sqrt{h}} = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} \text{ and } \varepsilon_t = \frac{e_t}{\sqrt{h}} \overset{i.i.d.}{\sim} \left(0, \sigma^2\right).$$

Define $\gamma^*(j) \equiv E\left(u_t^* u_{t-j}^*\right) = \gamma(j)/h$ for $j = 0, \pm 1, \pm 2, \ldots$. The long-run variance of u_t^* is

$$(\lambda^*)^2 \equiv \sum_{j=-\infty}^{\infty} \gamma^*(j) = \lambda^2/h = [C(1)]^2 \sigma^2.$$

If $c_j = 0$ for $j \geq 1$, Model (15) will reduce to Model (12).

Theorem 5.1 *Consider the general discrete-time model with weekly dependent errors defined in (14). When $T \to \infty$ with a fixed $\tau_0 = k_0/T$, the in-fill asymptotic distribution of $\hat{\tau}_{LS} = \hat{k}_{LS}/T$ with $\hat{k}_{LS}$ defined in (10) is*

$$\hat{\tau}_{LS} \Rightarrow \arg\max_{\tau \in (0,1)} \frac{\left\{\left[\widetilde{J}_{\tau_0}(\tau)\right]^2 - \left[\widetilde{J}_{\tau_0}(0)\right]^2 - \phi\tau\right\}^2}{\int_0^\tau \left[\widetilde{J}_{\tau_0}(r)\right]^2 dr} + \frac{\left\{\left[\widetilde{J}_{\tau_0}(1)\right]^2 - \left[\widetilde{J}_{\tau_0}(\tau)\right]^2 - \phi[1-\tau]\right\}^2}{\int_\tau^1 \left[\widetilde{J}_{\tau_0}(r)\right]^2 dr} \tag{16}$$

where $\phi = \gamma^*(0)/[C(1)\sigma]^2$ and $\widetilde{J}_\tau(r)$ is the Gaussian process defined in (6) with the initial value $\widetilde{J}_{\tau_0}(0) = x_0/[C(1)\sigma]$.

Remark 5.1 *Note that $\gamma^*(0) \equiv E\left[(u_t^*)^2\right] = \sum_{j=0}^{\infty} c_j^2 \sigma^2$. Therefore, $\phi = \sum_{j=0}^{\infty} c_j^2 / [C(1)]^2$ which is independent of σ^2. However, the in-fill asymptotic distribution given in (16) explicitly depends on x_0 and σ through the initial value of $\widetilde{J}_{\tau_0}(0) = x_0/[C(1)\sigma]$. Moreover, if $c_j = 0$ for $j \geq 1$, $[C(1)]^2 = c_0^2 = \sum_{j=0}^{\infty} c_j^2$ and $\phi = 1$. Then, the in-fill asymptotic distribution given in (16) becomes the same as the one given in (13) for the model with i.i.d. errors.*

Remark 5.2 *With the same reasons, we expect the in-fill asymptotic distribution in (16)* ***to be*** *asymmetric and trimodal.*

6 Monte Carlo Results

In this section, we first design seven Monte Carlo experiments to compare the performance of our in-fill asymptotic distribution with the corresponding long-span asymptotic distribution developed in the literature. In each experiment, we plot densities of the long-span distribution, our in-fill distribution, and the 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 $\tau_0 = 0.3, 0.5, 0.7$, $\sigma = 1$, $\varepsilon_t \overset{i.i.d.}{\sim} N(0,1)$, $T = 200$ (i.e. $h = 1/200$) and different combination of κ and δ. In all cases we set $x_0/\sigma = 1$. All the pdfs are obtained by simulation with 100,000 replications. When we calculate the in-fill distribution and the long-span distribution, the stochastic integrals are approximated over a very small grid size, namely 0.001. Let β_1 and β_1 denote the AR(1) coefficients before and after the break.

In the first experiment, we set $\kappa = 138$ and $\delta = 55$, implying $\beta_1 = 0.5$ and $\beta_2 = 0.38$. In this experiment, we assume the AR(1) coefficient switches from a stationary root to another stationary root. The corresponding long-span asymptotic distribution is given in (17) in Appendix A. The three densities are plotted in the upper panel in Figure 3.

In the second experiment, we set $\kappa = 21$ and $\delta = -21$, implying $\beta_1 = 0.9$ and $\beta_2 = 1$. In this experiment, we assume the AR(1) coefficient switches from a stationary root to a unit root. Hence, the corresponding long-span asymptotic distribution is the one in (18). The three densities are plotted in the middle panel in Figure 3.

In the third experiment, we set $\kappa = 0$ and $\delta = 10$, implying $\beta_1 = 1$ and $\beta_2 = 0.95$. In this experiment, we assume the AR(1) coefficient switches from a unit root to a stationary root. Hence, the corresponding long-span asymptotic distribution is the one given in (19). The three densities are plotted in the lower panel in Figure 3.

In the fourth experiment, we set $\kappa = 10$ and $\delta = -9$, implying $\beta_1 = 0.95$ and $\beta_2 = 0.995$. In this experiment, we assume the AR(1) coefficient switches from a stationary root to a local-to-unit-root. Hence, the corresponding long-span asymptotic distribution is the one given in (20) where we set $c = -1$. The three densities are plotted in the upper panel in Figure 4.

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) coefficient switches from a local-to-unit-root to a stationary root. Hence, the corresponding long-span asymptotic distribution is the one given in (21) where we set $c = -1$. The three densities are plotted in the lower panel in Figure 4.

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) coefficient switches from a mildly stationary root to a unit root. Hence, the corresponding long-span asymptotic distribution is the one given in (22) where we set $c = -1$ and $k_T = 20$. The three densities are plotted in the upper panel in Figure 5.

In the seventh experiment, we set $\kappa = 0$ and $\delta = 7$, implying $\beta_1 = 1$ and $\beta_2 = 0.96$. In this experiment, we assume the AR(1) coefficient switches from a unit root to a mildly stationary root. Hence, the corresponding long-span asymptotic distribution is the one given in (23) where we set $k_T = 30$ and $c = -1.2$. The three densities are plotted in the lower panel in Figure 5.

As discussed in Section 2, the long-span asymptotic distributions are unknown for many other interesting cases. We then design two Monte Carlo experiments to check the performance of our in-fill asymptotic distribution in cases where the long-span theory is unavailable. In both experiments, we plot densities of our in-fill asymptotic distribution and the finite sample distribution.

In the eighth experiment, we set $\kappa = 0$ and $\delta = -6$, implying $\beta_1 = 1$ and $\beta_2 = 1.03$. This case is important because it is related to the recent literature that estimates the bubble origination date; see, for example, Phillips, Wu, and Yu (2011) and Phillips and Yu (2011), and Phillips, Shi, and Yu (2015a, b). The two densities are plotted in the upper panel in Figure 6.

In the ninth experiment, we set $\kappa = -6$ and $\delta = 12$, implying $\beta_1 = 1.03$ and $\beta_2 = 0.97$. This case is also empirically important because it is related to the recent literature that estimates the bubble termination date; see, for example, Phillips and Shi (2017). The two densities are plotted in the lower panel in Figure 6.

Several features are apparent in these figures. First, the finite sample distribution is asymmetric about 0 even when $\tau_0 = 1/2$. 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 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 distribution when the long-span theory is available (as apparent in Figures 3-5), despite that the sample size is reasonably large ($T = 200$). It continues to provide accurate approximations to the finite sample distribution when the long-span theory is not available, as apparent in Figure 6.

We now turn our attention to the first moment of alternative distributions and hence the bias of $\hat{\tau}_{LS}$. Since our in-fill distribution is closer to the finite sample distribution

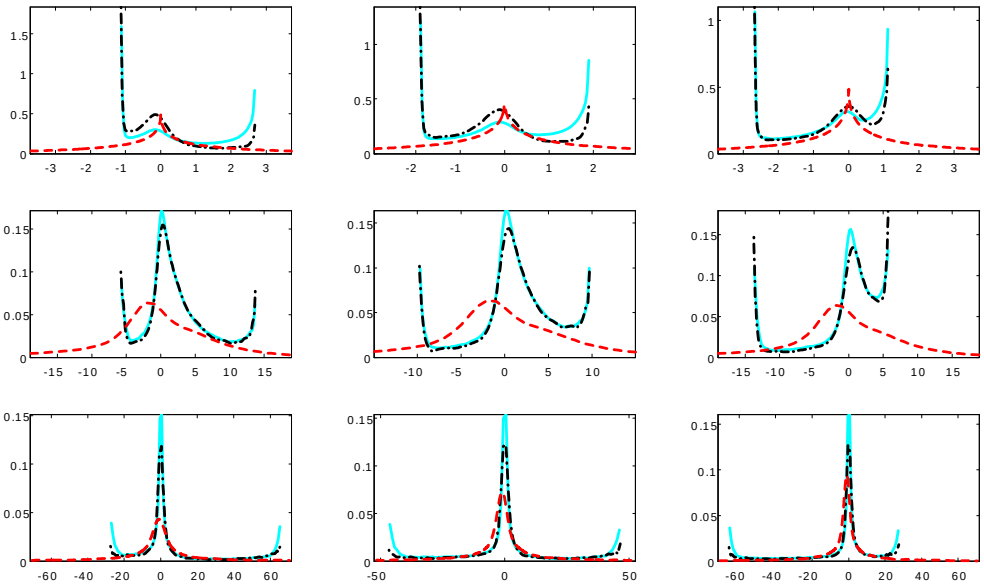

Figure 3: The pdf of $\frac{T(\beta_2-\beta_1)^2}{1-\beta_1^2}(\hat{\tau}_{LS}-\tau_0)$ when $\beta_1 = 0.5$, $\beta_2 = 0.38$; the pdf of $T(1-\beta_1)(\hat{\tau}_{LS}-\tau_0)$ when $\beta_1 = 0.9$, $\beta_2 = 1$; and the pdf of $T^2(\beta_2-1)^2(\hat{\tau}_{LS}-\tau_0)$ when $\beta_1 = 1$, $\beta_2 = 0.95$, (the upper, middle and lower panel respectively), with $x_0/\sigma = 1$ and $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively). Solid lines are finite sample distributions when $T = 200$; dashed lines are in-fill densities from Theorem 4.1; and broken lines are long-span limiting distributions.

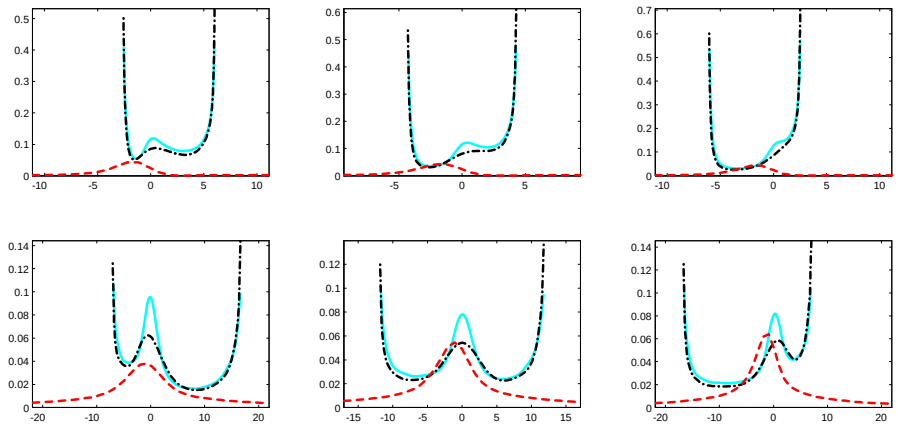

Figure 4: The pdf of $T(\beta_2-\beta_1)(\hat{\tau}_{LS}-\tau_0)$ when $\beta_1 = 0.95$, $\beta_2 = 0.995$, and the pdf of $T^2(\beta_2-\beta_1)^2(\hat{\tau}_{LS}-\tau_0)$ when $\beta_1 = 0.995$, $\beta_2 = 0.97$ (the upper and lower panel respectively), with $x_0/\sigma = 1$ and $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively). Solid lines are finite sample distributions when $T = 200$; dashed lines are in-fill densities from Theorem 4.1; and broken lines are long-span limiting distributions.

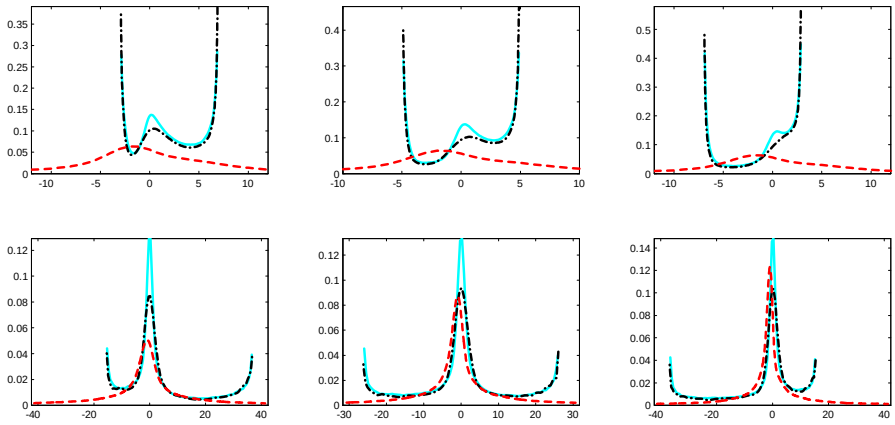

Figure 5: The pdf of $\frac{cT}{k_T}(\hat{\tau}_{LS} - \tau_0)$ when $\beta_1 = 0.95$, $\beta_2 = 1, c = 1, k_T = 20$, and the pdf of $\frac{c^2T^2}{k_T^2}(\hat{\tau}_{LS} - \tau_0)$ when $\beta_1 = 1, \beta_2 = 0.96, c = 1.2, k_T = 30$ (the upper and lower panel respectively), with $x_0/\sigma = 1$ and $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively). Solid lines are finite sample distributions when $T = 200$; dashed lines are in-fill densities from Theorem 4.1; and broken lines are long-span limiting distributions.

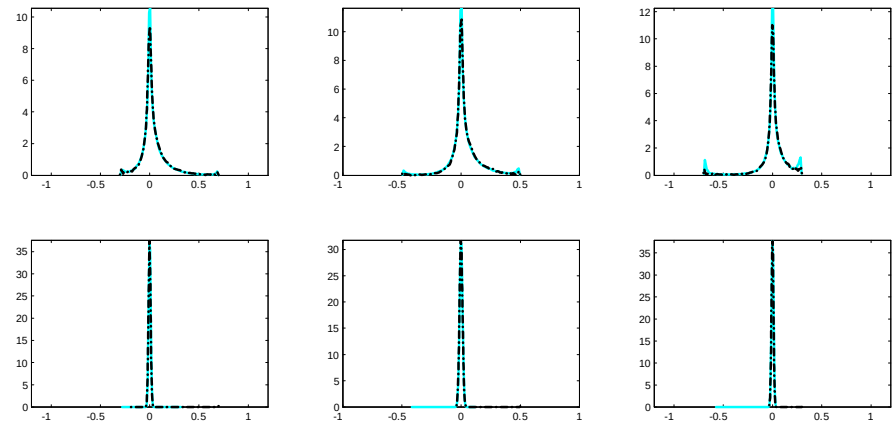

Figure 6: The pdf of $(\hat{\tau}_{LS} - \tau_0)$ when $\beta_1 = 1$, $\beta_2 = 1.03$, and the pdf of $(\hat{\tau}_{LS} - \tau_0)$ when $\beta_1 = 1.03$, $\beta_2 = 0.97$ (the upper and lower panel respectively), with $x_0/\sigma = 1$ and $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively). Solid lines are finite sample distributions when $T = 200$; dashed lines are in-fill densities from Theorem 4.1.

Table 2: The table shows the finite sample bias of $\hat{\tau}_{LS}$, the bias implied by the in-fill asymptotic distribution, and the bias implied by the long-span asymptotic distribution when the AR(1) process switches from a stationary root to another stationary root with different break sizes, $x_0/\sigma = 0.2$ and $T = 200$. The number of replications is 10,000.

	β_1	β_2				
τ_0	0.5	0.45	0.55	0.61	0.74	0.83
0.3	Finite	0.2113	0.2675	0.2648	0.1792	0.1093
0.3	In-fill	0.2871	0.3261	0.2899	0.1590	0.0887
0.3	Long span	0	0	0	0	0
0.5	Finite	0.0146	0.0745	0.0933	0.0743	0.0491
0.5	In-fill	0.1004	0.1321	0.1235	0.0768	0.0501
0.5	Long span	0	0	0	0	0
0.7	Finite	-0.1777	-0.1245	-0.0840	-0.0235	0.0029
0.7	In-fill	-0.0793	-0.0621	-0.0435	0.0044	0.0200
0.7	Long span	0	0	0	0	0

than the long-span distribution, it is expected that the bias implied by the in-fill theory should be closer to the true bias. To confirm this conjecture, we design an experiment where the AR(1) coefficient switches from a stationary root ($\beta_1 = 0.5$) to another stationary root ($\beta_2 = 0.45$, 0.55, 0.61, 0.74, or 0.83) with $\tau_0 = 0.3, 0.5, 0.7$, $x_0/\sigma = 0.2$, $T = 200$. Table 2 reports the true bias, the bias implied by the in-fill distribution and the bias implied by the long-span distribution. We may draw the following conclusions from Table 2. First, the LS estimate suffers from severe bias problem in nearly all cases. For example, when $\beta_1 = 0.5$, $\beta_2 = 0.55$, $\tau_0 = 0.3$, the bias is 0.2675 which is about 90% of the true value. Furthermore, the LS estimate is biased even when $\tau_0 = 0.5$. When $\beta_1 = 0.5$, $\beta_2 = 0.61$, $\tau_0 = 0.5$ (the same design that gives rise to Figure 1), the bias is 0.0933 which is about 20% of the true value. Second, there is no bias according to the long-span distribution. This is not surprising because the long-span distribution corresponds to $\arg\max_{u\in(-\infty,\infty)} \left\{W(u) - \frac{1}{2}|u|\right\}$ in this case. Hence, the long-span distribution not only fails to approximate the finite sample distribution but also fails to approximate the first moment. Third, the in-fill asymptotic distribution can approximate the true bias well in all cases considered.

In another experiment, we allow the model to switch from a unit root ($\beta_1 = 1$) to an explosive root ($\beta_2 = 1.01$, 1.02, 1.03, 1.04, or 1.05) with $\tau_0 = 0.3, 0.5, 0.7$, $x_0/\sigma = 0.2$, $T = 200$. In this case, the long-span asymptotic theory is not available. Table 3 reports the true bias and the bias implied by the in-fill distribution. Some remarks can be made.

Table 3: The table shows the finite sample bias of $\hat{\tau}_{LS}$, the bias implied by the in-fill asymptotic distribution, and the bias from the long-span asymptotic distribution when the AR(1) process switches from a unit root to mildly explosive root with different break sizes, $x_0/\sigma = 0.2$ and $T = 200$. The number of replications is 10,000.

	β_1	β_2				
τ_0	1	1.01	1.02	1.03	1.04	1.05
0.3	Finite	0.2247	0.2445	0.2291	0.1751	0.1223
0.3	In-fill	0.2112	0.2355	0.2322	0.1817	0.1297
0.5	Finite	0.0213	0.0588	0.0648	0.0496	0.0369
0.5	In-fill	0.0102	0.0500	0.0660	0.0570	0.0416
0.7	Finite	-0.1826	-0.1036	-0.0293	-0.0017	0.0095
0.7	In-fill	-0.1940	-0.1158	-0.0349	-0.0008	0.0119

First, the LS estimate can suffer from a bias problem in this case. For example, when $\beta_1 = 1$, $\beta_2 = 1.02$, $\tau_0 = 0.3$, the bias is 0.2445 which is about 80% of the true value. Given the importance of this estimator for bubble detection (see, for example, Phillips, Wu, and Yu, 2011), the bias reported here must have serious empirical implications. Second, the in-fill asymptotic distribution can approximate the true bias well in all cases considered.

We now shift our attention to the impact of the initial condition. While the long-span distribution is independent of the initial condition, both the finite sample distribution and the in-fill distribution depend on the initial condition. We have already shown that the in-fill distribution provides excellent approximations to the finite sample distribution and that the bias implied by the in-fill distribution is very close to the true bias in all cases. To examine the impact of the initial condition, we focus on the bias implied by the in-fill distribution. In particular, we plot the bias function (i.e., $E(\hat{\tau}_{LS})$ as a function of τ_0) implied by the in-fill asymptotics for Model (12) and examine the sensitivity of the function to the initial condition.

Figures 7-9 plot the bias function when $x_0 = 0.2, 0.4, 0.6, 0.8, 1$ and $\sigma = 1$. Figure 7 corresponds to the case where $\beta_1 = 0.9$, $\beta_1 = 1$; Figure 8 to $\beta_1 = 1$, $\beta_1 = 0.9$; Figure 9 to $\beta_1 = 1$, $\beta_1 = 1.03$. Several conclusions can be drawn. First, the initial condition can have a significant impact on the magnitude of bias. Specifically, when x_0/σ gets bigger, the bias becomes smaller generally. This result corroborates the result obtained in Perron (1991, Figure 4) in the context of AR(1) model without break. Second, it seems there exists a value of τ_0 (which depends on the values of β_1 and β_2), at which the bias may not be zero but becomes invariant of the initial condition.

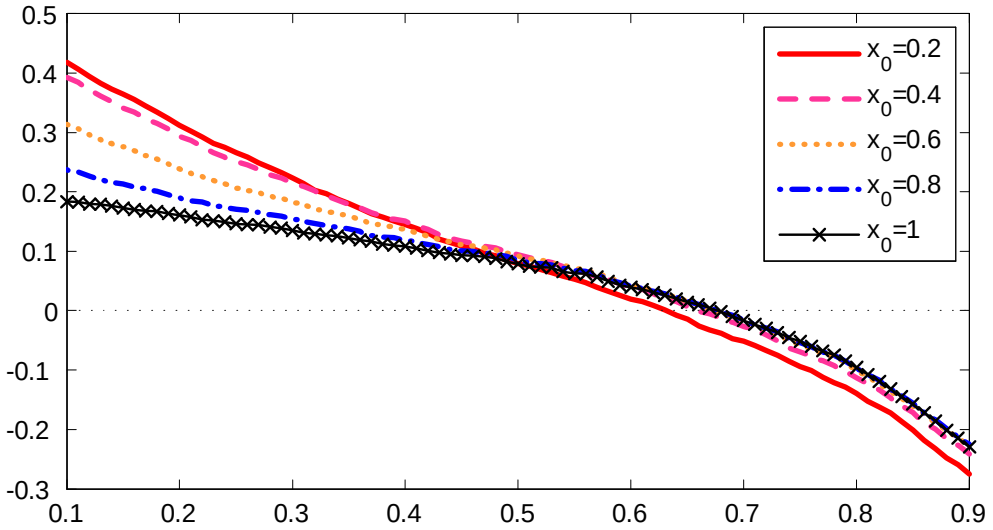

Figure 7: Bias functions implied by the in-fill asymptotic distribution given in Theorem 4.1 when $\beta_1 = 0.9$ and $\beta_2 = 1$ with various initial conditions.

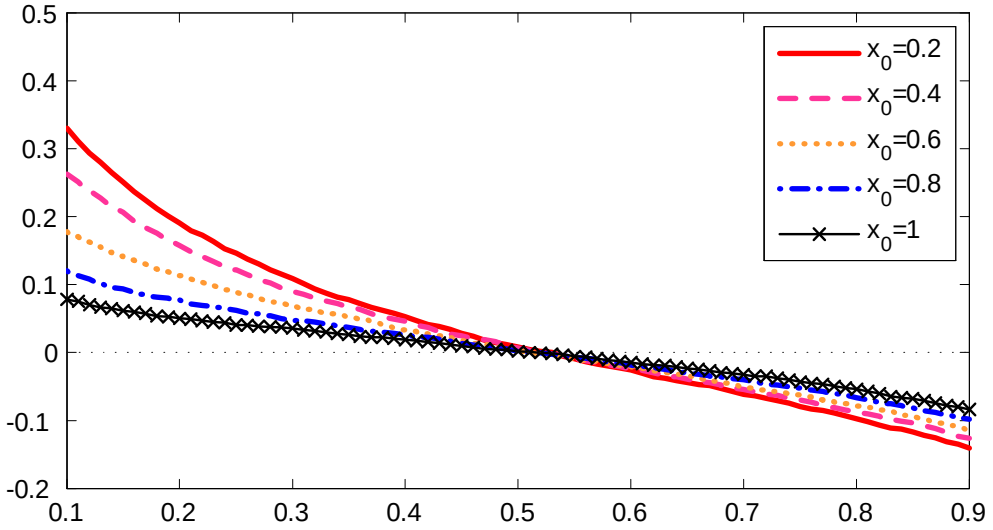

Figure 8: Bias functions implied by the in-fill asymptotic distribution given in Theorem 4.1 when $\beta_1 = 1$ and $\beta_2 = 0.9$ with various initial conditions.

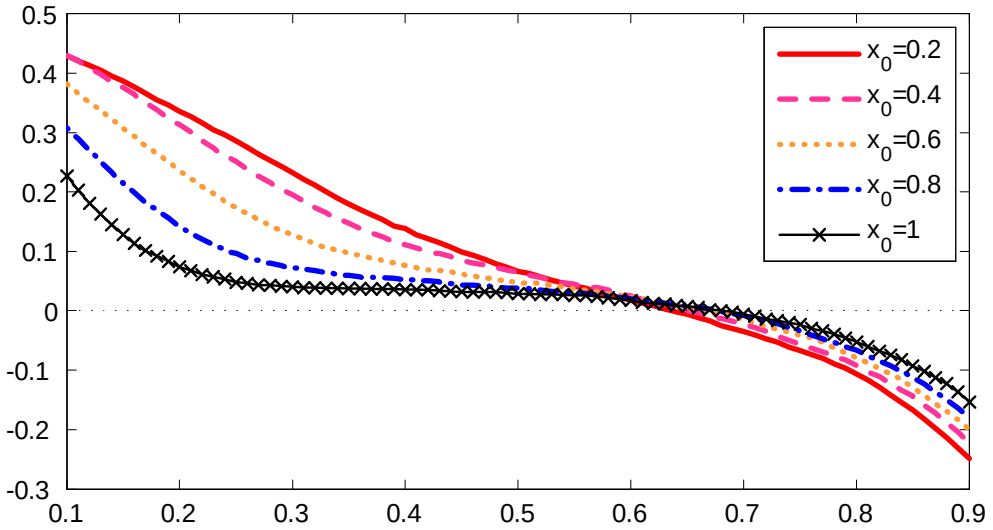

Figure 9: Bias functions implied by the in-fill asymptotic distribution given in Theorem 4.1 when $\beta_1 = 1$ and $\beta_2 = 1.03$ with various initial conditions.

7 Conclusions

This paper is concerned about the large sample approximation to the finite sample 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 exact distribution is asymmetric and trimodal. These two properties are also found in the finite sample distribution of the LS estimator of structural break point in AR models.

Unfortunately, the literature on the estimation of structural break point in AR models has always focused on developing asymptotic theory by assuming the time spans before and after the break go to infinity. We show that the long-span theory provides poor approximation to the finite sample distribution in many empirically relevant cases. Moreover, the long-span asymptotics developed in the literature are different, depending on the distance and the direction from the unity for underlying AR(1) coefficients. This discontinuity in the long-span asymptotic distributions makes it difficult to use in practice. Furthermore, the existing limiting theory is developed for a few cases only, leaving out some empirically interesting cases. Finally, the model considered in the literature is quite restrictive as the errors are independent.

This paper provides a unified limiting theory for the break point estimate in the AR(1) model with independent errors as well as the model with weakly dependent er-

rors. It develops an in-fill asymptotic theory for the LS estimator of structural break point. The developed in-fill asymptotic distribution is continuous in the underlying persistency parameters, regardless of their signs and values. We also show that this distribution is asymmetric and trimodal, and approximates the finite sample distribution better than the long-span distribution developed in the literature when the latter is known and provides excellent approximations to the finite sample distribution when the latter is unknown.

APPENDIX

A A detailed literature review

In the following, we review the main results on the long-span asymptotic distributions developed in the literature. In some cases, the AR roots, β_1 and β_2, are assumed to be functions of the sample size T. Then, we use β_{1T} and β_{2T} to replace β_1 and β_2 accordingly.

Chong (2001) first studied Model (1) with $|\beta_1| < 1$ and $|\beta_2| < 1$, where the AR(1) coefficient switches from a stationary root to another stationary root. To derive the long-span asymptotic distribution for the model with a small break size, Chong (2017) let β_2 depend on T, denoted as β_{2T}, and assumed that $\beta_{2T} - \beta_1 \to 0$ with $\sqrt{T}\,|\beta_{2T} - \beta_1| \to \infty$ as $T \to \infty$. Under the condition that $y_0 = O_p(1)$, he derived the long-span asymptotic distribution of $\hat{\tau}_{LS,T}$ as

$$\frac{T(\beta_{2T} - \beta_1)^2}{1 - \beta_1^2}(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg \max_{u \in (-\infty,\infty)} \left\{ W(u) - \frac{1}{2}|u| \right\}, \tag{17}$$

where $W(u)$ is a two-sided Brownian motion, defined as $W(u) = B_1(-u)$ if $u \leq 0$ and $W(u) = B_2(u)$ if $u > 0$, with B_1 and B_2 being two independent Brownian motions. The pdf and the cdf for this limiting distribution have been derived in Yao (1987).

Chong (2001) then studied Model (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. He let $\beta_1 = \beta_{1T}$, and assumed that $1 - \beta_{1T} \to 0$ with $T(1 - \beta_{1T}) \to \infty$ as $T \to \infty$. In this case, he proved that the the long-span asymptotic distribution of $\hat{\tau}_{LS,T}$ takes the form of

$$T(1 - \beta_{1T})(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg \max_{u \in (-\infty,\infty)} \left\{ \frac{W_a^*(u)}{R_1} - \frac{1}{2}|u| \right\}, \tag{18}$$

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 $R_1 = \int_0^\infty \exp(-s) dW_1(s)$.

Chong (2001) also studied Model (1) with $\beta_1 = 1$ and $|\beta_2| < 1$, where the AR model switches from a unit root to a stationary root. Assuming that $\beta_2 = \beta_{2T}$ with the condition $\sqrt{T}(1 - \beta_{2T}) \to 0$ and $T^{3/4}(1 - \beta_{2T}) \to \infty$ as $T \to \infty$, he derived a long-span asymptotic distribution of $\hat{\tau}_{LS,T}$ as

$$T^2(\beta_{2T} - 1)^2(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg \max_{u \in (-\infty, \infty)} \left\{ \frac{W(u)}{W_3(\tau_0)} - \frac{1}{2}|u| \right\}, \tag{19}$$

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

Pang, Zhang and Chong (2014) studied Model (1) with $|\beta_{1T}| < 1$ and $\beta_{2T} = 1 \pm c/T$. In this case the AR model switches from a stationary root to a local-to-unit-root. Under the assumptions that $y_0 = o_p(\sqrt{T})$, $|\beta_{2T} - \beta_{1T}| \to 0$ with $T(\beta_{2T} - \beta_{1T}) \to \infty$, they derived an asymptotic distribution of $\hat{\tau}_{LS,T}$ as

$$T(\beta_{2T} - \beta_1)(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg \max_{u \in (-\infty, \infty)} \left\{ \frac{W_b^*(u)}{R_1} - \frac{1}{2}|u| \right\}, \tag{20}$$

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

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

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 $R_1 = \int_0^\infty \exp(-s) dW_1(s)$.

Pang, Zhang and Chong (2014) also studied Model (1) with $\beta_{1T} = 1 + c/T$ and $\beta_{2T} < 1$. In this case the AR model switches from a local-to-unit-root to a stationary root. Under the assumptions that $y_0 = o_p\left(\sqrt{T}\right)$, $\sqrt{T}(\beta_{2T} - \beta_{1T}) \to 0$ with $T^{3/4}(\beta_{2T} - \beta_{1T}) \to \infty$, they proved that $\hat{\tau}_{LS,T}$ has the long-span asymptotic distribution as

$$T^2(\beta_2 - \beta_{1T})^2(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg \max_{u \in (-\infty, \infty)} \left\{ \frac{W(u)}{\exp\left(c(1 - \tau_0)\right) G(W_1, c, \tau_0)} - \frac{1}{2}|u| \right\}, \tag{21}$$

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

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

Liang et al. (2017) studied Model (1) with $\beta_{1T} = 1 - c/T^\alpha$ and $\beta_2 = 1$ where c is a positive constant and $\alpha \in (0,1)$.[6] In this case the AR model switches from a mildly stationary root to a unit root. Under the assumptions that $y_0 = o_p(\sqrt{T^\alpha})$, they derived a long-span asymptotic distribution of $\hat{\tau}_{LS,T}$ as

$$\frac{cT}{T^\alpha}(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg\max_{u\in(-\infty,\infty)} \left\{ \frac{W_c^*(u)}{R_c} - \frac{1}{2}|u| \right\}, \tag{22}$$

where $W_c^*(u) = W_1(-u)$ when $u \le 0$, and when $u > 0$

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

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

Liang et al. (2017) also studied Model (1) with $\beta_1 = 1$ and $\beta_{2T} = 1 - c/T^\alpha$. In this case the AR model switches from a unit root to a mildly stationary root. Under the assumptions that $y_0 = o_p(\sqrt{T^\alpha})$, $\sqrt{T}/T^\alpha \to 0$ and $T^{3/4}/T^\alpha \to \infty$ as $T \to \infty$, they derived a long-span asymptotic distribution of $\hat{\tau}_{LS,T}$ as

$$\frac{c^2T^2}{T^{2\alpha}}(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg\max_{u\in(-\infty,\infty)} \left\{ \frac{W(u)}{W_1(\tau_0)} - \frac{1}{2}|u| \right\}, \tag{23}$$

where $W(u)$ is a two-sided Brownian motion and $W_1(\cdot)$ is an independent standard Brownian motion.

B Proofs

Lemma B.1 *Consider the process y_t defined in (12) with the dynamics*

$$y_t = \left(\beta_1 1_{[t\le k_0]} + \beta_2 1_{[t>k_0]}\right) y_{t-1} + \varepsilon_t, \;\; \varepsilon_t \overset{i.i.d.}{\sim} \left(0, \sigma^2\right), \;\; y_0 = x_0/\sqrt{h}.$$

When $T = 1/h \to \infty$ with a fixed $\tau_0 = k_0/T$, for any $\tau \in [0,1]$,
(a) $T^{-1}\sum_{t=1}^{\lfloor T\tau\rfloor} y_{t-1}\varepsilon_t \Rightarrow \sigma^2 \int_0^\tau \widetilde{J}_{\tau_0}(r)dB(r)$;
(b) $T^{-2}\sum_{t=1}^{\lfloor T\tau\rfloor} y_{t-1}^2 \Rightarrow \sigma^2 \int_0^\tau \left[\widetilde{J}_{\tau_0}(r)\right]^2 dr$;

[6] Following Phillips and Magdalinos (2007), Liang et al. (2017) used k_T instead of T^α with the assumption that $k_T \to \infty$ and $k_T/T \to 0$.

(c) $\left[\widetilde{J}_{\tau_0}(\tau)\right]^2 - \left[\widetilde{J}_{\tau_0}(0)\right]^2 = 2\int_0^\tau \widetilde{J}_{\tau_0}(r)dB(r) - 2\int_0^\tau \left(\kappa + \delta 1_{[r>\tau_0]}\right)\left[\widetilde{J}_{\tau_0}(r)\right]^2 dr + \tau;$
(d) $\left[\widetilde{J}_{\tau_0}(1)\right]^2 - \left[\widetilde{J}_{\tau_0}(\tau)\right]^2 = 2\int_\tau^1 \widetilde{J}_{\tau_0}(r)dB(r) - 2\int_\tau^1 \left(\kappa + \delta 1_{[r>\tau_0]}\right)\left[\widetilde{J}_{\tau_0}(r)\right]^2 dr + (1-\tau),$
where $\lfloor T\tau \rfloor$ *denotes the integer part of* $T\tau$, $\widetilde{J}_{\tau_0}(r)$ *for* $r \in [0,1]$ *is a Gaussian process generated by* $d\widetilde{J}_{\tau_0}(r) = -\left(\kappa + \delta 1_{[r>\tau_0]}\right)\widetilde{J}_{\tau_0}(r)dr + dB(r)$ *with the initial value* $\widetilde{J}_{\tau_0}(0) = x_0/\sigma$, *and* $B(r)$ *is a standard Brownian motion.*

Lemma B.2 *Consider the process* y_t *defined in (15) with the dynamics*

$$y_t = \left(\beta_1 1_{[t\le k_0]} + \beta_2 1_{[t>k_0]}\right) y_{t-1} + u_t^*, \quad y_0 = x_0/\sqrt{h}$$

where

$$u_t^* = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j},\ \varepsilon_t \overset{i.i.d.}{\sim} \left(0, \sigma^2\right),\ c_0 = 1 \text{ and } \sum_{j=0}^{\infty} j\,|c_j| < \infty.$$

Define $\gamma^*(j) \equiv E\left(u_t^* u_{t-j}^*\right)$ *for* $j = 0, \pm 1, \pm 2, \ldots$ *and* $C(1) = \sum_{j=0}^{\infty} c_j$. *When* $T = 1/h \to \infty$ *with a fixed* $\tau_0 = k_0/T$, *for any* $\tau \in [0,1]$,
(a) $T^{-2}\sum_{t=1}^{\lfloor T\tau \rfloor} y_{t-1}^2 \Rightarrow [C(1)\sigma]^2 \int_0^\tau \left[\widetilde{J}_{\tau_0}(r)\right]^2 dr;$
(b) $T^{-1}\sum_{t=1}^{\lfloor T\tau \rfloor} y_{t-1}u_t^* \Rightarrow [C(1)\sigma]^2 \int_0^\tau \widetilde{J}_{\tau_0}(r)dB(r) + (\tau/2)\left\{[C(1)\sigma]^2 - \gamma^*(0)\right\};$
where $\lfloor T\tau \rfloor$ *denotes the integer part of* $T\tau$, $\widetilde{J}_{\tau_0}(r)$ *for* $r \in [0,1]$ *is a Gaussian process generated by* $d\widetilde{J}_{\tau_0}(r) = -\left(\kappa + \delta 1_{[r>\tau_0]}\right)\widetilde{J}_{\tau_0}(r)dr + dB(r)$ *with the initial value* $\widetilde{J}_{\tau_0}(0) = x_0/[C(1)\sigma]$, *and* $B(r)$ *is a standard Brownian motion.*

Proof of Lemma B.1: When $\tau \le \tau_0$, the process y_t for $t = 1, 2, \ldots, \lfloor T\tau \rfloor$ has no break. Then, the results in (a) and (b) can be obtained straightforwardly by using the large sample theory for local-to-unity process; see, for example, Perron (1991). When $\tau > \tau_0$, the AR root of y_t changes from β_1 to β_2 at the point $t = k_0 = T\tau_0$. We can apply the large sample theory for local-to-unity process separately on different sides of the break to get the result in (a) as

$$\begin{aligned} T^{-1}\sum_{t=1}^{\lfloor T\tau \rfloor} y_{t-1}\varepsilon_t &= T^{-1}\sum_{t=1}^{\lfloor T\tau_0 \rfloor} y_{t-1}\varepsilon_t + T^{-1}\sum_{t=\lfloor T\tau_0 \rfloor+1}^{\lfloor T\tau \rfloor} y_{t-1}\varepsilon_t \\ &\Rightarrow \sigma^2\left\{\int_0^{\tau_0} \widetilde{J}_{\tau_0}(r)dB(r) + \int_{\tau_0}^{\tau} \widetilde{J}_{\tau_0}(r)dB(r)\right\} = \sigma^2 \int_0^{\tau} \widetilde{J}_{\tau_0}(r)dB(r). \end{aligned}$$

Similarly, the result in (b) for $\tau > \tau_0$ can be obtained.

The results in (c) and (d) can be derived directly from the diffusion function

$$\begin{aligned} d\left[\widetilde{J}_{\tau_0}(r)\right]^2 &= 2\widetilde{J}_{\tau_0}(r)d\widetilde{J}_{\tau_0}(r) + dr \\ &= 2\widetilde{J}_{\tau_0}(r)dB(r) - 2\left(\kappa + \delta 1_{[r>\tau_0]}\right)\left[\widetilde{J}_{\tau_0}(r)\right]^2 dr + dr, \end{aligned}$$

where the first equation comes from Itô's lemma.

Proof of Lemma B.2: When $\tau \leq \tau_0$, the process y_t for $t = 1, 2, \ldots, \lfloor T\tau \rfloor$ has no break. Then, (a) and (b) are just extensions of the results in Phillips (1987b) from the case where $x_0 = 0$ to the case where $x_0 \neq 0$. These extensions can be done easily by using the approach proposed in Perron (1991).

When $\tau > \tau_0$, the AR root of y_t changes from β_1 to β_2 at the point $t = T\tau_0$. Then, the method to prove Lemma B.1 can be used again to get (a) and (b) in this lemma.

Proof of Theorem 3.1: Note that

$$\begin{aligned}
\hat{\tau}_{ML} &= \arg\max_{\tau\in(0,1)} \{\log \mathcal{L}(\tau)\} \\
&= \arg\max_{\tau\in(0,1)} \frac{1}{\sigma^2}\left\{-\int_0^1 (\kappa + \delta 1_{[t>\tau]})x(t)dx(t) - \frac{1}{2}\int_0^1 \left(\kappa + \delta 1_{[t>\tau]}\right)^2 [x(t)]^2 dt\right\} \\
&= \arg\max_{\tau\in(0,1)} \frac{1}{\sigma^2}\left\{-\int_0^1 \delta 1_{[t>\tau]}x(t)dx(t) - \frac{1}{2}\int_0^1 \left(2\kappa\delta + \delta^2\right) 1_{[t>\tau]} [x(t)]^2 dt\right\} \\
&= \arg\max_{\tau\in(0,1)} -\frac{\delta}{\sigma^2}\left\{\int_\tau^1 x(t)dx(t) + \frac{1}{2}\int_\tau^1 (2\kappa + \delta) [x(t)]^2 dt\right\} \\
&= \arg\max_{\tau\in(0,1)} -\frac{\delta}{\sigma^2}\left\{\int_\tau^1 x(t)dx(t) - \frac{1}{2}\int_0^\tau (2\kappa + \delta) [x(t)]^2 dt\right\}
\end{aligned}$$

where the third equation is obtained by deleting the terms independent of the choice of τ but appearing in the second equation. Applying Itô's lemma to the diffusion process $x(t)$ defined in (3) leads to

$$d[x(t)]^2 = 2x(t)dx(t) + \sigma^2 dt.$$

Hence,

$$\begin{aligned}
\int_\tau^1 x(t)dx(t) &= \frac{1}{2}\int_\tau^1 d[x(t)]^2 - \frac{1}{2}\int_\tau^1 \sigma^2 dt \\
&= \frac{1}{2}\left([x(1)]^2 - [x(\tau)]^2\right) - \frac{1}{2}\sigma^2(1-\tau).
\end{aligned}$$

We then have

$$\begin{aligned}
\hat{\tau}_{ML} &= \arg\max_{\tau\in(0,1)} -\frac{\delta}{\sigma^2}\left\{\frac{1}{2}\left([x(1)]^2 - [x(\tau)]^2\right) - \frac{1}{2}\sigma^2(1-\tau) - \frac{2\kappa+\delta}{2}\int_0^\tau [x(t)]^2 dt\right\} \\
&= \arg\max_{\tau\in(0,1)} -\frac{\delta}{\sigma^2}\left\{-[x(\tau)]^2 + \sigma^2\tau - (2\kappa+\delta)\int_0^\tau [x(t)]^2 dt\right\} \\
&= \arg\max_{\tau\in(0,1)} \delta\left\{\left[\widetilde{J}_{\tau_0}(\tau)\right]^2 - \tau + (2\kappa+\delta)\int_0^\tau \left[\widetilde{J}_{\tau_0}(t)\right]^2 dt\right\}
\end{aligned}$$

where the second equation is obtained by deleting the terms independent of the choice of τ but appearing in the first equation, and the third equation comes form the relationship of $\widetilde{J}_{\tau_0}(t) = x(t)/\sigma^2$ which can be obtained from the definitions of $\widetilde{J}_{\tau_0}(t)$ and $x(t)$ as in (6) and (3), respectively.

Proof of Theorem 4.1: First note that $\hat{k}_{LS}$ defined in (10) can be identically represented as

$$\hat{k}_{LS} = \arg \min_{k=1,\cdots,T-1} S(k), \text{ with } S(k) = \sum_{t=1}^{k} \left(y_t - \hat{\beta}_1(k) y_{t-1}\right)^2 + \sum_{t=k+1}^{T} \left(y_t - \hat{\beta}_2(k) y_{t-1}\right)^2$$

where $\hat{\beta}_1(k) = \sum_{t=1}^{k} y_t y_{t-1} / \sum_{t=1}^{k} y_{t-1}^2$, $\hat{\beta}_2(k) = \sum_{t=k+1}^{T} y_t y_{t-1} / \sum_{t=k+1}^{T} y_{t-1}^2$, and $y_t = x_t/\sqrt{h}$ is defined in (12). Define the $T \times 2$ matrix $Y(k) = \begin{bmatrix} Y_1(k) & Y_2(k) \end{bmatrix}$ with $Y_1(k) = \begin{bmatrix} y_0 & \cdots & y_{k-1} & 0 & \cdots & 0 \end{bmatrix}'$ and $Y_2(k) = \begin{bmatrix} 0 & \cdots & 0 & y_k & \cdots & y_{T-1} \end{bmatrix}'$. Let $Y = \begin{bmatrix} y_1 & \cdots & y_T \end{bmatrix}'$. Then, standard linear regression algebra gives an identical representation of the sum of squared residuals:

$$S(k) = Y'MY \text{ with } M = I - Y_1(k)\left[Y_1'(k) Y_1(k)\right]^{-1} Y_1'(k) - Y_2(k)\left[Y_2'(k) Y_2(k)\right]^{-1} Y_2'(k),$$

where I is an $T \times T$ identity matrix. From equation (12), we have

$$y_t = \beta_1 y_{t-1} + (\beta_2 - \beta_1) 1_{[t>k_0]} y_{t-1} + \varepsilon_t = \beta_1 y_{t-1} + \eta_t$$

where $\eta_t \equiv (\beta_2 - \beta_1) 1_{[t>k_0]} y_{t-1} + \varepsilon_t$. Let $Y_- = \begin{bmatrix} y_0 & \cdots & y_{T-1} \end{bmatrix}'$ and $\eta = \begin{bmatrix} \eta_1 & \cdots & \eta_T \end{bmatrix}'$. We then have

$$Y = Y_- \beta_1 + \eta.$$

Therefore,

$$\begin{aligned} S(k) &= Y'MY = Y'M'MY = (Y_-\beta_1 + \eta)' M'M (Y_-\beta_1 + \eta) = \eta' M \eta \\ &= \eta'\eta - \eta' Y_1(k)\left[Y_1'(k) Y_1(k)\right]^{-1} Y_1'(k)\eta - \eta' Y_2(k)\left[Y_2'(k) Y_2(k)\right]^{-1} Y_2'(k)\eta \end{aligned}$$

where the second equation is from $M'M = M$ and the fourth equation is because $MY_- = \mathbf{0}_{T\times 1}$. Note that

$$\eta'\eta = \sum_{t=1}^{k_0} \eta_t^2 + \sum_{t=k_0+1}^{T} \eta_t^2 = \sum_{t=1}^{k_0} \varepsilon_t^2 + \sum_{t=k_0+1}^{T} \left[(\beta_2 - \beta_1) y_{t-1} + \varepsilon_t\right]^2,$$

which is independent of the choice of k, and

$$\begin{aligned} \eta' Y_1(k)\left[Y_1'(k) Y_1(k)\right]^{-1} Y_1'(k)\eta &= \frac{\left(\sum_{t=1}^{k} y_{t-1}\eta_t\right)^2}{\sum_{t=1}^{k} y_{t-1}^2} \\ \eta' Y_2(k)\left[Y_2'(k) Y_2(k)\right]^{-1} Y_2'(k)\eta &= \frac{\left(\sum_{t=k+1}^{T} y_{t-1}\eta_t\right)^2}{\sum_{t=k+1}^{T} y_{t-1}^2}. \end{aligned}$$

Hence,

$$\hat{k}_{LS} = \arg \min_{k=1,\cdots,T-1} S(k) = \arg \max_{k=1,\cdots,T-1} \left\{ \frac{\left(\sum_{t=1}^{k} y_{t-1}\eta_t\right)^2}{\sum_{t=1}^{k} y_{t-1}^2} + \frac{\left(\sum_{t=k+1}^{T} y_{t-1}\eta_t\right)^2}{\sum_{t=k+1}^{T} y_{t-1}^2} \right\}. \tag{24}$$

The same transformation method has been used in Elliott and Müller (2007) for a general linear time series regression with a single break.

Note that

$$\hat{\tau}_{LS} = \hat{k}_{LS}/T = \arg \max_{\tau \in (0,1)} \left\{ \frac{\left(\sum_{t=1}^{\lfloor T\tau \rfloor} y_{t-1}\eta_t\right)^2}{\sum_{t=1}^{\lfloor T\tau \rfloor} y_{t-1}^2} + \frac{\left(\sum_{t=\lfloor T\tau \rfloor+1}^{T} y_{t-1}\eta_t\right)^2}{\sum_{t=\lfloor T\tau \rfloor+1}^{T} y_{t-1}^2} \right\}.$$

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

$$T^{-1} \sum_{t=1}^{\lfloor T\tau \rfloor} y_{t-1}\eta_t = T^{-1} \sum_{t=1}^{\lfloor T\tau \rfloor} y_{t-1}\varepsilon_t \Rightarrow \sigma^2 \int_0^{\tau} \widetilde{J}_{\tau_0}(r) dB(r)$$

and

$$\begin{aligned}
\frac{1}{T} \sum_{t=\lfloor T\tau \rfloor+1}^{T} y_{t-1}\eta_t &= \frac{1}{T} \left[\sum_{t=\lfloor T\tau \rfloor+1}^{\lfloor T\tau_0 \rfloor} y_{t-1}\eta_t + \sum_{t=\lfloor T\tau_0 \rfloor+1}^{T} y_{t-1}\eta_t \right] \\
&= \frac{1}{T} \left[\sum_{t=\lfloor T\tau \rfloor+1}^{\lfloor T\tau_0 \rfloor} y_{t-1}\varepsilon_t + (\beta_2 - \beta_1) \sum_{t=\lfloor T\tau_0 \rfloor+1}^{T} y_{t-1}^2 + \sum_{t=\lfloor T\tau_0 \rfloor+1}^{T} y_{t-1}\varepsilon_t \right] \\
&= \frac{1}{T} \sum_{t=\lfloor T\tau \rfloor+1}^{T} y_{t-1}\varepsilon_t + T(\beta_2 - \beta_1) \frac{1}{T^2} \sum_{t=\lfloor T\tau_0 \rfloor+1}^{T} y_{t-1}^2 \\
&\Rightarrow \sigma^2 \int_{\tau}^{1} \widetilde{J}_{\tau_0}(r) dB(r) - \delta\sigma^2 \int_{\tau_0}^{1} \left[\widetilde{J}_{\tau_0}(r)\right]^2 dr
\end{aligned}$$

where the limiting results are obtained from (a) and (b) in Lemma B.1 straightforwardly, from which we can also get

$$T^{-2} \sum_{t=1}^{\lfloor T\tau \rfloor} y_{t-1}^2 \Rightarrow \sigma^2 \int_0^{\tau} \left[\widetilde{J}_{\tau_0}(r)\right]^2 dr \quad \text{and} \quad T^{-2} \sum_{t=\lfloor T\tau \rfloor+1}^{T} y_{t-1}^2 \Rightarrow \sigma^2 \int_{\tau}^{1} \left[\widetilde{J}_{\tau_0}(r)\right]^2 dr.$$

Denoting $\Psi(\tau) = \left(\sum_{t=1}^{\lfloor T\tau \rfloor} y_{t-1}\eta_t\right)^2 / \sum_{t=1}^{\lfloor T\tau \rfloor} y_{t-1}^2 + \left(\sum_{t=\lfloor T\tau \rfloor+1}^{T} y_{t-1}\eta_t\right)^2 / \sum_{t=\lfloor T\tau \rfloor+1}^{T} y_{t-1}^2$, we then have

$$\Psi(\tau) \Rightarrow \sigma^2 \left\{ \frac{\left(\int_0^{\tau} \widetilde{J}_{\tau_0}(r) dB(r)\right)^2}{\int_0^{\tau} \left[\widetilde{J}_{\tau_0}(r)\right]^2 dr} + \frac{\left(\int_{\tau}^{1} \widetilde{J}_{\tau_0}(r) dB(r) - \delta \int_{\tau_0}^{1} \left[\widetilde{J}_{\tau_0}(r)\right]^2 dr\right)^2}{\int_{\tau}^{1} \left[\widetilde{J}_{\tau_0}(r)\right]^2 dr} \right\}.$$

Based on the results of (c) and (d) in Lemma B.1, we have

$$\begin{aligned}\frac{\left(\int_0^\tau \widetilde{J}_{\tau_0}(r)dB\left(r\right)\right)^2}{\int_0^\tau \left[\widetilde{J}_{\tau_0}(r)\right]^2 dr} &= \frac{\left(\left[\widetilde{J}_{\tau_0}(\tau)\right]^2-\left[\widetilde{J}_{\tau_0}(0)\right]^2-\tau+2\kappa\int_0^\tau\left[\widetilde{J}_{\tau_0}(r)\right]^2dr\right)^2}{4\int_0^\tau\left[\widetilde{J}_{\tau_0}(r)\right]^2dr}\\ &= \frac{\left(\left[\widetilde{J}_{\tau_0}(\tau)\right]^2-\left[\widetilde{J}_{\tau_0}(0)\right]^2-\tau\right)^2}{4\int_0^\tau\left[\widetilde{J}_{\tau_0}(r)\right]^2dr}+\kappa^2\int_0^\tau\left[\widetilde{J}_{\tau_0}(r)\right]^2dr\\ &\quad+\kappa\left(\left[\widetilde{J}_{\tau_0}(\tau)\right]^2-\left[\widetilde{J}_{\tau_0}(0)\right]^2-\tau\right)\end{aligned}$$

and

$$\begin{aligned}&\frac{\left(\int_\tau^1 \widetilde{J}_{\tau_0}(r)dB\left(r\right)-\delta\int_{\tau_0}^1\left[\widetilde{J}_{\tau_0}(r)\right]^2dr\right)^2}{\int_\tau^1\left[\widetilde{J}_{\tau_0}(r)\right]^2dr}\\ =\ &\frac{\left(\left[\widetilde{J}_{\tau_0}(1)\right]^2-\left[\widetilde{J}_{\tau_0}(\tau)\right]^2-(1-\tau)+2\kappa\int_\tau^1\left[\widetilde{J}_{\tau_0}(r)\right]^2dr\right)^2}{4\int_\tau^1\left[\widetilde{J}_{\tau_0}(r)\right]^2dr}\\ =\ &\frac{\left(\left[\widetilde{J}_{\tau_0}(1)\right]^2-\left[\widetilde{J}_{\tau_0}(\tau)\right]^2-(1-\tau)\right)^2}{4\int_\tau^1\left[\widetilde{J}_{\tau_0}(r)\right]^2dr}+\kappa^2\int_\tau^1\left[\widetilde{J}_{\tau_0}(r)\right]^2dr\\ &+\kappa\left(\left[\widetilde{J}_{\tau_0}(1)\right]^2-\left[\widetilde{J}_{\tau_0}(\tau)\right]^2-(1-\tau)\right)\end{aligned}$$

As a result,

$$\begin{aligned}\frac{\Psi\left(\tau\right)}{\sigma^2} \Rightarrow\ &\frac{\left(\left[\widetilde{J}_{\tau_0}(\tau)\right]^2-\left[\widetilde{J}_{\tau_0}(0)\right]^2-\tau\right)^2}{4\int_0^\tau\left[\widetilde{J}_{\tau_0}(r)\right]^2dr}+\frac{\left(\left[\widetilde{J}_{\tau_0}(1)\right]^2-\left[\widetilde{J}_{\tau_0}(\tau)\right]^2-(1-\tau)\right)^2}{4\int_\tau^1\left[\widetilde{J}_{\tau_0}(r)\right]^2dr}\\ &+\kappa^2\int_0^1\left[\widetilde{J}_{\tau_0}(r)\right]^2dr+\kappa\left(\left[\widetilde{J}_{\tau_0}(1)\right]^2-\left[\widetilde{J}_{\tau_0}(0)\right]^2-1\right).\end{aligned}$$

Following the same procedure above, when $\tau > \tau_0$, it can be proved that

$$\begin{aligned}\frac{\Psi\left(\tau\right)}{\sigma^2} \Rightarrow\ &\frac{\left(\left[\widetilde{J}_{\tau_0}(\tau)\right]^2-\left[\widetilde{J}_{\tau_0}(0)\right]^2-\tau\right)^2}{4\int_0^\tau\left[\widetilde{J}_{\tau_0}(r)\right]^2dr}+\frac{\left(\left[\widetilde{J}_{\tau_0}(1)\right]^2-\left[\widetilde{J}_{\tau_0}(\tau)\right]^2-(1-\tau)\right)^2}{4\int_\tau^1\left[\widetilde{J}_{\tau_0}(r)\right]^2dr}\\ &+\kappa^2\int_0^1\left[\widetilde{J}_{\tau_0}(r)\right]^2dr+\kappa\left(\left[\widetilde{J}_{\tau_0}(1)\right]^2-\left[\widetilde{J}_{\tau_0}(0)\right]^2-1\right).\end{aligned}$$

Therefore, deleting the common terms shared by the limit of $\Psi(\tau)$ when $\tau > \tau_0$ and $\tau \leq \tau_0$ which are independent of the choice of τ leads to the final in-fill asymptotic distribution of $\hat{\tau}_{LS}$ as

$$\begin{aligned}\hat{\tau}_{LS} &= \arg\max_{\tau\in(0,1)} \Psi(\tau) \\ &\Rightarrow \arg\max_{\tau\in(0,1)} \frac{\left[\left[\widetilde{J}_{\tau_0}(\tau)\right]^2 - \left[\widetilde{J}_{\tau_0}(0)\right]^2 - \tau\right]^2}{\int_0^\tau \left[\widetilde{J}_{\tau_0}(r)\right]^2 dr} + \frac{\left[\left[\widetilde{J}_{\tau_0}(1)\right]^2 - \left[\widetilde{J}_{\tau_0}(\tau)\right]^2 - [1-\tau]\right]^2}{\int_\tau^1 \left[\widetilde{J}_{\tau_0}(r)\right]^2 dr}.\end{aligned}$$

Proof of Theorem 5.1: With the use of the in-fill asymptotics given in Lemma B.2, the same procedure for the proof of Theorem 4.1 will lead to the result in Theorem 5.1. The details are omitted here.

References

[1] Bai, J., 1994. Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis* 15, 453-472.

[2] Bai, J., 1997. Estimation of a change point in multiple regression models. *Review of Economics and Statistics* 79, 551-563.

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

[4] 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.

[5] Liang, Y. L., Pang, T., Zhang, D., and Chong, T.T.L., 2017. Structural change in mildly integrated AR(1) models. *Econometric Theory* forthcoming.

[6] Jiang, L., Wang, X. H., and Yu, J., 2016. New Distribution Theory for the Estimation of Structural Break Point in Mean. *Journal of Econometrics* forthcoming.

[7] 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.

[8] 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.

[9] 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.

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

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

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

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

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

[15] Phillips, P.C.B. and Shi, S.P., 2017. Financial Bubble Implosion and Reverse Regression. *Econometric Theory* forthcoming.

[16] Phillips, P.C.B., Shi, S.P. and Yu, J., 2015a. Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review* 56, 1043-1078.

[17] Phillips, P.C.B., Shi, S.P. and Yu, J., 2015b. Testing For Multiple Bubbles: Limit Theory Of Real-Time Detectors. *International Economic Review* 56, 1079-1134.

[18] 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.

[19] 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.

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

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

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

[23] Tang, C.Y. and S.X. Chen, 2009, Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics* 149, 65-81.

[24] Wang, X. H. and Yu, J., 2016. Double asymptotics for explosive continuous time models. *Journal of Econometrics*, 193, 35-53.

[25] 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.

[26] Yu, J., 2012. Bias in the Estimation of the Mean Reversion Parameter in Continuous Time Models. *Journal of Econometrics* 169, 114-122.

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

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