Econometric Reviews

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lecr20

In-fill asymptotic theory for structural break point in autoregressions

Liang Jiang , Xiaohu Wang & Jun Yu

To cite this article: Liang Jiang , Xiaohu Wang & Jun Yu (2020): In-fill asymptotic theory for structural break point in autoregressions, Econometric Reviews

To link to this article: https://doi.org/10.1080/07474938.2020.1788822

Published online: 16 Jul 2020.

ECONOMETRIC REVIEWS
https://doi.org/10.1080/07474938.2020.1788822

Taylor & Francis
Taylor & Francis Group

In-fill asymptotic theory for structural break point in autoregressions

Liang Jiang[a], Xiaohu Wang[b,c], and Jun Yu[d]

[a]Fanhai International School of Finance, Fudan University, Shanghai, China; [b]School of Economics, Fudan University, Shanghai, China; [c]Shanghai Institute of International Finance and Economics, Shanghai, China; [d]School of Economics and Lee Kong Chian School of Business, Singapore Management University, Singapore

ABSTRACT

This article 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 article 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 roots 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. The article also proposes to use the highest density region to construct confidence intervals for structural break point.

KEYWORDS

Asymmetry; exact distribution; highest density region; long-span asymptotics; in-fill asymptotics; trimodality

JEL CLASSIFICATION

C11; C46

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 et al. (1987), Phillips et al. (2011), Phillips and Yu (2011), Homm and Breitung (2012), and Phillips et al. (2015a, 2015b). 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 et al. (2014) and Liang et al. (2018) 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

CONTACT Jun Yu yujun@smu.edu.sg School of Economics, Singapore Management University, 90 Stamford Rd, Singapore 178903.

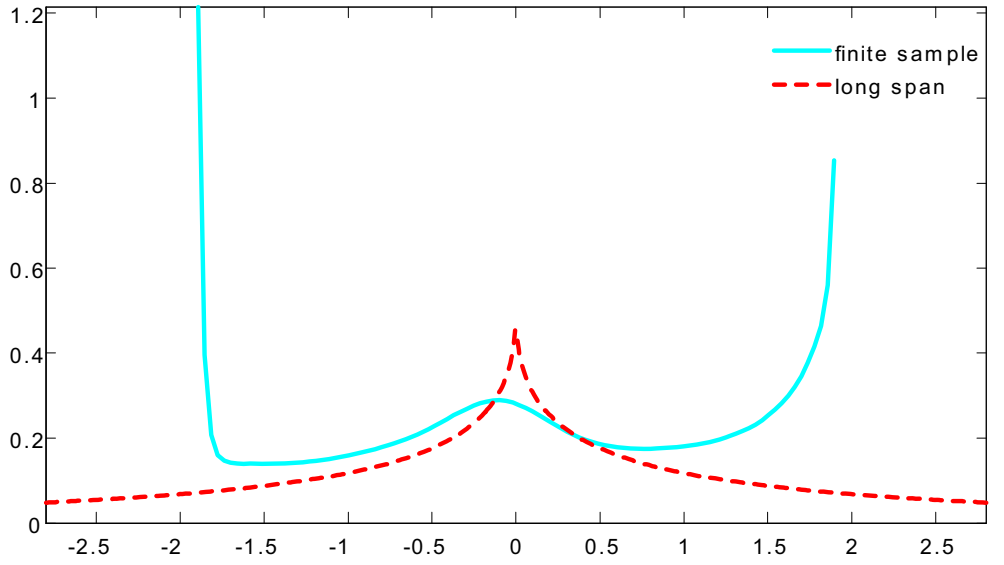

Figure 1. The pdf of the finite sample distribution of $\frac{T(\beta_2-\beta_1)^2}{\sigma^2}(\hat{\tau}_{LS,T}-\tau_0)$ when $T=200, \beta_1=0.5, \beta_2=0.38, \sigma=1$ and $\tau_0=0.5$ in Model (1) and the pdf of $\arg\max_{u\in(-\infty,\infty)}\{W(u)-\frac{1}{2}|u|\}$..

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 et al. (2011), Phillips and Yu (2011), Homm and Breitung (2012), Phillips et al. (2015a), and Phillips and Shi (2018), no asymptotic theory is available.

Second, to aggravate the matter, the derived asymptotic distribution often does not perform well in finite samples, even when both roots are much less than 1 (say 0.5 and 0.38, as will be shown in Fig. 1 below). It is discontinuous in the underlying AR(1) parameters as one or two roots pass the unity. 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 samples. Evidence from the simulations reported later strongly suggests that in many empirically relevant cases the long-span asymptotic 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]

[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.

[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.

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) 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 (2015) 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. Furthermore, it motivates Jiang, Wang and Yu (2018, JWY hereafter) to develop the in-fill asymptotic distribution in the estimation of structural break point in mean.

In this article, 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. 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 JWY (2018) 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 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 nonstandard and the density function is not available analytically. Hence, simulations are needed to obtain moments and quantiles. 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 article 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. In Section 7, we propose to use the highest density region to construct confidence intervals for the break point. Section 8 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, Bai et al. (1998), Bai (2000), Chong (2001), Pang et al. (2014) and Liang et al. (2018) 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 LS estimator is the focus of our article although other estimation methods have been also considered in the literature; see, for example, Harvey et al. (2020).

The model considered in these papers is

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

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 $1 \le k_0 < T$ 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

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

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 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

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.

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

$W(u)$ is a two-sided Brownian motion, whose definition and the meanings of other notations are introduced in Appendix A.

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 et al. (2014) and Liang et al. (2018) 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 article. 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. Another two interesting cases include changes from a mildly explosive process to a random walk or to a stationary process which are classified as "disturbing" and "smooth" crisis events in the literature (Huang et al., 2010; Phillips and Shi, 2018; Rosser, 2000).

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? Clearly these two asymptotic distributions have different expressions. Later we will report the evidence of large discrepancy between some of the long-span distributions.

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) for simulation evidence and Phillips (1987a) for analytical evidence in unit root models, Phillips (1987b) for evidence in local-to-unit root models, and Wang and Yu (2016) in mildly explosive models. 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 JWY (2018). 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 broken line in Fig. 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 samples 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 Fig. 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 $\beta_2 = 0.38$.[3] The finite sample distribution is obtained from simulated data with 100,000 replications. The red broken line in Fig. 1 plots the density of the long-span asymptotic distribution. In spite that a parameter setting is chosen to favor the long-span asymptotic theory, the long-span asymptotic distribution and the finite sample distribution are far from each other. In two aspects the finite sample distribution is remarkably different from the long-span asymptotic distribution $\left(\arg\max_{u\in(-\infty,\infty)}\left\{W(u)-\frac{|u|}{2}\right\}\right)$.[4] First, the finite sample distribution is asymmetric, indicating a downward 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 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).

[3]This parameter setting is chosen to favor the long-span asymptotic theory reported in the second row of Table 1 for two reasons. The first is that the two AR(1) coefficients are much smaller than one. The second is that the break size is moderately large.

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

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 $t \in [0,1]$,[5] $1_{[t>\tau_0]}$ 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 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.

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.

When a continuous record is available over an interval, no matter how short the interval is, the diffusion parameter can be estimated by the quadratic variation without estimation error (see, for example, Phillips and Yu, 2009). Therefore, if there is a break in σ, recursive calculation of the quadratic variation can perfectly recover the break point in σ. A break in mean can also be allowed in the model. In fact, the model considered in JWY (2018) allows for a structural break in mean. However, it is difficult for a continuous-time model to have a deterministic trend. Thus, it is not easy, if not impossible, to consider a break in the trend in our setup.

Assume that all parameters except for τ_0 are known. For any $\tau \in (0,1)$, the exact log-likelihood of Model (3) can then 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 $\hat{\tau}_{ML}$.

[5] A different length of time interval, such as $[0,N]$, may be assumed without qualitatively changing the results derived in the present paper. Tao et al. (2019) also consider a continuous time OU process with random coefficients where the randomness is generated by a Brownian motion.

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*

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

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

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

with the initial condition $\tilde{J}_{\tau_0}(0) = x(0)/\sigma$, and B(r) is a standard Brownian motion which is the same as in (3).[6]

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 $\tilde{J}_{\tau_0}(r)$ 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 . The feature is the same as that of the LS estimator of persistency parameter in a continuous time model obtained in Phillips (1987a).

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

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

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 $\tilde{J}_{\tau_0}(r)$ 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

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

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

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

[6]Note that δ is in the limiting distribution which cannot be got rid of as the sign of δ can be important to the shape of the distribution of $\hat{\tau}_{ML}$.

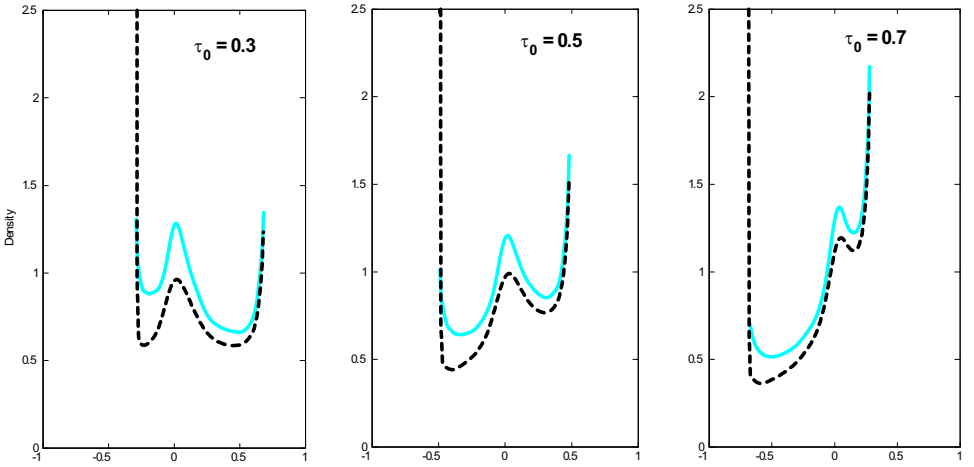

Figure 2. Densities of $\hat{\tau}_{ML} - \tau_0$ given in Eq. (7) when $\kappa = 138, \delta = -20, \sigma = 1$ and $\tau_0 = 0.3, 0.5, 0.7$, respectively. Solid lines are densities for $x(0) = 0.2$; broken lines are densities for $x(0) = 1$.

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

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

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

suggesting that on average $Z_1(u)$ has the unique maximum at the origin. Similarly, $E(Z_2(u))$ has a supremum at the origin. This property is similar to that of $E(W(u) - |u|/2)$, as explained in JWY (2018). That the expectation of the objection function in (7) is maximized at the origin explains why the origin is a mode in $\hat{\tau}_{ML} - \tau_0$. 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 by $-\tau_0$ and $1 - \tau_0$. In an argument similar to that in JWY (2018), there are two modes at the boundary points in the distribution of $\hat{\tau}_{ML}$.

In Fig. 2, we plot the density of $\hat{\tau}_{ML} - \tau_0$ 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 Fig. 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 $0(= \hat{\tau}_{ML} - \tau_0)$ 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 discrete time model has the form of

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

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, $t = 1,\ldots,T$ with T being the sample size, and $h = 1/T$.[7] 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

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

where $t = 1,\ldots,T$ 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

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

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

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

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

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 the LS estimators of β_1 and β_2, respectively. The LS estimator of the fractional break point is defined as

$$\hat{\tau}_{LS} = \hat{k}_{LS}/T. \tag{11}$$

Based on $\hat{k}_{LS}$, one can define

$$\hat{\kappa}_{LS} = -T\ln\left(\hat{\beta}_1\left(\hat{k}_{LS}\right)\right),\ \ \hat{\delta}_{LS} = T\ln\left(\hat{\beta}_2\left(\hat{k}_{LS}\right)\right) - T\ln\left(\hat{\beta}_1\left(\hat{k}_{LS}\right)\right). \tag{12}$$

Let the residuals be $\{\hat{\varepsilon}_t\}_{t=1}^{T}$. Then one can estimate σ by

$$\hat{\sigma}_{LS} = \sqrt{\sum_{t=1}^{T}\hat{\varepsilon}_t^2}. \tag{13}$$

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 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

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

$$y_t = \left(\beta_1 1_{[t\le k_0]} + \beta_2 1_{[t>k_0]}\right) y_{t-1} + \varepsilon_t,\ \ \varepsilon_t \overset{i.i.d.}{\sim} (0,\sigma^2), y_0 = x_0/\sqrt{h} = O_p(T^{1/2}). \tag{14}$$

Model (14) is almost the same as the model in (1) except for three important differences. First, the initial condition of y_t in (14) diverges at the rate of $T^{1/2}$ as $T \to \infty$, whereas the initial condition in Model (1) is set to be $o_p(T^{1/2})$ 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 (14), $\beta_1 = \exp\{-\kappa/T\} \to 1$ 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 (14) 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 samples.

The third difference lies in the order of break size. The break size is $\beta_2 - \beta_1 = O(T^{-1})$ in Model (14) while it is $O(T^{-\alpha})$ 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 $O(T^{-\alpha})$ 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 $O(T^{-1/2})$. JWY (2018) provided evidence that, when the break size is $O(T^{-1/2})$, the asymptotic distribution is closer to the finite sample distribution.[8] The present article 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*

$$\hat{\tau}_{LS} \Rightarrow \arg\max_{\tau\in(0,1)} \frac{\left\{\left[\tilde{J}_{\tau_0}(\tau)\right]^2 - \left[\tilde{J}_{\tau_0}(0)\right]^2 - \tau\right\}^2}{\int_0^\tau \left[\tilde{J}_{\tau_0}(r)\right]^2 dr} + \frac{\left\{\left[\tilde{J}_{\tau_0}(1)\right]^2 - \left[\tilde{J}_{\tau_0}(\tau)\right]^2 - [1-\tau]\right\}^2}{\int_\tau^1 \left[\tilde{J}_{\tau_0}(r)\right]^2 dr} \tag{15}$$

where $\tilde{J}_{\tau_0}(r)$, for $r \in [0,1]$, is the Gaussian process defined in (6) with the initial condition $\tilde{J}_{\tau_0}(0) = x_0/\sigma$, 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 (15) 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 $\tilde{J}_{\tau_0}(r)$, the in-fill asymptotic distribution given in (15) explicitly depends on the initial condition x_0/σ. Moreover, it also depends on the persistency parameters κ and δ. Since $\tilde{J}_{\tau_0}(r)$ is continuous in κ and δ, the in-fill asymptotic distribution in (15) should also be continuous in κ and δ. The feature is the same as that of the LS estimator of AR(1) parameter obtained in Phillips (1987a) in unit root models.

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

[8]To calculate the power of tests for structural break, Sowell (1996) considered a sequence of local alternatives which shrink to the null hypothesis at the rate of $O(T^{-1/2})$.

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

As in the exact distribution, both the asymmetry in $(-\tau_0, 1-\tau_0)$ and the asymmetry in $\tilde{J}_{\tau_0}(r)$ 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[9]:

$$x_t = \left(\beta_1 1_{[t \le k_0]} + \beta_2 1_{[t > k_0]}\right) x_{t-1} + u_t, x_0 = O_p(1), \tag{16}$$

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

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

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 $j = 0, \pm 1, \pm 2, ...,$ and $C(1) \equiv \sum_{j=0}^{\infty} c_j$. Note that the long-run variance of u_t goes to zero as $h \to 0$ since

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

It is also clear that Model (16) reduces to Model (8) if $c_j = 0$ for $j \ge 1$. Clearly an AR(p) model with a structural break occurring in the AR(1) coefficient is a special case of Model (16). No long-span asymptotic theory has been derived 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

$$y_t = \left(\beta_1 1_{[t \le k_0]} + \beta_2 1_{[t > k_0]}\right) y_{t-1} + u_t^*, \text{ with } y_0 = x_{0/}\sqrt{h} = O_p(T^{1/2}), \tag{17}$$

where

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

[9]This specification extends Model (1) of Phillips (1987b) by allowing for a structural break.

Define $\gamma^*(j) \equiv E(u_t^* u_{t-j}^*) = \gamma(j)/h$ 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 (17) will reduce to Model (14).

Theorem 5.1. *Consider the general discrete-time model with weakly dependent errors defined in (16). 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*

$$\hat{\tau}_{LS} \Rightarrow \arg\max_{\tau \in (0,1)} \frac{\left\{ \left[\tilde{J}_{\tau_0}(\tau)\right]^2 - \left[\tilde{J}_{\tau_0}(0)\right]^2 - \phi\tau \right\}^2}{\int_0^\tau \left[\tilde{J}_{\tau_0}(r)\right]^2 dr} + \frac{\left\{ \left[\tilde{J}_{\tau_0}(1)\right]^2 - \left[\tilde{J}_{\tau_0}(\tau)\right]^2 - \phi[1-\tau] \right\}^2}{\int_\tau^1 \left[\tilde{J}_{\tau_0}(r)\right]^2 dr} \tag{18}$$

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

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 (18) explicitly depends on x_0 and σ through the initial value of $\tilde{J}_{\tau_0}(0) = x_0/[C(1)\sigma]$. 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 (18) becomes the same as the one given in (15) for the model with i.i.d. errors. Hence, we expect the in-fill asymptotic distribution in (18) to be asymmetric and trimodal.

6. Monte Carlo results

In this section, we first design two Monte Carlo experiments for cases where long-span asymptotic distributions have been developed in the literature so that their performance in fitting finite sample distributions can be compared to that of the in-fill asymptotic distributions derived in the article. In both cases, 100,000 sample paths are generated from Model (14) with $\sigma = 1$, $\varepsilon_t \overset{i.i.d.}{\sim} N(0,1)$, $T = 200$ (i.e. $h = 1/200$), $x_0 = 1$ and $\tau_0 = 0.3, 0.5, 0.7$.[10]

To obtain the density of the in-fill distribution by simulation, we first generate $\tilde{J}_{\tau_0}(r)$ with $r \in [0,1]$ from its exact discretization at a very fine grid; then plug $\tilde{J}_{\tau_0}(r)$ and other parameters into the in-fill distribution given in Theorem 4.1 to obtain the argmax over [0.1]. This is one random draw from the in-fill distribution. We repeat the exercise 100,000 times to obtain a sufficiently large number of i.i.d. draws from the in-fill distribution. To obtain the density of the long-span distribution, we adopt the same procedure. For example, for case 1 in Table 1, we first generate $W(u)$ with $u \in [-C, C]$ at a very fine grid, where C is a large constant; then obtain the argmax of $W(u) - \frac{|u|}{2}$ over $[-C, C]$. In both cases, the grid size is set at 0.001.

In the first experiment, we set $\kappa = 138$ and $\delta = 55$, which makes the AR roots in Model (14) on different sides of the break to be $\beta_1 = 0.5$ and $\beta_2 = 0.38$. To use the long-span asymptotic distribution, we assume that the AR process switches from a stationary root to a different stationary root. This assumption seems to be weak and should favor the long-span asymptotic distribution because β_1 and β_2 are much smaller than one. The long-span asymptotic distribution for this case has been developed as in (A.1) in Appendix A. Figure 3 plots the three densities, corresponding to the long-span distribution, the in-fill distribution, and the finite-sample distribution.

[10]Simulation studies for cases where ε_t follow other distributions, such as the t distribution, have also been carried out. It is found that all the findings under the normality as explained below are robust to the distribution assumption. To save space, we decide not to report the results.

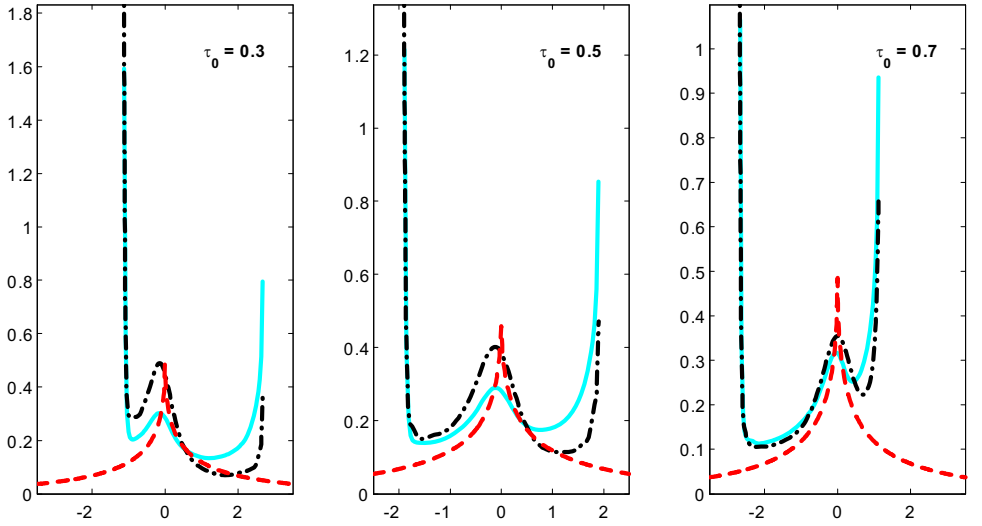

Figure 3. Densities of $\frac{T(\beta_2-\beta_1)^2}{1-\beta_1^2}(\hat{\tau}_{LS}-\tau_0)$ when $\beta_1 = 0.5$, $\beta_2 = 0.38$ with $x_0/\sigma = 1$. Solid lines are finite sample distributions when $T = 200$; dot-dashed lines are in-fill densities from Theorem 4.1; and dashed lines are long-span limiting distributions given in (A.1).

As it can be seen clearly, the densities of the long-span distributions are remarkably different with those of the finite sample distributions. There are two key characteristics in the finite sample distributions which are missing in the long-span distributions. First, the finite sample distributions are asymmetric about 0 even when $\tau_0 = 1/2$. Second, the finite sample distributions have trimodality in general. One mode is at the origin and the other two are at the two boundary points. On the other hand, the densities of the in-fill distributions given in Theorem 4.1 well capture the two key characteristics, and hence, provide much better approximations to the finite sample distributions.

In the second experiment, we set $\kappa = 21$ and $\delta = -21$, implying $\beta_1 = 0.9$ and $\beta_2 = 1$. Depending on whether the AR root $\beta_1 = 0.9$ is treated as a stationary root or a mildly stationary root, there are two different long-span asymptotic distributions that can be used. If it is treated as a stationary root, then the long-span asymptotic distribution is the one developed in (A.2). If it is treated as a mildly stationary root, then the long-span asymptotic distribution is the one developed in (A.6). Unfortunately, the literature provides no guidance about which one should be used in practice. Figure 4 plots both long-span asymptotic distributions, together with the densities of the finite sample distributions and in-fill asymptotic distributions given in Theorem 4.1.[11] Several features can be found. First, there is a big gap between the two long-span asymptotic distributions, making the choice of a long-span theory critical. Second and more importantly, both long-span asymptotic distributions are remarkably different from the finite sample distributions. Third, our in-fill asymptotic distributions very well approximate the finite sample distributions.

It is worth pointing out that the shape of finite sample distribution depends on break size. If we increase the break size in simulations, the two modes at the boundary points would become less pronounced. However, as argued by Elliott and Müller (2007), small breaks are important for two reasons. First, breaks that are small in a statistical sense could be a very large one in an economic sense. Second, a small break size is prevalent in economic applications. This is particularly true for time series which can be well fitted by AR models. For example, what commonly seen in

[11]The rates of convergency for the two long-span asymptotic distributions are different. Therefore, we draw densities of $\hat{\tau}_{LS} - \tau_0$ here. When plotting the long-span densities of (25), we set $c = 1$ and $T^{\alpha} = 20$.

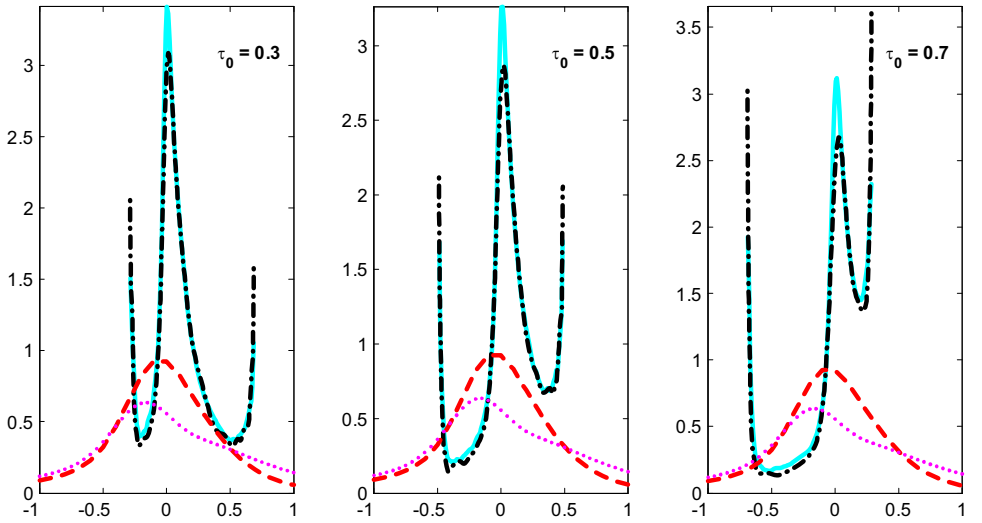

Figure 4. The pdf of $\hat{\tau}_{LS} - \tau_0$ when $\beta_1 = 0.9$, $\beta_2 = 1$ with $x_0/\sigma = 1$. Solid lines are finite sample distributions when $T = 200$; dot-dashed lines are in-fill densities from Theorem 4.1; dashed lines are long-span limiting distributions given in (A.2); and dotted lines are long-span limiting distributions given in (A.6).

practice is some time series changing from a near unit root process to a unit root process and vice versa. The case of changing from a very stationary process to a unit root process is rare.

The level of trimodality could also be reduced if the break point is estimated strictly inside the sample by truncating the first and last ends of the sample. This method is sometimes used in practice. However, as long as the break size is small, trimodality may still be observed with two modes at the truncation points. More importantly, the finite sample distribution remains seriously asymmetric. As a result, the in-fill asymptotic distribution continues providing better approximations than the long-span asymptotic distribution.

We now design three more experiments to investigate the performance of in-fill asymptotic distributions in approximating finite sample distributions when no long-span theory is available. In the third experiment, we set $\kappa = 0$ and $\delta = -6$, implying $\beta_1 = 1$ and $\beta_2 = 1.03$. This case is closely related to the literature of bubble testing where the break point is often referred to as the bubble origination date; see, for example, Phillips et al. (2011), Phillips and Yu (2011), and Phillips et al. (2015a, 2015b). In the fourth experiment, we set $\kappa = -6$ and $\delta = 6$, implying $\beta_1 = 1.03$ and $\beta_2 = 1$. In the fifth experiment we set $\kappa = -6$ and $\delta = 12$, implying $\beta_1 = 1.03$ and $\beta_2 = 0.97$. The break point in these two experiments is often referred to as the bubble termination date; see, for example, Phillips and Shi (2018), and Harvey et al. (2020).

The densities of the finite sample distributions and the in-fill distributions for these three experiments are plotted in three panels of Fig. 5. Several observations can be found. First, comparing with those in Figs. 3 and 4, the estimates of the break date $\hat{\tau}_{LS}$ become more concentrated around the true value τ_0. This is because an explosive AR process is involved in either side of the break.[12] Second, the trimodality almost disappears. Third, the asymmetry remains, especially in the upper panel. In the other panels, if we draw densities over the interval (–0.2, 0.2), the asymmetry can be easily seen. Fourth, in all cases, the in-fill asymptotic distributions provide excellent approximations to the finite sample distributions.

[12]This finding is reasonable as an explosive process is very different from a unit root process or a stationary process and an explosive process increases at an exponential rate. Hence, once an explosive process appears in one side of the break, the break is easier to identify even when the break size is small.

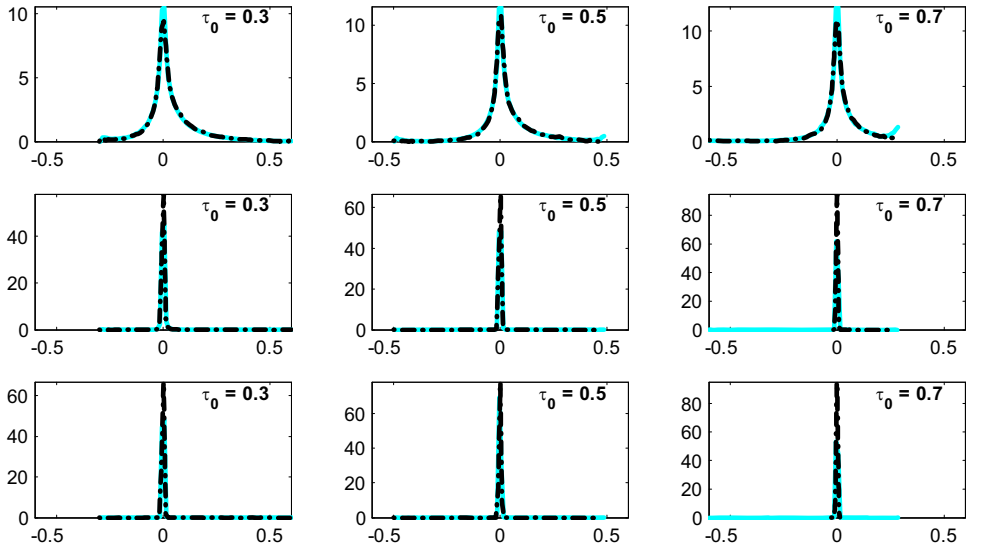

Figure 5. Densities of $\hat{\tau}_{LS} - \tau_0$ when $(\beta_1 = 1, \beta_2 = 1.03), (\beta_1 = 1.03, \beta_2 = 1)$, and $(\beta_1 = 1.03, \beta_2 = 0.97)$ are in the upper, middle and lower panel respectively. Solid lines are finite sample distributions when $T = 200$; dot-dashed lines are in-fill densities from Theorem 4.1.

We now turn our attention to the bias of $\hat{\tau}_{LS}$, which might be a serious problem given the findings from the Monte Carlo simulations. Considering that our in-fill distribution is closer to the finite sample distribution 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 root in Model (14) 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. Several conclusions can be drawn. 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 Fig. 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, corresponding to the case where the AR process switches from a stationary one to another stationary one, is symmetric about τ_0 as shown in (A.1) in Appendix A. Third, the in-fill asymptotic distribution approximates the true bias well in most 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. First, the LS estimator still suffers from severe bias problem in all cases. 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 et al., 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

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 $T = 200$.

		β_2				
τ_0	β_1 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

The number of replications is 10,000.

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., $E(\hat{\tau}_{LS} - \tau_0)$ as a function of τ_0) implied by the in-fill asymptotics for Model (14) and examine the sensitivity of the function to the initial condition.

Figures 6 and 7 plot the bias function when $x_0 = 0.2, 0.4, 0.6, 0.8, 1$ and $\sigma = 1$. Figure 6 corresponds to the case where $\beta_1 = 0.9$, $\beta_2 = 1$; Fig. 7 to $\beta_1 = 1$, $\beta_2 = 1.03$. 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, Fig. 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 insensitive of the initial condition.

7. Constructing confidence intervals

Constructing statistical confidence inferences about τ based on the in-fill asymptotic theory is challenging since the limiting distribution is non-pivotal, nonstandard, and has the features of asymmetry and trimodality. Clearly, a confidence interval which is symmetric around the median cannot be the optimal choice. One possible solution is to use the highest density region (HDR) implied by the in-fill asymptotic distribution.

Let $f(x)$ be the density function of a random variable X. The $100(1-\alpha)\%$ HDR is defined as

$$R(f_\alpha) = \{x : f(x) \geq f_\alpha\},$$

where f_α is the largest constant such that $\Pr(X \in R(f_a)) \geq 1 - \alpha$. The HDR is designed to contain points with relatively high density in $f(x)$ and is considered as the most credible region. Moreover, the HDR gives the shortest confidence interval on the density for any fixed confidence level.

If $f(x)$ is symmetric about a unique mode, the HDR contains a single interval with endpoints a and b that are symmetric about the mode. In this case, $f(a) = f(b)$. Whereas, if $f(x)$ is asymmetric, the HDR is asymmetric about the mode. It skews to the side with a fatter tail. Moreover, if $f(x)$ has multiple modes, the HDR may consist of several disjoint intervals. As shown in Hyndman (1996), the HDR is especially useful when $f(x)$ displays asymmetry and multimodality, and therefore, is suitable for the developed in-fill asymptotic distribution.

To obtain the HDR based on the in-fill asymptotic distribution, we need to make random draws from the in-fill asymptotic distribution. Since the in-fill asymptotic distribution depends on unknown parameters τ, κ, δ, and σ, we first calculate $\hat{\tau}_{LS}, \hat{\kappa}_{LS}, \hat{\delta}_{LS}$, and $\hat{\sigma}_{LS}$ based on (11), (12), and (13). We use $\hat{\tau}_{LS}, \hat{\kappa}_{LS}, \hat{\delta}_{LS}$, and $\hat{\sigma}_{LS}$ to replace τ, κ, δ, and σ in the in-fill asymptotic

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 $T = 200$.

	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

The number of replications is 10,000.

distribution given in Theorem 4.1. Then, we can make random draws from the plug-in in-fill asymptotic distribution. The $100(1-\alpha)\%$ HDR can be computed from the empirical density of the random draws with the use of the density quantile algorithm of Hyndman (1996).[13]

The long-span asymptotic distributions developed in the literature also depend on the unknown parameters τ, κ, δ, and σ (or a subset of them in some cases). Constructing confidence intervals based on various long-span asymptotic distributions also need to plug-in the estimates, $\hat{\tau}_{LS}, \hat{\kappa}_{LS}$, and $\hat{\delta}_{LS}$. As the in-fill asymptotic distribution is much closer to the finite sample distribution than any of the long-span asymptotic distributions, the HDR obtained from the plug-in in-fill asymptotic distribution is expected to be more powerful in finite samples than that based on the plug-in long-span asymptotic distributions when the long-span asymptotic distributions are available. In most cases, the long-span asymptotic distributions are unknown and hence, confidence intervals cannot be obtained via the long-span asymptotic distributions. However, it is not clear if the proposed HDR generates the most powerful confidence interval for hypothesis testing. This issue needs to be examined carefully, but it is beyond the scope of the present article.

8. Conclusions

This article 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 article provides a new limiting theory for the break point estimate in the AR(1) model with independent errors as well as the model with weakly dependent errors. It develops an in-fill

[13]An open-source package to calculate the HDR from random draws, which is implemented in R, is available at https://pkg.robjhyndman.com/hdrcde/reference/hdr.html.

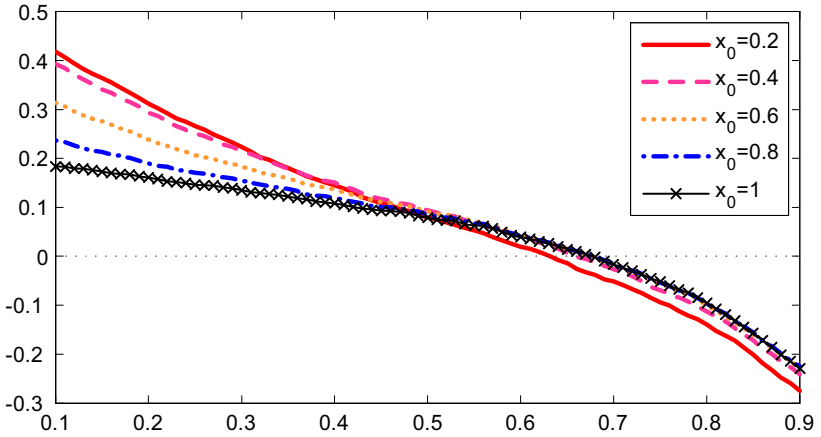

Figure 6. Bias functions implied by the in-fill asymptotic distribution given in Theorem 4.1 when $\beta_1 = 0.9$ and $\beta_2 = 1$ with various initial conditions.

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.

There is a cost to applying the in-fill limiting theory. The asymmetry and trimodality of the in-fill distribution make the construction of confidence intervals difficult. In the article, we suggest using the HDR to construct confidence intervals.

It is possible to extend the in-fill asymptotic theory to vector autoregressive (VAR) processes. We can begin with a multivariate diffusion process over the interval $[0,1]$ with one-time structural break in the drift function as

$$dX(t) = -\left(\kappa + \delta 1_{[t>\tau_0]}\right)X(t)dt + \Sigma dB(t),$$

where $X(t)$ is an $N \times 1$ random vector, κ, δ, and Σ are three $N \times N$ parameter matrices, $B(t)$ denotes an N -dimensional standard Brownian motion, and τ_0 is the time point at which the drift matrix changes. Assume that the process $X(t)$ has discrete-time observations at T equally spaced points over the interval $[0,1]$ and denote these observations as $\{X_t\}$ for $t = 1, 2, ..., T$. The discrete-time process X_t follows a VAR(1) model with one-time break in the autoregressive matrix as

$$X_t = \left(B_1 1_{[t\le k_0]} + B_2 1_{[t>k_0]}\right)x_{t-1} + \sqrt{h}\Omega^{1/2}\varepsilon_t, \varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, I_N), \;\; x_0 = O_p(1), \tag{19}$$

where $B_1 = \exp\{-\kappa/T\}$ and $B_2 = \exp\{-(\kappa+\delta)/T\}$ are the coefficient matrices before and after the break, $k_0 = T\tau_0$ is an integer denoting the break point, $h = 1/T$, Ω is an N-dimensional positive definite matrix, I_N is the N-dimensional identity matrix. In the VAR(1) model, the persistency and the cointegrating properties of elements in X_t before and after the break are determined by the eigenvalues and the ranks of B_1 and B_2, respectively. As long as $B_1 \neq B_2$, a break occurs in the model.[14] Letting $h = 1/T$, the in-fill asymptotic distribution of the LS estimator of

[14]In the VAR(1) model, various possibilities of structural breaks appear. For example, shifts among the stationary roots, unit roots, explosive roots are all possible. Moreover, if the rank of B_1 is different from that of B_2 and both are less than N, a change in the number of cointegrating relationships happens.

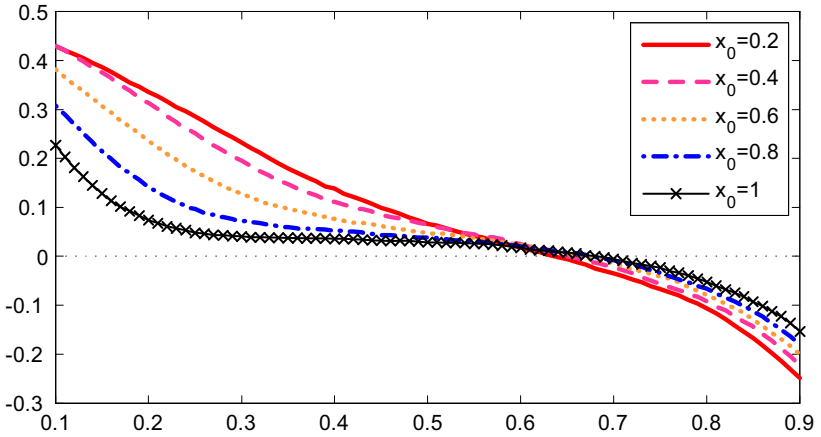

Figure 7. Bias functions implied by the in-fill asymptotic distribution given in Theorem 4.1 when $\beta_1 = 1$ and $\beta_2 = 1.03$ with various initial conditions.

the break point $\hat{\tau}_{LS}$ can be obtained. The in-fill asymptotic distribution should depend on a multivariate Gaussian process $\tilde{J}_{\tau_0}(r)$ that is a function of matrices κ and δ.

However, the in-fill asymptotic theory of $\hat{\tau}_{LS}$ for the VAR(1) model is expected to be more complicated than that in the univariate AR(1) model. There is a variety of possibilities of structural breaks that may lead to different in-fill asymptotics. For example, consider B_1 and B_2 that make X_t to be cointegrated both before and after the break. The change from B_1 and B_2 may cause the following three scenarios: (i) no change in the cointegrating relationships and their numbers, but only changes in the loadings for some cointegrating relationships, (ii) no change in the number of the cointegrating relationships, but changes in the values of some cointegrating vectors, and (iii) the change in cointegrating numbers. It is well-known in the literature that the estimators of cointegrating vectors, cointegrating ranks, and the loadings of cointegrating relationships have different convergence rates (see, for example, Hansen (2003); Johansen (1988, 1991) Wang (2019)). Hence, it is possible that the three scenarios lead to different in-fill asymptotic theories for $\hat{\tau}_{LS}$.

Moreover, comparing to univariate models, the implementation of the in-fill asymptotic distribution of $\hat{\tau}_{LS}$ for the VAR(1) model is also more complicated. The implementation of the in-fill asymptotic distribution will depend on κ and δ in the corresponding continuous-time model. However, the identification and estimation methods for κ and δ will be different, depending on whether the VAR process X_t is stationary, pure unit root, or cointegrated (see, for example, Hansen and Sargent (1983); Kessler and Rahbek (2004); McCrorie (2003, 2009); Phillips (1973)). The full study of the in-fill asymptotics for VAR models deserves a full-length paper and will be reported in our future works.

Acknowledgment

We acknowledge helpful comments from the editor, two referees, Peter C.B. Phillips, Yichong Zhang, and seminar participants at UC, Berkeley, Princeton, Singapore Management University and UC, San Diego.

Appendix 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 (2001) 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

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

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 long-span asymptotic distribution of $\hat{\tau}_{LS,T}$ takes the form of

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

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

$$T^2(\beta_{2T} - 1)^2(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg\max_{u \in (-\infty, \infty)} \left\{ \frac{W(u)}{W_3(\tau_0)} - \frac{1}{2}|u| \right\}, \tag{A.3}$$

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

Pang et al. (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

$$T(\beta_{2T} - \beta_1)(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg\max_{u \in (-\infty, \infty)} \left\{ \frac{W_b^*(u)}{R_1} - \frac{1}{2}|u| \right\}, \tag{A.4}$$

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

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

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

Pang et al. (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 $y_0 = o_p(\sqrt{T})$, $\sqrt{T}(\beta_{2T} - \beta_{1T}) \to 0$ with $T^{3/4}(\beta_{2T} - \beta_{1T}) \to \infty$, they proved that $\hat{\tau}_{LS,T}$ has the long-span asymptotic distribution as

$$T^2(\beta_2 - \beta_{1T})^2(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg\max_{u \in (-\infty, \infty)} \left\{ \frac{W(u)}{\exp(c(1 - \tau_0)) G(W_1, c, \tau_0)} - \frac{1}{2}|u| \right\}, \tag{A.5}$$

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

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

Liang et al. (2018) studied Model (1) with $\beta_{1T} = 1 - c/T^{\alpha}$ and $\beta_2 = 1$ where c is a positive constant and $\alpha \in (0,1)$.[15] In this case the AR model switches from a mildly stationary root to a unit root. Under the assumptions that $y_0 = o_p\left(\sqrt{T^{\alpha}}\right)$, they derived a long-span asymptotic distribution of $\hat{\tau}_{LS,T}$ as

$$\frac{cT}{T^{\alpha}}(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg\max_{u\in(-\infty,\infty)} \left\{\frac{W_c^*(u)}{R_c} - \frac{1}{2}|u|\right\}, \tag{A.6}$$

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. (2018) 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 $y_0 = o_p\left(\sqrt{T^{\alpha}}\right)$, $\sqrt{T}/T^{\alpha} \to 0$ 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

$$\frac{c^2T^2}{T^{2\alpha}}(\hat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg\max_{u\in(-\infty,\infty)} \left\{\frac{W(u)}{W_1(\tau_0)} - \frac{1}{2}|u|\right\}, \tag{A.7}$$

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

Appendix B: Proofs

Lemma B.1. *Consider the process y_t defined in (14) with the dynamics*

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

When $T = 1/h \to \infty$ with a fixed $\tau_0 = k_0/T$, for any $\tau \in [0,1]$,

(a) $T^{-1}\sum_{t=1}^{\lfloor T\tau\rfloor} y_{t-1}\varepsilon_t \Rightarrow \sigma^2\int_0^{\tau}\tilde{J}_{\tau_0}(r)dB(r)$;
(b) $T^{-2}\sum_{t=1}^{\lfloor T\tau\rfloor} y_{t-1}^2 \Rightarrow \sigma^2\int_0^{\tau}\left[\tilde{J}_{\tau_0}(r)\right]^2 dr$;
(c) $\left[\tilde{J}_{\tau_0}(\tau)\right]^2 - \left[\tilde{J}_{\tau_0}(0)\right]^2 = 2\int_0^{\tau}\tilde{J}_{\tau_0}(r)dB(r) - 2\int_0^{\tau}(\kappa + \delta 1_{[r>\tau_0]})\left[\tilde{J}_{\tau_0}(r)\right]^2 dr + \tau$;
(d) $\left[\tilde{J}_{\tau_0}(1)\right]^2 - \left[\tilde{J}_{\tau_0}(\tau)\right]^2 = 2\int_{\tau}^{1}\tilde{J}_{\tau_0}(r)dB(r) - 2\int_{\tau}^{1}(\kappa + \delta 1_{[r>\tau_0]})\left[\tilde{J}_{\tau_0}(r)\right]^2 dr + (1-\tau)$,

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

Lemma B.2. *Consider the process y_t defined in (17) with the dynamics*

$$y_t = \left(\beta_1 1_{[t\le k_0]} + \beta_2 1_{[t>k_0]}\right)y_{t-1} + u_t^*, y_0 = x_{0/}\sqrt{h}$$

where

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

Define $\gamma^*(j) \equiv E(u_t^* u_{t-j}^*)$ for $j = 0, \pm1, \pm2, \ldots$ 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) $T^{-2}\sum_{t=1}^{\lfloor T\tau\rfloor} y_{t-1}^2 \Rightarrow [C(1)\sigma]^2\int_0^{\tau}\left[\tilde{J}_{\tau_0}(r)\right]^2 dr$;
(b) $T^{-1}\sum_{t=1}^{\lfloor T\tau\rfloor} y_{t-1}u_t^* \Rightarrow [C(1)\sigma]^2\int_0^{\tau}\tilde{J}_{\tau_0}(r)dB(r) + (\tau/2)\left\{[C(1)\sigma]^2 - \gamma^*(0)\right\}$;

[15] Following Phillips and Magdalinos (2007), Liang et al. (2018) used k_T instead of T^{α} with the assumption that $k_T \to \infty$ and $k_T/T \to 0$.

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

Proof of Lemma B.1. When $\tau \leq \tau_0$, the process y_t for $t = 1, 2, ..., \lfloor T\tau \rfloor$ 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

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

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

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

where the first equation comes from it is lemma.

Proof of Lemma B.2. When $\tau \leq \tau_0$, the process y_t for $t = 1, 2, ..., \lfloor T\tau \rfloor$ 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

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

where the third equation is obtained by deleting the terms independent of the choice of τ but appearing in the second equation. Applying It'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,

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

We then have

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

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 $\tilde{J}_{\tau_0}(t)=x(t)/\sigma^2$ which can be obtained from the definitions of $\tilde{J}_{\tau_0}(t)$ 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

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

where $\hat{\beta}_1(k)=\sum_{t=1}^{k}y_ty_{t-1}/\sum_{t=1}^{k}y_{t-1}^2$, $\hat{\beta}_2(k)=\sum_{t=k+1}^{T}y_ty_{t-1}/\sum_{t=k+1}^{T}y_{t-1}^2$, and $y_t=x_t/\sqrt{h}$ is defined in (14). 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 Eq. (14), we have

$$y_t=\beta_1y_{t-1}+(\beta_2-\beta_1)1_{[t>k_0]}y_{t-1}+\varepsilon_t=\beta_1y_{t-1}+\eta_t$$

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 $MY_-=0_{T\times 1}$. Note that

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

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,

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

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 $\tau \le \tau_0$, we have

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

and

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

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

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

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

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

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

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

and

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

As a result,

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

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

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

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

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

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.

Funding

Jiang gratefully acknowledges the financial support from MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Project No.18YJC790063). Part of this article was done when Wang worked at the Department of Economics, The Chinese University of Hong Kong and Wang acknowledges the support from the Hong Kong Research Grants Council General Research Fund under No. 14503718. Yu thanks the financial support from the Lee Foundation and the Singapore Ministry of Education for Academic Research Fund under grant number MOE2013-T3-1-009.

References

Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis* 15(5):453–472. doi:10.1111/j.1467-9892.1994.tb00204.x

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics* 79(4):551–563. doi:10.1162/003465397557132

Bai, J. (2000). Vector autoregressive models with structural change in regression coefficients and in variance-covariance matrix. *Annals of Economics and Finance* 1:303–339.

Bai, J., Lumsdaine, R. L., Stock, J. H. (1998). Testing for and dating common breaks in multivariate time series. *Review of Economic Studies* 65(3):395–432. doi:10.1111/1467-937X.00051

Chong, T. T. L. (2001). Structural change in AR (1) models. *Econometric Theory* 17(1):87–155. doi:10.1017/S0266466601171045

Evans, G. B. A., Savin, N. E. (1981). Testing for unit roots: 1. *Econometrica* 49(3):753-779. doi:10.2307/1911521

Elliott, G., Müller, U. K. (2007). Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics* 141(2):1196–1218. doi:10.1016/j.jeconom.2007.02.001

Hansen, L. P., Sargent, T. J. (1983). The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica* 51(2):377–388. doi:10.2307/1911996

Hansen, P. R. (2003). Structural changes in the cointegrated vector autoregressive model. *Journal of Econometrics* 114(2):261–295. doi:10.1016/S0304-4076(03)00085-X

Harvey, D. I., Leybourne, S. J., Zu, Y. (2020). Sign-based unit root tests for explosive financial bubbles in the presence of nonstationary volatility. *Econometric Theory* 36(1):122–169. doi:10.1017/S0266466619000057

Homm, U., Breitung, J. (2012). Testing for speculative bubbles in stock markets: a comparison of alternative methods. *Journal of Financial Econometrics* 10(1):198–231. doi:10.1093/jjfinec/nbr009

Huang, W., Zheng, H., Chia, W. M. (2010). Financial crises and interacting heterogeneous agents. *Journal of Economic Dynamics and Control* 34(6):1105–1122. doi:10.1016/j.jedc.2010.01.013

Hyndman, R. J. (1996). Computing and graphing highest density regions. *The American Statistician* 50(2):120–126. doi:10.2307/2684423

Jiang, L., Wang, X. H., Yu, J. (2018). New distribution theory for the estimation of structural break point in mean. *Journal of Econometrics* 205(1):156–176. doi:10.1016/j.jeconom.2018.03.009

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12(2–3):231–254. doi:10.1016/0165-1889(88)90041-3

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59(6):1551–1580. doi:10.2307/2938278

Kessler, M., Rahbek, A. (2004). Identification and inference for multivariate cointegrated and ergodic Gaussian diffusions. *Statistical Inference for Stochastic Process* 36:153–188.

Liang, Y. L., Pang, T., Zhang, D., Chong, T. T. L. (2018). Structural change in mildly integrated AR(1) models. *Econometric Theory* 34(5):985–1017. doi:10.1017/S0266466617000317

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.

Mankiw, N. G., Miron, J. A., 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.

McCrorie, J. R. (2003). The problem of aliasing in identifying finite parameter continuous time stochastic models. *Acta Applicandae Mathematicae* 79(1/2):9–16. doi:10.1023/A:1025858121378

McCrorie, J. R. (2009). Estimating continuous time models on the basis of discrete data via an exact discrete analog. *Econometric Theory* 25(4):1120–1137. doi:10.1017/S0266466608090452

Pang, T., Zhang, D., 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. doi:10.1111/jtsa.12055

Perron, P. (1991). A continuous time approximation to the unstable first order autoregressive processes: The case without an intercept. *Econometrica* 59(1):211–236. doi:10.2307/2938247

Phillips, P. C. B. (1973). The problem for identification in finite parameter continuous time models. *Journal of Econometrics* 1(4):351–362. doi:10.1016/0304-4076(73)90021-3

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica* 55(2):277–301. doi:10.2307/1913237

Phillips, P. C. B. (1987b). Toward a unified asymptotic theory for autoregression. *Biometrika* 74(3):535–547. doi:10.1093/biomet/74.3.535

Phillips, P. C. B. (1991). To criticize the critics: An objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics* 6(4):333–364. doi:10.1002/jae.3950060402

Phillips, P. C. B., Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136(1):115–130. doi:10.1016/j.jeconom.2005.08.002

Phillips, P. C. B., Shi, S. P. (2018). Financial bubble implosion and reverse regression. *Econometric Theory* 34(4):705–753. doi:10.1017/S0266466617000202

Phillips, P. C. B., Shi, S. P., Yu, J. (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review* 56(4):1043–1078. doi:10.1111/iere.12132

Phillips, P. C. B., Shi, S. P., Yu, J. (2015b). Testing for multiple bubbles: Limit theory of real-time detectors. *International Economic Review* 56(4):1079–1134. doi:10.1111/iere.12131

Phillips, P. C. B., Wu, Y., Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review* 52(1):201–226. doi:10.1111/j.1468-2354.2010.00625.x

Phillips P. C. B., Yu J. (2009). Maximum likelihood and Gaussian estimation of continuous time models in finance. In: Mikosch, T., Kreiß, JP., Davis, R., Andersen, T., eds. *Handbook of Financial Time Series*. Berlin, Heidelberg: Springer, pp. 497–530.

Phillips, P. C. B., Yu, J. (2011). Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics* 2(3):455–491. doi:10.3982/Qe82

Rosser, J. B. (2000). *From Catastrophe to Chaos: A General Theory of Economic Discontinuities: Mathematics, Microeconomics and Finance*, Vol. 1. New York City: Springer.

Sims, C. A. (1988). Bayesian skepticism on unit root econometrics. *Journal of Economic Dynamics and Control* 12(2–3):463–474. doi:10.1016/0165-1889(88)90050-4

Sims, C. A., Uhlig, H. (1991). Understanding unit rooters: A helicopter tour. *Econometrica* 59(6):1591–1599. doi:10.2307/2938280

Sowell, F. (1996). Optimal tests for parameter instability in the generalized method of moments framework. *Econometrica* 64(5):1085–1107. doi:10.2307/2171957

Tang, C. Y., Chen, S. X. (2009). Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics* 149(1):65–81. doi:10.1016/j.jeconom.2008.11.001

Tao, Y., Phillips, P. C. B., Yu, J. (2019). Random coefficient continuous systems: Testing for extreme sample path behaviour. *Journal of Econometrics* 209(2):208–237. doi:10.1016/j.jeconom.2019.01.002

Wang, X. (2019). *Estimating the Persistency Matrix of Multivariate Diffusion Processes, Working Paper*. Chinese University of Hong Kong.

Wang, X, Yu, J. (2016). Double asymptotics for explosive continuous time models. *Journal of Econometrics* 193(1): 35–53. doi:10.1016/j.jeconom.2016.02.014

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(3):1321–1328. doi:10.1214/aos/1176350509

Yu, J. (2012). Bias in the estimation of the mean reversion parameter in continuous time models. *Journal of Econometrics* 169(1):114–122. doi:10.1016/j.jeconom.2012.01.004

Yu, J. (2014). Econometric analysis of continuous time models: A survey of Peter Phillips' work and some new results. *Econometric Theory* 30(4):737–774. doi:10.1017/S0266466613000467

Zhou, Q., Yu, J. (2015). Asymptotic theory for linear diffusions under alternative sampling schemes. *Economics Letters* 128:1–5. doi:10.1016/j.econlet.2014.12.015