New Distribution Theory for the Estimation of Structural Break Point in Mean*

Liang Jiang
Singapore Management University

Xiaohu Wang
The Chinese University of Hong Kong

Jun Yu
Singapore Management University

September 22, 2016

Abstract

Based on the Girsanov theorem, this paper obtains the exact distribution of the maximum likelihood estimator of structural break point in a continuous time model. The exact distribution is asymmetric and tri-modal, indicating that the estimator is biased. These two properties are also found in the finite sample distribution of the least squares (LS) estimator of structural break point in the discrete time model, suggesting the classical long-span asymptotic theory is inadequate. The paper then builds a continuous time approximation to the discrete time model and develops an in-fill asymptotic theory for the LS estimator. The in-fill asymptotic distribution is asymmetric and tri-modal and delivers good approximations to the finite sample distribution. To reduce the bias in the estimation of both the continuous time and the discrete time models, a simulation-based method based on the indirect estimation (IE) approach is proposed. Monte Carlo studies show that IE achieves substantial bias reductions.

JEL classification: C11; C46
Keywords: Structural break, Bias reduction, Indirect estimation, Exact distribution, In-fill asymptotics

---

*We would like to thank the editors and 3 referees, participants of many conferences and seminars for their constructive comments that have significantly improved the paper. This research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund Tier 3 grant with the MOE's official grant number MOE2013-T3-1-009. Corresponding author: Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University; email: yujun@smu.edu.sg.

1 Introduction

Statistical inference of structural breaks has received a great deal of attention both in the econometrics and in the statistics literature over the last several decades. Tremendous efforts have been made in developing the asymptotic theory for the estimation of the fractional structural break point (the absolute structural break point divided by the total sample size), including the consistency, the rate of convergence, and the limiting distribution; see, for example, Yao (1987) and Bai (1994, 1997b), among others. Asymptotic theory has been developed under the long-span asymptotic scheme under which the time span of data is assumed to infinity. This long-span asymptotic distribution is the distribution of the location of the extremum of a two-sided Brownian motion with triangular drift over the interval $(-\infty, +\infty)$. It is symmetric with the true break point being the unique mode, indicating that the estimators have no asymptotic bias. Interestingly and rather surprisingly, how well the asymptotic distribution works in finite sample is largely unknown.

Focusing on simple models with a shift in mean, this paper systematically investigates the performance of the long-span asymptotic distribution, the exact distributional properties, and the bias problem in the estimation of the structural break point. To the best of our knowledge, our study is the first systematic analysis of the exact distribution theory in the literature.

Our paper makes several contributions to the literature. First, by using the Girsanov theorem, we develop the exact distribution of the maximum likelihood (ML) estimator of the structural break point in a continuous time model, assuming that a continuous record over a finite time span is available. It is shown that the exact distribution is asymmetric when the true break point is not in the middle of the sample. Moreover, the exact distribution has trimodality when the signal-to-noise ratio (the break size over the standard deviation of the error term) is not very large, regardless of the location of the true break point. Asymmetry together with trimodality makes the ML estimator biased and suggests that the long-span asymptotic distribution does not conform to the exact distribution. It is also found that upward (downward) bias is obtained when the fractional structural break point is smaller (larger) than 50%, and the further the fractional structural break point away from 50%, the larger the bias.

Second, the properties of asymmetry and trimodality are found to be shared by the finite sample distribution of the LS estimator of the structural break point in the discrete time model, suggesting a substantial bias in the LS estimator and the inadequacy of the long-span asymptotic distribution in finite sample approximations. To better approximate the finite sample distribution, we consider a continuous time approximation to the discrete time model with a structural break in mean and develop an in-fill asymptotic theory for the LS estimator. The developed in-fill asymptotic distribution retains the properties of asymmetry and trimodality, and, hence, provides better approximations than the long-span asymptotic distribution. The in-fill asymptotic scheme leads to a break size of a smaller order than that assumed in Bai (1994). It is this important difference in the break size that leads to a different asymptotic distribution.

Third, an indirect estimation (IE) procedure is proposed to reduce bias in the estimation of the structural break point. One standard method for bias reduction is to obtain an analytical form to approximate the bias and then bias-correct the original estimator via the analytical approach as in Kendall (1954) and Yu (2012). However, it is difficult to use the analytical approach here, as the bias formula is difficult to obtain analytically. The primary advantage of IE lies in its merit in calibrating the binding function via simulations and avoiding the need to obtain an analytical expression for the bias function. It is shown that IE, without using the analytical form of the bias, achieves substantial bias reduction.

The in-fill asymptotic treatment is not new in the literature.[1] Recently, Yu (2014) and Zhou and Yu (2015) demonstrated that the in-fill asymptotic distribution provides better approximations to the finite sample distribution than the long-span asymptotic distribution in persistent autoregressive models. To the best of our knowledge, it is the first time in the literature of structural breaks that the in-fill asymptotic distribution is derived. As in Yu (2014) and Zhou and Yu (2015), we also find that the in-fill asymptotic distribution conforms better to the finite sample distribution than the long-span counterpart.

The rest of the paper is organized as follows. Section 2 gives a brief review of the literature and provides the motivations of the paper. Section 3 develops the exact distribution of the ML estimator of structural break point in a continuous time model. Section 4 establishes a continuous time approximation to the discrete time model previously considered in the literature and develops the in-fill asymptotic theory for the LS estimator under different settings. The IE procedure and its applications in the continuous time and the discrete time models with a structural break are introduced in Section 5. In Section 6, we provide simulation results and compare the finite sample performance of IE with that of the traditional estimation methods and of

[1] Phillips (1987) and Perron (1991) developed the in-fill asymptotic distributions of the LS estimator of the autoregressive parameter. Barndorff-Nielsen and Shephard (2004) developed the in-fill asymptotic distribution of the LS estimator in regression models.

other simulation-based methods. Section 7 concludes. All proofs are contained in the Appendix.

2 Literature Review and Motivations

The literature on estimating structural break points is too extensive to review. A partial list of contributions include Hinkley (1970), Hawkins et al. (1986), Yao (1987), Bai (1994, 1995, 1997a, 1997b), Bai and Perron (1998) and Bai et al. (1998). In these studies, large sample theories for different estimators under various model settings are established.

A simplified model considered in Hinkley (1970) is

$$Y_t = \begin{cases} \mu + \epsilon_t & \text{if } \ t \leq k_0 \\ (\mu + \delta) + \epsilon_t & \text{if } \ t > k_0 \end{cases}, \quad t = 1, \ldots, T, \tag{1}$$

where T denotes the number of observations, ϵ_t is a sequence of independent and identically distributed (i.i.d.) random variables with $E(\epsilon_t) = 0$ and $Var(\epsilon_t) = \sigma^2$. Let k denotes the break point with true value k_0. The condition of $1 \leq k_0 < T$ is assumed to ensure that one break happens. The fractional break point is defined as $\tau = k/T$ with true value $\tau_0 = k_0/T$. Constant μ measures the mean of Y_t before break and δ is the break size. Let the probability density function (pdf) of Y_t be $f(Y_t, \mu)$ for $t \leq k_0$ and $f(Y_t, \mu + \delta)$ for $t > k_0$. Under the assumption that the functional form of $f(\cdot, \cdot)$ and the parameters μ and δ are all known, the ML estimator of k is defined as

$$\widehat{k}_{ML,T} = \arg \max_{k=1,\ldots,T-1} \left\{ \sum_{t=1}^{k} \log f(Y_t, \mu) + \sum_{t=k+1}^{T} \log f(Y_t, \mu + \delta) \right\}. \tag{2}$$

The corresponding estimator of τ is $\widehat{\tau}_{ML,T} = \widehat{k}_{ML,T}/T$. Yao (1987) developed a long-span limiting distribution under the scheme of $T \to \infty$ followed by $\delta \to 0$ which takes the form of

$$\delta^2 I(\mu) \left(\widehat{k}_{ML,\infty} - k_0 \right) \xrightarrow{d} \arg \max_{u \in (-\infty, \infty)} \left\{ W(u) - \frac{1}{2}|u| \right\}, \tag{3}$$

where $I(\mu)$ is the Fisher information of the density function $f(y, \mu)$, $W(u)$ is a two-sided Brownian motion which will be defined below, and $\xrightarrow{d}$ denotes convergence in distribution. The closed-form expressions for the pdf and the cumulative distribution function (cdf) of the long-span limiting distribution were derived in Yao (1987).

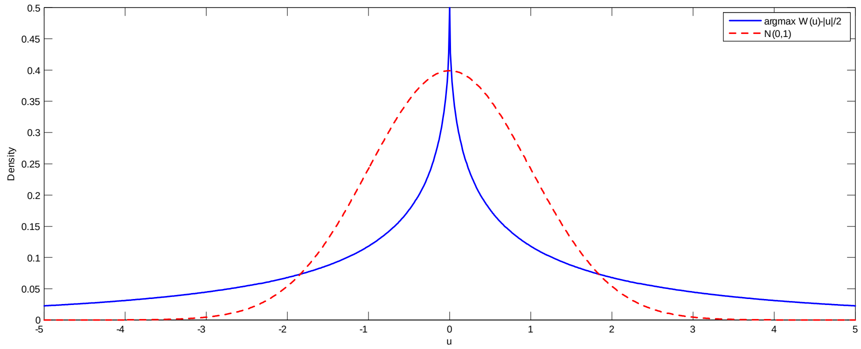

Figure 1: The pdfs of $\arg\max_{u\in(-\infty,\infty)}\left\{W(u)-\frac{1}{2}|u|\right\}$ and a standard normal distribution.

For the same model as in Equation (1) with unknown parameters μ and δ, Hawkins et al. (1986) and Bai (1994) studied the long-span asymptotic behavior of the LS estimator of the break point. The LS estimator takes the form of

$$\widehat{k}_{LS,T}=\arg\min_{k=1,\ldots,T-1}\left\{S_k^2\right\}=\arg\max_{k=1,\ldots,T-1}\left\{\left[V_k\left(Y_t\right)\right]^2\right\}, \tag{4}$$

where $S_k^2=\sum_{t=1}^{k}\left(Y_t-\overline{Y}_k\right)^2+\sum_{t=k+1}^{T}\left(Y_t-\overline{Y}_k^*\right)^2$ with $\overline{Y}_k$ ($\overline{Y}_k^*$) being the sample mean of the first k (last $T-k$) observations and $\left[V_k\left(Y_t\right)\right]^2=\frac{T(T-k)}{T^2}\left(\overline{Y}_k^*-\overline{Y}_k\right)^2$. The corresponding estimator of τ is $\widehat{\tau}_{LS,T}=\widehat{k}_{LS,T}/T$. Hawkins et al. (1986) showed that $T^{\alpha}\left(\widehat{\tau}_{LS,T}-\tau_0\right)\xrightarrow{p}0$ for any $\alpha<1/2$, where $\xrightarrow{p}$ denotes convergence in probability. Bai (1994) improved the rate of convergence by showing that $\widehat{\tau}_{LS,T}-\tau_0=O_p\left(\frac{1}{T\delta^2}\right)$. In addition, by letting the break size depend on T (denoted by δ_T), and assuming that $\delta_T\to0$ with $\frac{\sqrt{T}\delta_T}{\sqrt{\log T}}\to\infty$ as $T\to\infty$, Bai (1994) derived an asymptotic distribution as

$$T\left(\delta_T/\sigma\right)^2\left(\hat{\tau}_{LS,T}-\tau_0\right)\xrightarrow{d}\arg\max_{u\in(-\infty,\infty)}\left\{W(u)-\frac{1}{2}|u|\right\}, \tag{5}$$

which is the same as in (3). This long-span asymptotic distribution in (5) is widely used as an approximation to the finite sample distribution for models with a small break. Note that when ϵ_t is normally distributed, the Fisher information $I\left(\mu\right)$ in Equation (3) is σ^{-2}. In this case, the asymptotic theory for $\widehat{\tau}_{ML,T}$ in Yao (1987) is exactly the same as that for $\widehat{\tau}_{LS,T}$ in Bai (1994). However, Bai's results were obtained without assuming Gaussian errors, and, hence, an invariance principle applies.

Figure 1 plots the pdf of the limiting distribution given in (3) and (5). For the purpose of comparison, the pdf of a standard normal distribution is also plotted. It

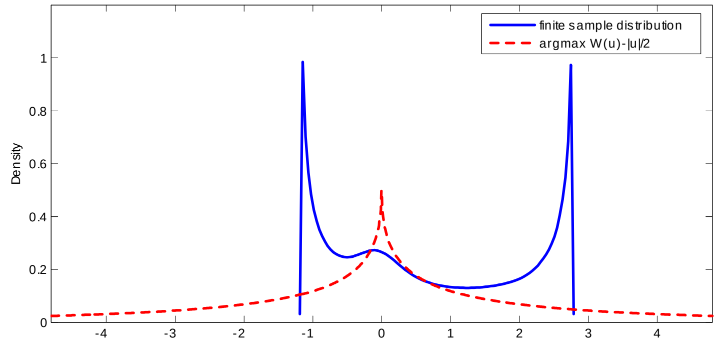

Figure 2: The pdf of the finite sample distribution of $T\left(\frac{\delta_T}{\sigma}\right)^2(\widehat{\tau}_{LS,T}-\tau_0)$ when $T=100$, $\delta_T=0.2$, $\sigma=1$ and $\tau_0=0.3$ in Model (1) and the pdf of $\arg\max_{u\in(-\infty,\infty)}\left\{W(u)-\frac{1}{2}|u|\right\}$.

can be seen that, relative to the standard normal distribution, the limiting distribution obtained in the literature has much fatter tails and a much higher peak. More importantly, the limiting distribution has a unique mode at the origin and is symmetric about it, suggesting that both the ML estimator and the LS estimator have no asymptotic bias, no matter what the true value of the structural break point is.

Unfortunately, the long-span asymptotic distribution developed in the literature does not perform well in many empirically relevant cases. To see this problem, in Figure 2 we plot the pdf of the long-span asymptotic distribution listed in (5) and the finite sample distribution of $T\left(\frac{\delta}{\sigma}\right)^2(\widehat{\tau}_{LS,T}-\tau_0)$ when $T=100$, $\delta=0.2$, $\sigma=1$ and $\tau_0=0.3$ in Model (1). The finite sample distribution is obtained from simulated data. It is clear that the two distributions are very different from each other. Three striking distinctions can be found. First, the finite sample distribution is asymmetric, whereas the long-span asymptotic distribution is symmetric. Second, the finite sample distribution displays trimodality while the long-span asymptotic distribution has a unique mode. Third, the finite sample distribution indicates that the LS estimator $\widehat{\tau}_{LS,T}$ is seriously biased. The simulation result shows that the bias is 0.1704, which is about 57% of the true value. In contrast, there is no bias suggested by the long-span asymptotic distribution. It is this inadequacy of the long-span asymptotic distribution for approximating the finite sample distribution that motivates us to develop an alternative distribution theory for the estimation of the structural break point.

3 A Continuous Time Model

In this section we focus our attention on a continuous time model with a structural break in the drift function. The model considered here is

$$dX(t) = \left(\mu + \left(\frac{\delta^*}{\varepsilon} \right) 1_{[t>\tau_0]} \right) dt + \sigma dB(t), \tag{6}$$

where $t \in [0,1]$, $1_{[t>\tau_0]}$ is an indicator function, μ, δ^*, ε and τ_0 are constants with δ^*/ε being the break size, σ is another constant capturing the noise level, and $B(t)$ denotes a standard Brownian motion. The condition of $\tau_0 \in [\alpha, \beta]$ with $0 < \alpha < \beta < 1$ is assumed to ensure that one break happens during the time interval $(0, 1)$. We further assume that a continuous record is available and all parameters are known except for τ_0. With a continuous record, assuming a more complex structure for σ such as a time varying diffusion will not change the analysis because the diffusion function can be estimated by quadratic variation without estimation error.

The continuous time diffusion model is a natural choice to study the asymmetry of the sample information before and after the break point. This is because it is well-known in the continuous time literature that the longer the time span over which a continuous record is available, the more information that the continuous record contains about the parameters in the drift function; see Phillips and Yu (2009a, 2009b). Hence, when $\tau_0 \neq 1/2$, the amount of information contained by observations over the time interval $[0, \tau_0]$ is different from that over the time interval $[\tau_0, 1]$. This difference is captured by the asymmetry in the length of the time span before and after the break point. Therefore, the exact distribution of the ML estimator of the structural break point is expected to be asymmetric.

For any $\tau \in (0, 1)$ we can obtain the exact log-likelihood function of Model (6) via the Girsanov Theorem as[2]

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

where P_τ is the probability measure corresponding to Model (6) with τ_0 being replaced by τ for any $\tau \in (0, 1)$ and P_B is the probability measure corresponding to $B(t)$. This leads to the ML estimator of τ_0 as

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

[2] See also Phillips and Yu (2009b) for a recent usage of the Girsanov Theorem in estimating continuous time models.

Following the literature, we now define a two-sided Brownian motion as

$$W(u) = \begin{cases} W_1(-u) = B(\tau_0) - B(\tau_0 - (-u)) & \text{if } u \le 0 \\ W_2(u) = B(\tau_0) - B(\tau_0 + u) & \text{if } u > 0 \end{cases}, \tag{8}$$

where $W_1(s) = B(\tau_0) - B(\tau_0 - s)$ and $W_2(s) = B(\tau_0) - B(\tau_0 + s)$ are two independent Brownian motions composed by increments of the standard Brownian motion $B(\cdot)$ before and after τ_0, respectively. Theorem 3.1 reports the exact distribution of $\widehat{\tau}_{ML}$.

Theorem 3.1 *Consider Model (6) with a continuous record being available. For the ML estimator $\widehat{\tau}_{ML}$ defined in (7),*
(a) when ε is a constant, we have the exact distribution as

$$\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2 (\widehat{\tau}_{ML} - \tau_0) \stackrel{d}{=} \arg \max_{u \in \left(-\tau_0\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2, (1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2\right)} \left\{W(u) - \frac{|u|}{2}\right\}; \tag{9}$$

(b) when $\varepsilon \to 0$, the break size $\delta^*/\varepsilon \to \infty$, we have the small-$\varepsilon$ distribution as

$$\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2 (\widehat{\tau}_{ML} - \tau_0) \xrightarrow{d} \arg \max_{u \in (-\infty,\infty)} \left\{W(u) - \frac{|u|}{2}\right\},$$

where $W(u)$ is the two-sided Brownian motion defined in (8), and $\stackrel{d}{=}$ denotes equivalence in distribution.

Part (a) of Theorem 3.1 gives the exact distribution of $\widehat{\tau}_{ML}$ when a continuous record over a finite time span is available. It is different from the long-span limiting distribution developed in the literature as in (5) in two obvious aspects. First, the limiting distribution in (5) corresponds to the location of the extremum of $W(u) - \frac{1}{2}|u|$ over the interval of $(-\infty, \infty)$. As the interval is symmetric about zero, the limiting distribution is symmetric too. However, the exact distribution in (9) corresponds to the interval of $\left(-\tau_0\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2, (1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2\right)$, which depends on the true value of the fractional break point τ_0. Only when the true break point is exactly in the middle of the sample, i.e., $\tau_0 = 1/2$, does the interval become $\left(-\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2/2, \left(\frac{\delta^*}{\sigma\varepsilon}\right)^2/2\right)$, being symmetric about the origin. In this case the exact distribution is symmetric. However, if τ_0 is not 1/2, the interval and hence the exact distribution will be asymmetric, indicating that $\widehat{\tau}_{ML}$ is biased. It is easy to see that the exact distribution in (9) suggests upward bias when $\tau_0 < 1/2$ and downward bias when $\tau_0 > 1/2$, and the further τ_0 away from 1/2, the larger the bias. In addition, the signal-to-noise ratio $\frac{\delta^*}{\sigma\varepsilon}$ contributes to the degree of asymmetry of the interval, and, hence, affects the exact distribution and the magnitude

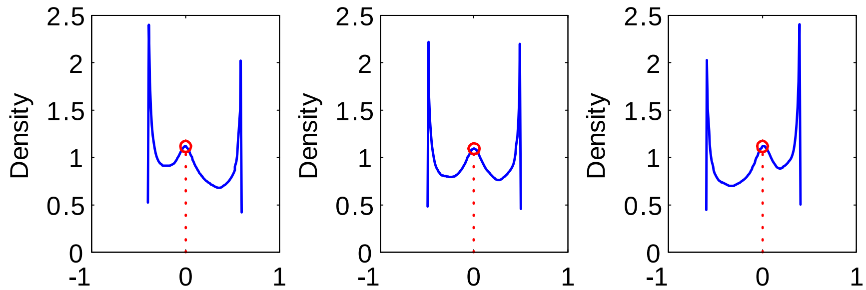

Figure 3: The density of $\widehat{\tau}_{ML} - \tau_0$ given in Equation (9) when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively) and the signal-to-noise ratio $(\frac{\delta^*}{\sigma \epsilon})$ is 1.

of bias. These findings are confirmed by the simulation results reported in Figures 3-4, in which we plot the density functions of the exact distribution when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively), $\varepsilon = 1$ and $\frac{\delta^*}{\sigma \varepsilon}$ being 1 and 4, respectively.

Second, the interval to locate the argmax in the exact distribution in (9) is always bounded. Whereas, the interval to locate the argmax in the long-span limiting distribution in (5) is unbounded. Such a difference has an implication for the modality of the distribution. As shown in Figure 1, the long-span limiting distribution has a unique mode at the origin. Whereas, Figures 3-4 shows that the exact distribution displays trimodality. One mode is at the origin. The other two modes are at the two boundary points, $-\tau_0 \left(\frac{\delta^*}{\sigma \varepsilon}\right)^2$ and $(1-\tau_0)\left(\frac{\delta^*}{\sigma \varepsilon}\right)^2$. When $\tau_0 = 1/2$, the two modes at the boundary points have the same height. When $\tau_0 \neq 1/2$, the two modes at the boundary points do not have the same height. From the comparison of Figures 3-4, we can also find that the modes at the two boundary points are higher when the signal-to-noise ratio is smaller. As a result, for the case where $\tau_0 \neq 1/2$, the exact distribution is more skewed and leads to a larger bias of $\widehat{\tau}_{ML}$ when the signal-to-noise ratio is smaller.

The mode at the origin is well expected. This is because the drift term and the random term in $W(u) - \frac{1}{2}|u|$ are $-\frac{1}{2}|u|$ and $W(u) = O_p(\sqrt{|u|})$, respectively. When $|u|$ is large, the negative drift term dominates the random term. As a result, the probability for $W(u) - \frac{1}{2}|u|$ to reach the maximum at a large value of $|u|$ should be small, and decreasing as $|u|$ getting larger. In the mean time, because of the randomness in $W(u)$, it is still possible for $W(u) - \frac{1}{2}|u|$ to reach the maximum at any large value of $|u|$. This also explains the shape of the long-span limiting distribution in (5) as apparent in Figure 1.

When the interval of the exact distribution in (9) is bounded with a comparatively small value of $\frac{\delta^*}{\sigma \varepsilon}$, $\frac{1}{2}|u|$ takes small values even at the boundary points. This means that the negative drift term becomes less dominant, hence, it is more likely for $W(u) - \frac{1}{2}|u|$

to reach the maximum in the neighborhoods of the two boundary points. To explain why there are two modes at the two boundary points, take the right boundary point $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$ as an example. Being a mode at this boundary point means that it is more likely for $W(u)-\frac{1}{2}|u|$ to reach the maximum at $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$ than at any point arbitrarily close to but strictly less than $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$. Given the randomness of $W(u)$, the probability for $W(u)-\frac{1}{2}|u|$ to reach the maximum in any small left neighborhood of $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$ is nonzero. Conditional on the event that $W(u)-\frac{1}{2}|u|$ reaches the maximum in a small left neighborhood, for $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$ to be the arg max point, the value of $W(u)-\frac{1}{2}|u|$ at $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$ only needs to be larger than the value of $W(u)-\frac{1}{2}|u|$ at the points smaller than $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$. However, for any interior point to be the arg max, we have to compare the value of $W(u)-\frac{1}{2}|u|$ at this interior point with that at both sides of this interior point. Therefore, $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$ is more likely to be the arg max of $W(u)-\frac{1}{2}|u|$ than any interior point. Similar arguments apply to the other boundary point, $-\tau_0\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$.

When the signal-to-noise ratio $\frac{\delta^*}{\sigma\varepsilon}$ gets smaller, the values of the two boundary points become smaller too. Hence, the probabilities of $W(u)-\frac{1}{2}|u|$ reaching its maximum in the neighborhoods of the two boundary points get larger, leading to larger values of the modes at the two boundary points. Similar arguments explain the reason why the boundary point closer to the origin has a larger mode than the other boundary point. Moreover, when $\frac{\delta^*}{\sigma\varepsilon}$ is very small, the length of the interval over which $W(u)-\frac{1}{2}|u|$ is maximized is very small. In this case, the negative drift term is stochastically dominated by the random term in $W(u)-\frac{1}{2}|u|$. This explains why the origin may not be the highest mode when the signal-to-noise ratio is very small, as apparent in Figure 3.

Part (b) of Theorem 3.1 shows that the asymptotic distribution of $\widehat{\tau}_{ML}$ when $\varepsilon \to 0$ is the same as the long-span asymptotic distribution developed in Yao (1986) and Bai (1994). The same small-ε asymptotic distribution is also obtained in Ibragimov and Has'minskii (1981).

4 Continuous Time Approximation to Discrete Time Models

Motivated by the findings in the exact distribution in the continuous time model, in this section we first build a continuous time approximation to the discrete time structural break model widely studied in the literature. Then, we develop the in-fill asymptotic theory for the LS estimator of the break point, and show that the in-fill asymptotic distribution provides better approximations to the finite sample distribution than the

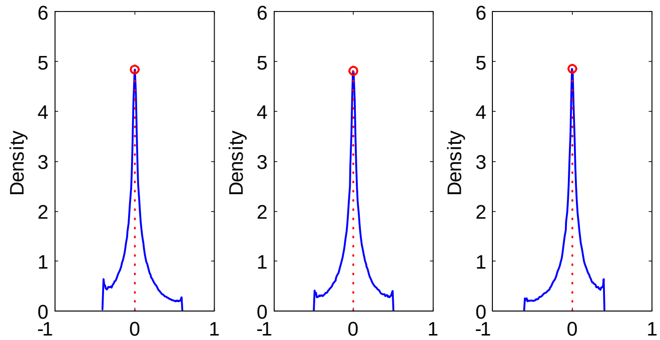

Figure 4: The density of $\widehat{\tau}_{ML} - \tau_0$ given in Equation (9) when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively) and the signal-to-noise ratio $\frac{\delta^*}{\sigma\epsilon}$ is 4.

long-span asymptotic distribution developed in the literature.

Consider the continuous time process $X(t)$ defined in (6). We now assume that the observations are only available at discrete time points, say at T equally spaced points $\{th\}_{t=1}^{T}$, where h is the sampling interval and $T = 1/h$ is the sample size. For simplicity, we assume the structural break point $T\tau_0$ to be an integer, denoted by k_0. Let $\{X_{th}\}_{t=1}^{T}$ denote the discrete time observations. Then, the exact discretization of the continuous time process defined in (6) can be written as

$$X_{th} - X_{(t-1)h} = \begin{cases} \mu h + \sqrt{h}\epsilon_t & \text{for } \ t = 1, \cdots, k_0, \\ (\mu + \delta^*/\varepsilon)h + \sqrt{h}\epsilon_t & \text{for } \ t = k_0 + 1, \cdots, T, \end{cases}$$

where $\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$. Letting $Z_t = \left(X_{th} - X_{(t-1)h}\right)/\sqrt{h}$, we have

$$Z_t = \begin{cases} \mu\sqrt{h} + \epsilon_t & \text{if } \ t \leq k_0, \\ (\mu + \delta^*/\varepsilon)\sqrt{h} + \epsilon_t & \text{if } \ t > k_0. \end{cases} \tag{10}$$

It can be seen that, whenever h is fixed, the discrete time model in Equation (10) is the same as the one studied in Yao (1987) and Bai (1994) given in Equation (1) with ϵ_t being normally distributed and the shift in mean being $\delta = (\delta^*/\varepsilon)\sqrt{h}$.

We now develop the asymptotic theory of the LS estimator of $\tau_0 = k_0/T$ under the in-fill asymptotic scheme where $h \to 0$ with a fixed time span $Th = 1$. Clearly, if $h \to 0$, the sample size $T \to \infty$. In the limit of $h \to 0$, a continuous record is available. As it can be seen clearly in the proofs in Appendix, the development of the in-fill asymptotic theory does not require the assumption of Gaussian errors. Therefore, an invariance principle applies. Moreover, the in-fill asymptotic theory continues to hold when $\mu\sqrt{h}$ in Model (10) is replaced with μ. In other words, making the means of Z_t before and after break to be around a constant different from zero, instead of converging to zero when $h \to 0$ as required in Model (10), would not change the in-fill asymptotics developed in the section.

With a fixed ε, the in-fill asymptotic scheme implies that the break size $(\delta^*/\varepsilon)\sqrt{h}$ goes to zero at the rate of $1/\sqrt{T}$. This rate is faster than that assumed in Bai (1994). This key difference makes our in-fill asymptotic theory different from the long-span asymptotic theory developed in Bai (1994). When μ and δ^*/ε are known, the in-fill asymptotic distribution is shown to be the same as the exact distribution of the ML estimator when a continuous record is available, as given in Part (a) of Theorem 3.1. When μ and δ^*/ε are unknown, we derive an in-fill asymptotic distribution which is asymmetric if $\tau_0 \neq 1/2$, and has trimodality. In both cases, simulation results show that the in-fill asymptotic distribution provides better approximations to the finite sample distribution.

We also consider the in-fill asymptotic scheme with $\varepsilon \to 0$ and $(\delta^*/\varepsilon)\sqrt{h} \to 0$. In this case the break size goes to zero but at a rate slower than $1/\sqrt{T}$. It is shown that the in-fill asymptotic distribution with $\varepsilon \to 0$ is the same as the long-span asymptotic distribution obtained in Yao (1987) and Bai (1994). Hence, our setup and results generalize and connect naturally with those in the literature.

4.1 In-fill asymptotics when only τ is unknown

When μ and δ^*/ε are known, the LS estimator of k is defined as

$$\begin{aligned}
\widehat{k}_{LS,T} &= \arg\min_{k=1,\ldots,T-1} \left\{ \sum_{t=1}^{k} \left(Z_t - \mu\sqrt{h} \right)^2 + \sum_{t=k+1}^{T} \left(Z_t - (\mu + \delta^*/\varepsilon)\sqrt{h} \right)^2 \right\} \\
&= \arg\max_{k=1,\ldots,T-1} \left\{ -(\delta^*/\varepsilon)\sqrt{h} \sum_{t=1}^{k} \left(Z_t - \mu\sqrt{h} \right) + (\delta^*/\varepsilon)^2 hk/2 \right\}. \qquad (11)
\end{aligned}$$

The corresponding estimator of τ is $\widehat{\tau}_{LS,T} = \widehat{k}_{LS,T}/T$. When the errors in Model (10) are normally distributed, the LS estimators of k and τ are identical to the ML estimators as defined in Yao (1987). Compared to Yao's long-span asymptotic distribution, the in-fill asymptotic distribution given in Part (a) of Theorem 4.1 provides an alternative asymptotic approximation to the finite sample distribution of $\widehat{\tau}_{LS,T}$. Part (b) of the theorem connects our in-fill asymptotics to Yao's long-span asymptotics.

Theorem 4.1 *Consider Model (10) with known μ and δ^*/ε. Denote the LS estimator $\widehat{\tau}_{LS,T} = \widehat{k}_{LS,T}/T$ with $\widehat{k}_{LS,T}$ defined in (11). Then,* *(a) when $h \to 0$ with a fixed ε, we have the in-fill asymptotic distribution as*

$$T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS,T}-\tau_0)\xrightarrow{d}\arg\max_{u\in\left(-\tau_0\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2,(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2\right)}\left\{W(u)-\frac{|u|}{2}\right\};$$

(b) when $h \to 0$ and $\varepsilon \to 0$ simultaneously with $(\delta^*/\varepsilon)\sqrt{h} \to 0$, we have the small-$\varepsilon$ in-fill asymptotic distribution as

$$T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS,T}-\tau_0)\xrightarrow{d}\arg\max_{u\in(-\infty,\infty)}\left\{W(u)-\frac{|u|}{2}\right\},$$

where $W(u)$ is the two-sided Brownian motion defined in (8).

Remark 4.1 *Note that $T = 1/h$ implies $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2 = (\delta^*/(\sigma\varepsilon))^2$. Hence, the in-fill asymptotic distribution of $\widehat{\tau}_{LS,T}$ in Theorem 4.1 is the same as the exact distribution of $\widehat{\tau}_{ML}$ obtained in Theorem 3.1.*

Remark 4.2 *When $h \to 0$ with a fixed ε, $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2 = (\delta^*/(\sigma\varepsilon))^2$ is a constant. In this case, Part (a) of Theorem 4.1 shows that $\widehat{\tau}_{LS,T}$ is inconsistent and $\widehat{k}_{LS,T}-k_0$ diverges at the rate of T. When $h \to 0$ and $\varepsilon \to 0$ simultaneously with $(\delta^*/\varepsilon)\sqrt{h} \to 0$, the break size shrinks to zero but at a rate slower than $1/\sqrt{T}$. In this case, $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2 \to \infty$ and $\widehat{\tau}_{LS,T}$ becomes consistent as shown by Part (b) of Theorem 4.1. Moreover, the small-ε in-fill asymptotic distribution obtained in Part (b) of Theorem 4.1 is the same as the long-span asymptotic distribution obtained in Bai (1994). Clearly, by relaxing the assumption of Bai, we get the same asymptotic distribution.*

Remark 4.3 *The proof of Theorem 4.1 does not depend on the assumption of Gaussian errors. Therefore, an invariance principle applies to the in-fill asymptotics. Moreover, the proof of Theorem 4.1 can be easily extended to the case where the errors in Model (10) follow a weakly stationary process with a long-run variance $[a(1)]^2$. In this case, the results in Theorem 4.1 still hold but with σ^2 being replaced by $[a(1)]^2$.*

Figure 5 plots the finite sample distribution of $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS,T}-\tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) obtained from simulations, the density of the in-fill asymptotic distribution given in Part (a) of Theorem 4.1 and the density of the long-span limiting distribution given in Yao (1987). The data are simulated from Model (10) with $\mu = 0$, $\delta^* = 2$, $\varepsilon = 1$, $\sigma = 1$ and $h = 1/100$. So the break size is $\left(\frac{\delta^*}{\varepsilon}\right)\sqrt{h} = 0.2$. The experiment is replicated 100,000 times to obtain the density. The first part of Table 1 reports the finite sample bias of $\widehat{\tau}_{LS,T}$, the bias

Table 1: The table shows the finite sample bias of $\hat{\tau}_{LS,T}$, the bias from the in-fill asymptotic distribution, and the bias from the long-span asymptotic distribution. These three kinds of bias are denoted by FS1, IF1, LS1, respectively, when τ_0 is the only unknown parameter; and are denoted by FS2, IF2, LS2, respectively, when more parameters are unknown. The number of replications is 100,000.

$\frac{\delta^*}{\sigma\varepsilon}$	2	2	4	4	6	6
τ_0	0.3	0.7	0.3	0.7	0.3	0.7
FS1	.0909	-.0921	.0307	-.0305	.0078	-.0080
IF1	.0911	-.0903	.0299	-.0302	.0073	-.0072
LS1	0	0	0	0	0	0
FS2	.1704	-.1717	.1068	-.1062	.0511	-.0495
IF2	.1738	-.1741	.1140	-.1142	.0549	-.0555
LS2	0	0	0	0	0	0

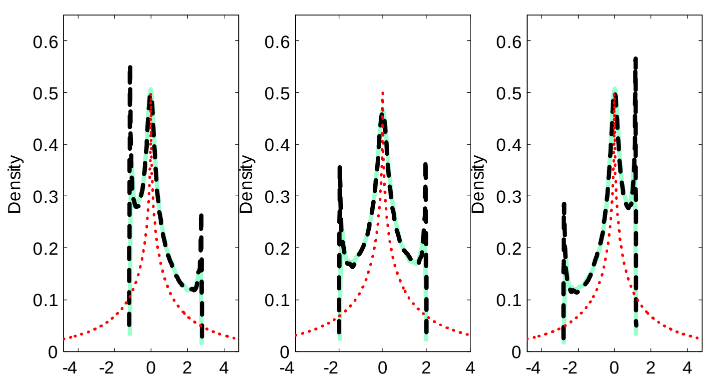

Figure 5: The pdf of $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS,T}-\tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\frac{\delta^*}{\sigma\varepsilon} = 2$. The blue solid line is the finite sample distribution when $T = 100$; the black broken line is the density given in Part (a) of Theorem 4.1; and the red dotted line is the long-span limiting distribution in Yao (1987).

implied by the in-fill asymptotic distribution in Part (a) of Theorem 4.1, and the bias implied by the long-span limiting distribution in Yao (1987), for the cases where the signal-to-noise ratio $\frac{\delta^*}{\sigma\varepsilon} = 2, 4, 6$, respectively.

Several features are apparent in Figure 5 and the first part of Table 1. First, the finite sample distribution is not symmetric about 0 when $\tau_0 \neq 1/2$. In particular, if τ_0 is smaller (larger) than 1/2, the density is positively (negatively) skewed, indicating an upward (downward) bias in $\widehat{\tau}_{LS,T}$. The bias is 30% above the true value when $\tau_0 = 0.3$ which is substantial. Second, the finite sample distribution has trimodality. The origin is one of the three modes and the two boundary points, $-\tau_0 \left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$ and $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$, are the other two. Third and most importantly, the in-fill asymptotic distribution given in Part (a) of Theorem 4.1 shares the two important features of the finite sample distribution, namely, asymmetry and trimodality, and captures the finite sample bias very well. Not surprisingly, it provides much better approximations to the finite sample distribution than the long-span asymptotic distribution. Fourth, as revealed by the first part of Table 1, as the signal-to-noise ratio $\frac{\delta^*}{\sigma\varepsilon}$ increases, the magnitude of asymmetry in the finite sample distribution decreases, and, hence, the finite sample bias becomes smaller. This property is also well captured by the in-fill asymptotics.

4.2 In-fill asymptotics with more unknown parameters

When μ and δ^*/ε are unknown, the means before and after the break point have to be estimated. As in Bai (1994), the LS estimator of the break point is now defined as,

$$\widehat{k}_{LS,T} = \arg \min_{k=1,\dots,T-1} \left\{ \sum_{t=1}^{k} \left(Z_t - \overline{Z}_k\right)^2 + \sum_{t=k+1}^{T} \left(Z_t - \overline{Z}_k^*\right)^2 \right\} = \arg\max_k \left\{ \left[V_k\left(Z_t\right)\right]^2 \right\}, \tag{12}$$

where $\overline{Z}_k$ ($\overline{Z}_k^*$) is the sample mean of the first k (last $T-k$) observations and $\left[V_k\left(Z_t\right)\right]^2 = \frac{T(T-k)}{T^2}\left(\overline{Z}_k^* - \overline{Z}_k\right)^2$. Similarly, $\widehat{\tau}_{LS,T} = \widehat{k}_{LS,T}/T$.

Theorem 4.2 *Consider Model (10) with unknown parameters of μ and δ^*/ε. For the LS estimator $\widehat{\tau}_{LS,T} = \widehat{k}_{LS,T}/T$ with $\widehat{k}_{LS,T}$ defined in (12),*
(a) when $h \to 0$ with a fixed ε, we have the in-fill asymptotic distribution as

$$T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2 \left(\widehat{\tau}_{LS,T} - \tau_0\right) \xrightarrow{d} \left(\frac{\delta^*}{\sigma\varepsilon}\right)^2 \arg\max_{u\in(-\tau_0, 1-\tau_0)} \left[\tilde{B}\left(u\right)\right]^2, \tag{13}$$

with

$$\tilde{B}\left(u\right) = \begin{cases} B_1\left(1-\tau_0-u\right) - B_2\left(\tau_0+u\right) - \frac{(1-\tau_0)\sqrt{\tau_0+u}}{\sqrt{1-\tau_0-u}}\frac{\delta^*}{\sigma\varepsilon} & \textit{for } u \leq 0 \\ B_1\left(1-\tau_0-u\right) - B_2\left(\tau_0+u\right) - \frac{\tau_0\sqrt{1-\tau_0-u}}{\sqrt{\tau_0+u}}\frac{\delta^*}{\sigma\varepsilon} & \textit{for } u > 0 \end{cases},$$

where $B_1(s)$ is a standard Brownian motion and $B_2(1-s) \equiv B_1(1) - B_1(s)$;
(b) when $h \to 0$ and $\varepsilon \to 0$ simultaneously with $(\delta^*/\varepsilon)\sqrt{h} \to 0$, we have the small-$\varepsilon$ in-fill asymptotic distribution as

$$T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2 (\widehat{\tau}_{LS,T} - \tau_0) \xrightarrow{d} \arg\max_{u\in(-\infty,\infty)} \left\{W(u) - \frac{|u|}{2}\right\},$$

where $W(u)$ is the two-sided Brownian motion defined in (8).

Remark 4.4 *The in-fill asymptotic distribution reported in Part (a) of Theorem 4.2 is new to the literature. When $\tau_0 \neq 1/2$, the interval $(-\tau_0, 1-\tau_0)$ is asymmetric about zero and, not surprisingly, the in-fill asymptotic distribution is asymmetric too. When $\tau_0 = 1/2$, the interval becomes symmetric, and we have*

$$\tilde{B}(u) = \begin{cases} B_1(1/2-u) - B_2(1/2+u) - \frac{\sqrt{1/2+u}}{2\sqrt{1/2-u}}\frac{\delta^*}{\sigma\varepsilon} & \text{for } u \leq 0 \\ B_1(1/2-u) - B_2(1/2+u) - \frac{\sqrt{1/2-u}}{2\sqrt{1/2+u}}\frac{\delta^*}{\sigma\varepsilon} & \text{for } u > 0 \end{cases},$$

which is symmetrically distributed about zero. As a result, the distribution in Part (a) of Theorem 4.2 is symmetric about zero when $\tau_0 = 1/2$. In practice, one needs to estimate τ and the signal-to-noise ratio $\frac{\delta^*}{\sigma\varepsilon}$ and then insert the estimated values into the in-fill asymptotic distribution reported in Part (a) of Theorem 4.2 for the purpose of making statistical inference.

Remark 4.5 *By using the Beveridge-Nelson decomposition and the functional central limit theory for serially dependent processes, Theorem 4.2 can be extended to the case where the errors in Model (10) follow a weakly stationary process with a long-run variance $[a(1)]^2$. In this case, the results in Theorem 4.2 still apply with σ^2 being replaced by $[a(1)]^2$.*

Figures 6 and 7 plot the finite sample distribution of $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2 (\widehat{\tau}_{LS,T} - \tau_0)$, obtained from simulated data, when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively), the density of the in-fill asymptotic distribution given in Part (a) of Theorem 4.2 and the density of the long-span limiting distribution given in Bai (1994). The data are simulated from Model (10) with $\mu = 0$, $\delta^* = 2$, $\varepsilon = 1$, $\sigma = 1$ and $h = 1/100$ and so the break size is $\left(\frac{\delta^*}{\varepsilon}\right)\sqrt{h} = 0.2$ in Figure 6. Figure 7 corresponds to $\frac{\delta^*}{\varepsilon} = 4$ and so the break size is $\left(\frac{\delta^*}{\varepsilon}\right)\sqrt{h} = 0.4$. The experiment is replicated 100,000 times. The finite sample bias of $\widehat{\tau}_{LS,T}$, the bias implied by the in-fill asymptotic distribution, and the bias implied by the long-span limiting distribution are reported in the second part of Table 1.

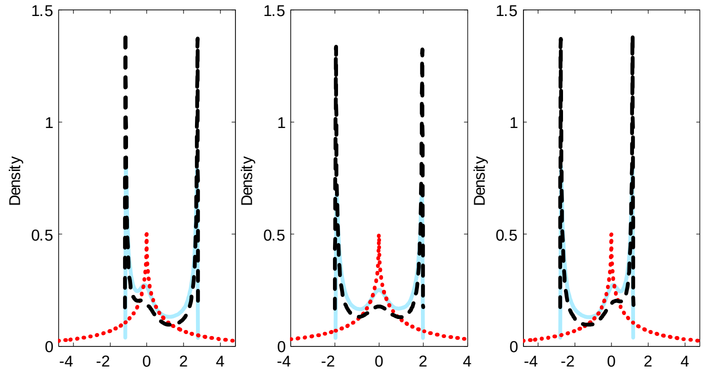

Figure 6: The pdf of $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS}-\tau_0)$ when $\tau_0=0.3,0.5,0.7$ (the left, middle and right panel respectively) and $\frac{\delta^*}{\sigma\varepsilon}=2$. The blue solid line is the finite sample distribution when $T=100$; the black broken line is the density given in Part (a) of Theorem 4.2; and the red dotted line is the long-span limiting distribution in Bai (1994).

Several features are apparent in Figures 6-7 and the second part of Table 1. First, the finite sample distribution is asymmetric about 0 when $\tau_0 \neq 1/2$, and, hence, $\widehat{\tau}_{LS,T}$ is biased. In particular, if τ_0 is less (greater) than 1/2, the density is positively (negatively) skewed, leading to an upward (downward) bias in $\widehat{\tau}_{LS,T}$. The bias is more than 50% of the true value if $\tau_0=0.3$, which is very substantial. Second, the finite sample distribution is not as concentrated around zero as suggested by the long-span limiting distribution. The finite sample distribution has trimodality. The origin is one of the three modes and the two boundary points, $-\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2\tau_0$ and $\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2(1-\tau_0)$, are the other two. The peak at the origin can be smaller than those at the boundary points when $\frac{\delta^*}{\sigma\varepsilon}$ is small. Third and most importantly, the in-fill asymptotic distribution given in Part (a) of Theorem 4.2 has trimodality, and is asymmetric about zero when $\tau_0 \neq 1/2$. It provides better approximations to the finite sample distribution than the long-span limiting distribution. Comparing two parts in Table 1, it can be seen that when other parameters are unknown, the bias in $\widehat{\tau}_{LS,T}$ increases. In spite of the increased bias in $\widehat{\tau}_{LS,T}$, it can be seen from the second part of Table 1 that the in-fill asymptotic distribution also captures the finite sample bias very well.

5 Bias Correction via IE

Indirect estimation (IE) is a simulation-based method, first introduced by Smith (1993), Gouriéroux et al. (1993), and Gallant and Tauchen (1996). This method is particularly useful for estimating parameters of a model where moments and likelihood function are

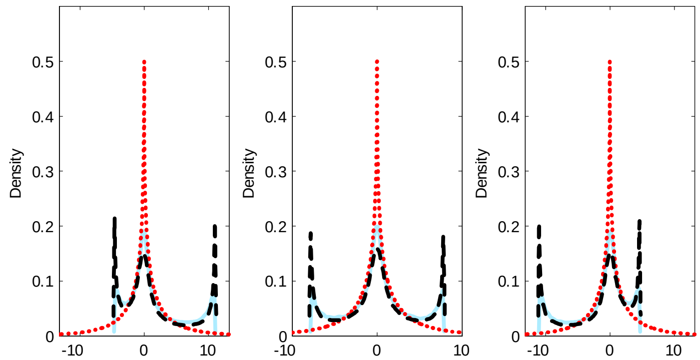

Figure 7: The pdf of $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS}-\tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\frac{\delta^*}{\sigma\varepsilon} = 4$. The blue solid line is the finite sample distribution when $T = 100$; the black broken line is the density given in Part (a) of Theorem 4.2; and the red dotted line is the long-span limiting distribution in Bai (1994).

difficult to calculate, but the model is easy to simulate. It uses an auxiliary model to capture aspects of the data upon which to base the estimation. The parameters of the auxiliary model can be estimated using either the observed data or the data simulated from the true model. Then, IE estimates are obtained by minimizing the distance between the two sets of parameter estimates. Typically, one chooses an auxiliary model that is amenable to estimate and well approximates the true model at the same time.

To improve finite sample properties of the original estimator, McKinnon and Smith (1998) and Gouriéroux et al. (2000) developed an IE procedure, where the auxiliary model is chosen to be the true model. In this section, we apply this IE procedure to do bias correction in estimating τ and k. It is important to obtain the bias function via simulations because the bias formula and the bias expansion of the ML and LS estimators studied in this paper are difficult to obtain. The same IE procedure was also used to do bias correction in continuous time models by Phillips and Yu (2009a, c) and in dynamic panel data models by Gouriéroux et al. (2010).

The application of IE for estimating the structural break point proceeds as follows. Given a parameter θ (say τ), we simulate data $\tilde{\boldsymbol{y}}(\theta) = \{\tilde{y}_0^s, \tilde{y}_1^s, \ldots, \tilde{y}_T^s\}$ from the true model, such as, Equation (6) or (10), where $s = 1, ..., S$, and S is the number of simulated paths. Note that T in $\tilde{\boldsymbol{y}}(\theta)$ should be chosen as the same number of the actual data under analysis so that the bias of the original estimator from the actual observations can be calibrated by simulated data. IE then matches the estimate from the actual data with that from the simulated data. To be specific, let $\hat{\theta}_T$ be an estimator of θ from the actual data and $\tilde{\theta}_T^s(\theta)$ be the estimator of θ based on the sth simulated path for some fixed θ. The IE estimator is then defined as

$$\hat{\theta}_{IE,T,S} = \arg\min_{\theta \in \Theta} \left\| \hat{\theta}_T - \frac{1}{S} \sum_{s=1}^{S} \tilde{\theta}_T^s(\theta) \right\|, \tag{14}$$

where $\|\cdot\|$ is some finite-dimensional distance metric and Θ is the compact parameter space. When $S \to \infty$, it is expected that $\frac{1}{S} \sum_{s=1}^{S} \tilde{\theta}_T^s(\theta) \xrightarrow{p} E(\tilde{\theta}_T^s(\theta)) := b_T(\theta)$, where $b_T(\theta)$ is known as the binding function. Then the IE estimator becomes $\hat{\theta}_{IE,T} = \arg\min_{\theta \in \Theta} \left\| \hat{\theta}_T - b_T(\theta) \right\|$. Gouriéroux et al (2000) showed that if $b_T(\theta)$ is an affine function in θ for any T, $\hat{\theta}_{IE,T}$ is exactly mean-unbiased. When the auxiliary model is identical to the true model and $\hat{\theta}_T$ is consistent, Gouriéroux et al (2000) gives non primitive conditions for the second order bias corrections by $\hat{\theta}_{IE,T}$. Arvanitis and Demos (2014) provided more primitive conditions to ensure the validity of moment expansions and the second order bias correction by $\hat{\theta}_{IE,T}$.

In our setup, if τ is the only unknown parameter, we can easily obtain $\widehat{\tau}_{IE,T}$ based on Equation (14). And the IE estimator of k can be obtained as $\hat{k}_{IE,T} = \widehat{\tau}_{IE,T} \times T$. Let the corresponding binding function be $b_T(k) = b_T(\tau) \times T$. Since $\widehat{\tau}_{ML}$ in the continuous time model and $\widehat{\tau}_{LS,T}$ in the discrete time model are consistent when $\varepsilon \to 0$, we can establish the second order bias correction by the IE estimator under some regularity conditions. To derive the asymptotic distribution of the IE estimator, one needs to verify that the binding function is asymptotically locally relatively equicontinuous (Phillips, 2012). If the binding function is indeed asymptotically locally relatively equicontinuous and $\lim_{T\to\infty} E(\widehat{\tau}_T) = \tau_0$ where $\widehat{\tau}_T$ is either $\widehat{\tau}_{ML}$ or $\widehat{\tau}_{LS,T}$, the Delta method can be applied to the original estimator $\widehat{\tau}_{ML}$ and $\widehat{\tau}_{LS,T}$ and the asymptotic theory (including the rate of convergence and the limiting distribution) should be the same as that of the original estimator. Unfortunately, since the pdf of $\widehat{\tau}_T$ is unknown analytically, finding the binding function is only possible numerically. As a result, calculating the derivative of the binding function and verifying asymptotically locally relative equicontinuity of the binding function are very difficult, if not impossible.

When ε is fixed, if the binding function is invertible, that is, $\widehat{\tau}_{IE,T} = b_T^{-1}(\widehat{\tau}_T)$, one may informally apply the Delta method to study the efficiency of the indirect estimator as $\mathrm{Var}(\widehat{\tau}_{IE,T}) \approx \left(\frac{\partial b_T(\tau_0)}{\partial \tau}\right)^{-2} \mathrm{Var}(\widehat{\tau}_T)$. Hence, the efficiency loss (or gain) is measured by $\frac{\partial b_T(\tau_0)}{\partial \tau}$. If $\left|\frac{\partial b_T(\tau_0)}{\partial \tau}\right| < 1$, $\widehat{\tau}_{IE,T}$ has a bigger variance than $\widehat{\tau}_T$. However, if $\left|\frac{\partial b_T(\tau_0)}{\partial \tau}\right| > 1$, $\widehat{\tau}_{IE,T}$ will have a smaller variance than $\widehat{\tau}_T$. As both the simulation results and the large sample theory suggest that τ is over estimated when $\tau_0 < 1/2$ and is under estimated when $\tau_0 > 1/2$, the binding function is expected to be flatter than the 45 degrees line. As a result, $\widehat{\tau}_{IE,T}$ is expected to lose some efficiency compared to $\widehat{\tau}_T$.

As suggested by a referee, we also consider other simulation-based methods to do bias correction, and compare their performance to the IE approach. One alternative bias correction method is the so-called median unbiased estimator (denoted by $\widehat{\tau}_{MU,T}$) as in Andrews (1993) which is obtained by replacing the sample mean in Equation (14) with the sample median. As the finite sample distribution of $\widehat{\tau}_T$ is asymmetric, median might be able to better measure the location than the mean. When the binding function is invertible and monotonic, $\widehat{\tau}_{MU,T}$ is exactly median unbiased. Another bias correction method is the bootstrap method of Efron (1979). Hall (1992) showed that the parametric bootstrap method is an effective method for bias correction. The idea of parametric bootstrap is to generate many bootstrap sample paths, each of which having the same structure as the estimated path from the initial estimation, and then to obtain a new estimate from each bootstrap sample path by applying the same estimation procedure, denoted as $\tau^s_{*T}(\widehat{\tau}_T)$ for $s = 1, ..., S$. Let $\overline{\tau}_{*T}(\widehat{\tau}_T) = \frac{1}{S}\sum_{s=1}^{S}\tau^s_{*T}(\widehat{\tau}_T)$. Then, the bias of $\widehat{\tau}_T$ when $\tau_0 = \widehat{\tau}_T$ is approximated by $\overline{\tau}_{*T}(\widehat{\tau}_T) - \widehat{\tau}_T$, and, hence, the bootstrap estimator is defined as $\widehat{\tau}_{BS,T} = \widehat{\tau}_T - (\overline{\tau}_{*T}(\widehat{\tau}_T) - \widehat{\tau}_T) = 2\widehat{\tau}_T - \overline{\tau}_{*T}(\widehat{\tau}_T)$. Many other simulation-based methods and their comparisons are discussed in Forneron and Ng (2015).

In Model (10) with unknown parameters other than τ_0, using Equation (14) to obtain $\widehat{\tau}_{IE,T}$ and the IE estimators of other parameters simultaneously will be numerically very time consuming. This is because the binding function now becomes a system of multivariate functions and has to be computed via simulations for combinations of some chosen values of all parameters. Given that τ is the parameter of interest, we propose a way to reduce the computational cost in calculating the binding function $b_T(\tau)$. First, it has been shown in the subsection 4.2 that the developed in-fill asymptotic distribution well approximates the finite sample distribution. We therefore suggest to approximate the binding function $b_T(\tau)$ by its limit under the in-fill asymptotic scheme, which is $b(\tau) = E\left(\tau + \arg\max_{u\in(-\tau,1-\tau)}\left[\tilde{B}(u)\right]^2\right)$ where $\tilde{B}(u)$ is defined as in (13). To reduce the dimensionality of the binding function, note that the in-fill asymptotic distribution of $\widehat{\tau}_{LS,T}$ given in (13) depends on the signal-to-noise ratio $\delta^*/(\sigma\varepsilon)$ as a whole, not on the break size δ^*/ε and the standard variance σ individually. We hence propose to replace $\delta^*/(\sigma\varepsilon)$ in $\tilde{B}(u)$ with its LS estimate, and treat it as known when $\tilde{B}(u)$ and $b(\tau)$ are simulated.

6 Monte Carlo Results

In this section, we design three Monte Carlo experiments to examine the bias of the ML estimator of τ in the continuous time model (6) and the LS estimator of k in the discrete time model (1), and compare their performance to the estimators from IE and other simulation-based bias-correction methods.

In the first experiment, data are generated from Model (6), with $\mu = 0$, $\sigma = 1$, $\varepsilon = 1$, $\delta^* = 2, 4, 6$, $\tau_0 = 0.3, 0.5, 0.7$, $dB(t) \overset{iid}{\sim} N(0, h)$ and $h = \frac{1}{10000}$. For each combination of δ^* and τ_0, we obtain the ML estimate of τ from (7) and several biased corrected estimates of τ with $S = 10,000$.[3] Table 2 reports the bias, the standard error, and the root mean squared errors (RMSE) of the ML estimator, the IE estimator, the median unbiased (MU) estimator, and the parametric bootstrap (PB) estimator, obtained from 100,000 replications. Some observations can be obtained from the table. First, when $\tau_0 = 0.5$, the ML estimator does not have any noticeable bias in all cases. However, when $\tau_0 \neq 0.5$, the ML estimator suffers from a bias problem. For example, when $\tau_0 = 0.3$ and $\frac{\delta^*}{\sigma\varepsilon} = 2$, the bias is 0.0912, which is about 30% of the true value. This is very substantial. In general, the bias becomes larger when τ_0 is further away from 0.5, or when the signal-to-noise ratio gets smaller. To the best of our knowledge, such a bias has not been discussed in the literature. Second, in all cases when $\tau_0 \neq 0.5$, the IE approach substantially reduces the bias. For example, when $\frac{\delta^*}{\sigma\varepsilon} = 2$ and $\tau_0 = 70\%$, IE removes about two thirds of the bias in the ML estimator. Third, the bias reduction by IE comes with a cost of a higher variance, which causes the RMSE of the IE estimator slightly higher than its ML counterpart. Finally. compared with IE, the MU estimator is less effective for bias reduction but is more efficient in terms of variance. In terms of RMSE, the MU estimator performs better. This finding is consistent with what was reported in Tables 7-8 of Phillips and Yu (2009a) for a continuous time model. However, compared with IE, the PB estimator performs similarly in terms of bias reduction but increases the variance more in almost all cases.

In the second experiment, data are generated from Model (1), with $\mu = 0$, $\sigma = 1$, $\delta = 0.2$, 0.4, 0.6, $\tau_0 = 0.3, 0.5, 0.7$, $\epsilon_t \overset{iid}{\sim} N(0, 1)$, $T = 100$.[4] For each combination of δ and τ_0, we obtain the LS estimate of k from (11) and several biased corrected estimates of k with $S = 10,000$. Table 3 reports the bias, the standard error, and the RMSE of the LS estimator, defined in (11), and the three simulation-based estimators, obtained

[3]We also try other values for H, such as $H = 1,000$ and $5,000$. The results are almost unchanged.

[4]We also try other values for T in the second and the third experiments, such as $T = 80$ and 120. The results remain qualitatively unchanged.

Table 2: Monte Carlo comparisons of bias and RMSE of the ML estimator, the MU estimator, the PB estimator, and the IE estimator for the continuous time model (6). The number of simulated paths is set to be 10,000. The number of replications is set at 100,000.

Case		Bias				Standard Error				RMSE			
$\frac{\delta^*}{\sigma\varepsilon}$	τ_0	ML	MU	PB	IE	ML	MU	PB	IE	ML	MU	PB	IE
2	0.3	0.0912	0.0751	0.0433	0.0378	0.2768	0.3128	0.4101	0.3677	0.2914	0.3217	0.4124	0.3696
2	0.5	-0.0006	-0.0004	0.0042	-0.0008	0.2737	0.3090	0.4031	0.3635	0.2737	0.3090	0.4031	0.3635
2	0.7	-0.0907	-0.0739	-0.0341	-0.0373	0.2763	0.3122	0.4044	0.3670	0.2908	0.3208	0.4058	0.3689
4	0.3	0.0313	0.0282	0.0070	0.0026	0.1874	0.1918	0.2229	0.2151	0.1900	0.1939	0.2231	0.2152
4	0.5	0.0002	0.0001	0.0033	0.0001	0.1902	0.1945	0.2215	0.2190	0.1902	0.1945	0.2215	0.2190
4	0.7	-0.0305	-0.0272	0.0001	-0.0012	0.1865	0.1911	0.2177	0.2146	0.1889	0.1930	0.2177	0.2146
6	0.3	0.0079	0.0075	0.0001	-0.0003	0.1180	0.1185	0.1254	0.1242	0.1183	0.1187	0.1254	0.1242
6	0.5	0.0007	0.0006	0.0013	0.0008	0.1228	0.1233	0.1287	0.1288	0.1228	0.1233	0.1287	0.1288
6	0.7	-0.0074	-0.0069	0.0017	0.0012	0.1176	0.1182	0.1238	0.1242	0.1179	0.1184	0.1238	0.1242

Table 3: Monte Carlo comparisons of bias and RMSE of the LS estimator, the MU estimator, the PB estimator, and IE estimator for the discrete time model (1) when only τ_0 is unknown. The sample size is set to be $T = 100$. The number of simulated paths is set to be 10,000. The number of replications is set at 100,000.

Case			Bias				Standard Error				RMSE			
$\frac{\delta}{\sigma}$	τ_0	k_0	LS	MU	PB	IE	LS	MU	PB	IE	LS	MU	PB	IE
0.2	0.3	30	9.0912	7.1469	4.1359	4.0070	27.2351	30.4276	40.3154	36.3728	28.7124	31.2557	40.5270	36.5928
0.2	0.5	50	-0.0806	-0.4141	0.2712	0.1831	26.9619	30.0310	40.0871	36.0119	26.9620	30.0339	40.0880	36.0124
0.2	0.7	70	-9.2084	-8.0359	-3.4942	-3.5321	27.3270	30.3804	40.5756	36.4723	28.8368	31.4253	40.7258	36.6429
0.4	0.3	30	3.0708	2.8233	1.0153	0.0599	18.3912	18.8147	21.9138	21.2435	18.6458	19.0253	21.9273	21.2436
0.4	0.5	50	-0.0283	-0.0392	0.8136	-0.0271	18.8135	19.1832	22.1478	21.7718	18.8135	19.1832	22.1628	21.7718
0.4	0.7	70	-3.0490	-2.8399	0.5832	-0.0335	18.4068	18.8136	21.6883	21.3029	18.8135	19.0267	21.6961	21.3029
0.6	0.3	30	0.7750	0.7670	0.9197	-0.1841	11.6529	11.6739	12.4560	12.3339	11.6786	11.6990	12.4899	12.3353
0.6	0.5	50	0.0031	0.0022	0.9562	0.0245	12.1705	12.1820	12.9329	12.7771	12.1705	12.1820	12.9682	12.7771
0.6	0.7	70	-0.8016	-0.7974	0.9121	0.2016	11.6490	11.6631	12.6650	12.3273	11.6765	11.6903	12.6978	12.3289

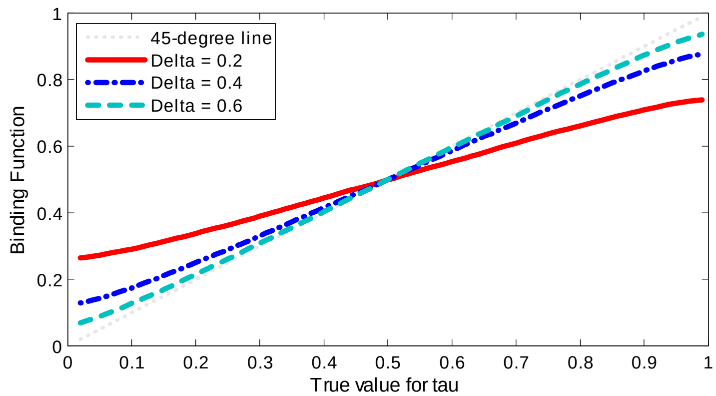

Figure 8: Binding functions of the LS estimator for discrete time model with $T = 100$

from 100,000 replications. The conclusions drawn from Table 3 are nearly identical to those from Table 2.

To understand why IE increases the variance relative to the original estimator in these two experiments, we plot the binding function in the second experiment in Figure 8. In Figure 8 we also plot the 45 degrees line for the purpose of comparison. Several conclusions can be made. First, every binding function passes through the 45 degrees line when $\tau_0 = 0.5$, suggesting that no bias exists when $\tau_0 = 0.5$. Second, the binding functions are flatter than the 45 degrees line in all cases, explaining why the variance of the IE estimator is larger than that of the ML estimator. The smaller the signal-to-noise ratio, the flatter the binding function and hence the bigger the loss in efficiency. Third, no binding function is exactly a straight line. Nonlinearity can be found near the two boundary points. Consequently, according to Gouriéroux et al. (2000), the IE estimator is not exactly mean unbiased. Although not plotted, the binding function in the first experiment shares the same characteristics.

In the third experiment, data are generated from Model (1), with $\mu = 0$, $\sigma = 1$, $\delta = 0.2,\ 0.4,\ 0.6$, $\tau_0 = 0.3, 0.5, 0.7$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, $T = 100$. Different from the second experiment, all the parameters, including τ, are assumed to be unknown. For each combination of $\frac{\delta}{\sigma}$ and τ_0, we obtain the LS estimate of k from (12) and the indirect estimate of k with $S = 10,000$. Table 4 reports the bias, the standard error, and the RMSE of $\widehat{k}_{LS,T}$ defined in (12) and $\widehat{k}_{IE,T}$ proposed in the end of Section 5, obtained from 10,000 replications. Several conclusions can be made from Table 4. First, when $\tau_0 = 0.5$, the LS estimator does not have any noticeable bias. However, when $\tau_0 \neq 0.5$, the LS estimator suffers from a severe bias problem. In general, the bias becomes larger when τ_0 is further away from 0.5 or when the signal-to-noise ratio gets smaller. Second,

Table 4: Monte Carlo comparisons of bias, standard error and RMSE of the LS estimator and the IE estimator when more parameters other than τ_0 are unknown. The sample size is set to be $T = 100$. The number of simulated paths is set at 10,000 for indirect estimation. The number of replications is set at 10,000.

Case			Bias		Standard Error		RMSE	
$\frac{\delta}{\sigma}$	τ_0	k_0	LS	IE	LS	IE	LS	IE
0.2	0.3	30	17.6198	16.8012	34.3571	33.0711	38.6117	37.0941
0.2	0.5	50	0.5371	0.5797	33.2781	31.8026	33.2824	31.8078
0.2	0.7	70	-16.6796	-15.6972	34.2376	32.9234	38.0844	36.4741
0.4	0.3	30	10.7315	8.1827	28.3439	28.8414	30.3074	29.9797
0.4	0.5	50	0.1786	0.2900	25.8011	25.9128	25.8017	25.9144
0.4	0.7	70	-11.0609	-8.4541	28.5325	28.8550	30.6015	30.0679
0.6	0.3	30	5.2745	2.1591	20.3822	21.4932	21.0536	21.6014
0.6	0.5	50	0.0426	0.0837	17.5939	18.3072	17.5939	18.3073
0.6	0.7	70	-5.2904	-2.1938	20.4234	21.3943	21.0975	21.5064

IE can reduce the bias in all cases. For example, when $\frac{\delta}{\sigma} = 0.6$ and $\tau_0 = 0.7$, IE removes about 59% of the bias of the LS estimator. Moreover, the variance of the IE estimator is comparable to that of the LS estimator. Overall, the RMSE of the IE estimator is similar to its LS counterpart. Unfortunately, analyzing the behavior of the binding function here is complicated for two reasons. First, we replace the signal-to-noise ratio $\delta^*/(\sigma\varepsilon)$ by its LS estimator which inevitably changes the curvature of the binding function in obtaining the IE estimator. Second, in general the binding function is a system of functions that depend on all unknown parameters.

7 Conclusions

This paper is concerned about the in-fill asymptotic approximation to the exact distribution in the estimation of structural break point in mean. We find that the exact distributions of the traditional estimators of structural break point are often asymmetric and have trimodality both in the continuous time model and in the discrete time model. It is also found that the traditional estimators are biased. Unfortunately, the literature on structural breaks has always focused the attention on developing asymptotic theory with a time span being assumed to go to infinity. The long-span limiting distribution developed in the literature is symmetric and has the true break point as the unique mode. As a result, it provides poor approximations to the exact distribution in many empirically relevant cases.

In this paper we address the finite sample problem in several aspects. First, we derive the exact distribution of the ML estimator of the structural break point in a continuous time model when a continuous record is available. It is shown that the exact distribution has trimodality, regardless of the location of the break. When the true break point is in the middle of the sample, the exact distribution is symmetric. However, when the true break point occurs earlier (later) than the middle of the sample, the exact distribution is skewed to the right (left), leading to a positive (negative) bias in the ML estimator.

In a discrete time model with a break in mean, we continue to find the trimodality and asymmetry in the finite sample distribution of the LS estimator of the structural break point. To better approximate the finite sample distribution, we deviate from the literature by considering a continuous time approximation to the discrete time model and developing an in-fill asymptotic theory. For the discrete time model with the break point being the only unknown parameter, the in-fill asymptotic distribution is the same as the exact distribution in the continuous time model. For the discrete time model with more unknown parameters, the in-fill asymptotic distribution is new to the literature. We show that this distribution has trimodality and is asymmetric when the true break point is not in the middle of the sample and the in-fill asymptotic distribution better approximates the finite sample distribution than the long-span limiting distribution developed in the literature.

Given that the exact distribution suggests a substantial bias in the ML/LS estimators, to reduce the bias, we propose to use the IE technique to estimate the break point. Indirect estimation inherits the asymptotic properties of the original estimator but reduces the finite sample bias. Monte Carlo results show that the IE procedure is effective in reducing the bias in the commonly used break point estimators.

The models considered in this paper are very simple in nature. Also, the estimators considered are based on the full sample. Real time (and hence subsample) estimators tend to have more serious finite sample problems. Further studies on developing better approximations to the finite sample distribution for more realistic models and real time estimators are needed.

Appendix

Proof of Theorem 3.1: (a) Note that

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

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

$$\log\left(\frac{dP_\tau}{dP_{\tau_0}}\right) = \int_0^1 \frac{\delta^*}{\sigma\varepsilon}\left(1_{[t>\tau]} - 1_{[t>\tau_0]}\right) dB(t) - \frac{1}{2}\int_0^1 \left(\frac{\delta^*}{\sigma\varepsilon}\right)^2 \left(1_{[t>\tau]} - 1_{[t>\tau_0]}\right)^2 dt.$$

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

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

When $\tau > \tau_0$, we have

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

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

$$\log\left(\frac{dP_\tau}{dP_{\tau_0}}\right) = \frac{\delta^*}{\sigma\varepsilon}\left(B(\tau_0) - B(\tau)\right) - \frac{1}{2}\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2 |\tau - \tau_0|.$$

This implies that the ML estimator of break point is

$$\widehat{\tau}_{ML} = \arg\max_{\tau\in(0,1)} \left\{\frac{\delta^*}{\sigma\varepsilon}\left(B(\tau_0) - B(\tau)\right) - \frac{1}{2}\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2 |\tau - \tau_0|\right\},$$

which leads to

$$\widehat{\tau}_{ML} - \tau_0 = \arg\max_{s\in(-\tau_0, 1-\tau_0)} \left\{\frac{\delta^*}{\sigma\varepsilon}\left(B(\tau_0) - B(\tau_0 + s)\right) - \frac{1}{2}\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2 |s|\right\}.$$

Let $W(\cdot)$ be the two-sided Brownian motion defined in (8). We then have

$$
\begin{aligned}
\widehat{\tau}_{ML}-\tau_0 &= \arg\max_{s\in(-\tau_0,1-\tau_0)}\left\{\frac{\delta^*}{\sigma\varepsilon}W(s)-\frac{1}{2}\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2|s|\right\}\\
&\overset{d}{=} \arg\max_{s\in(-\tau_0,1-\tau_0)}\left\{W\left(s\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2\right)-\frac{1}{2}\left|s\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2\right|\right\}\\
&\overset{d}{=} \left(\frac{\delta^*}{\sigma\varepsilon}\right)^{-2}\arg\max_{u\in\left(-\tau_0\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2,(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2\right)}\left\{W(u)-\frac{|u|}{2}\right\},
\end{aligned}
$$

which gives the result in Part (a) of Theorem 3.1 immediately.

(b) It is a straightforward result of Part (a).

Proof of Theorem 4.1: (a) Let $\Gamma(k)=-(\frac{\delta^*}{\varepsilon})\sqrt{h}\sum_{t=1}^{k}\left(Z_t-\mu\sqrt{h}\right)+\left(\frac{\delta^*}{\varepsilon}\right)^2hk/2$. Then, the LS estimator $\widehat{k}_{LS,T}$ defined in (11) can be expressed as

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

As $T=1/h$, $\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^2\left(\widehat{k}_{LS,T}-k_0\right)=(\delta^*/\varepsilon)^2(\widehat{\tau}_{LS,T}-\tau_0)=O_p(1)$ takes values in the interval of $\left(-\tau_0(\delta^*/\varepsilon)^2,(1-\tau_0)(\delta^*/\varepsilon)^2\right)$. Therefore, to study the in-fill asymptotic distribution of $\widehat{k}_{LS,T}$ we only need to examine the behavior of $\Gamma(k)-\Gamma(k_0)$ for those k in the neighborhood of k_0 such that $k=\left\lfloor k_0+s\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\rfloor$ with $s\in\left(-\tau_0(\delta^*/\varepsilon)^2,(1-\tau_0)(\delta^*/\varepsilon)^2\right)$, where $\lfloor\cdot\rfloor$ is the integer-valued function.

When $k\le k_0$, $h\to 0$ with a fixed ε, we have, for any $s\in\left(-\tau_0(\delta^*/\varepsilon)^2,0\right]$,

$$
\begin{aligned}
&\Gamma(k)-\Gamma(k_0)\\
=\;&(\delta^*/\varepsilon)\sqrt{h}\sum_{t=k+1}^{k_0}\left(Z_t-\mu\sqrt{h}\right)-(\delta^*/\varepsilon)^2\frac{k_0-k}{2}h\\
=\;&(\frac{\delta^*}{\varepsilon})\sqrt{h}\sum_{t=\left\lfloor k_0+s\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\rfloor+1}^{k_0}\epsilon_t-\left(\frac{\delta^*}{\varepsilon}\right)^2\frac{\left(k_0-\left\lfloor k_0+s\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\rfloor\right)}{2}h\\
\Rightarrow\;&\sigma W_1(-s)-\frac{|s|}{2},
\end{aligned}
$$

where $W_1(\cdot)$ is a standard Brownian motion, the second equation is from the fact that $Z_t-\mu\sqrt{h}=\epsilon_t\sim$ i.i.d.$(0,\sigma^2)$ for $t\le k_0$, and the last convergence result comes from a straightforward application of the functional central limit theory (FCLT) for the i.i.d. sequence.

When $k > k_0$, for any $s \in \left(0, (1-\tau_0)(\delta^*/\varepsilon)^2\right)$, we have

$$
\begin{aligned}
& \Gamma(k) - \Gamma(k_0) \\
= & -(\delta^*/\varepsilon)\sqrt{h} \sum_{t=k_0+1}^{k} \left(Z_t - \mu\sqrt{h}\right) + (\delta^*/\varepsilon)^2 \frac{k-k_0}{2} h \\
= & -(\delta^*/\varepsilon)\sqrt{h} \sum_{t=k_0+1}^{k} \left(Z_t - \mu\sqrt{h} - (\delta^*/\varepsilon)\sqrt{h}\right) - (\delta^*/\varepsilon)^2 \frac{k-k_0}{2} h \\
= & -(\frac{\delta^*}{\varepsilon})\sqrt{h} \sum_{t=k_0+1}^{\left\lfloor k_0 + s\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\rfloor} \epsilon_t - \left(\frac{\delta^*}{\varepsilon}\right)^2 \frac{\left(\left\lfloor k_0 + s\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\rfloor - k_0\right)}{2} h \\
\Rightarrow & -\sigma W_2(s) - \frac{|s|}{2} \overset{d}{=} \sigma W_2(s) - \frac{|s|}{2},
\end{aligned}
$$

where $W_2(\cdot)$ is a standard Brownian motion, and the third equation comes from the fact that $Z_t - \mu\sqrt{h} - (\delta^*/\varepsilon)\sqrt{h} = \epsilon_t \sim$ i.i.d.$(0, \sigma^2)$ for $t > k_0$.

It can be seen that $W_1(\cdot)$ and $W_2(\cdot)$ are determined by ϵ_t before and after k_0 respectively. Therefore, they are two independent Brownian motions. Let $W(\cdot)$ be the two-sided Brownian motion defined in (8). We then have

$$
\Gamma(k) - \Gamma(k_0) = \Gamma\left(\left\lfloor k_0 + s\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\rfloor\right) - \Gamma(k_0) \Rightarrow \sigma W(s) - \frac{|s|}{2}.
$$

Applying the continuous mapping theorem to the arg max function leads to

$$
\begin{aligned}
T\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^2 (\widehat{\tau}_{LS,T} - \tau_0) & \xrightarrow{d} \arg \max_{s \in \left(-\tau_0(\delta^*/\varepsilon)^2, (1-\tau_0)(\delta^*/\varepsilon)^2\right)} \left\{\sigma W(s) - \frac{|s|}{2}\right\} \\
& = \arg \max_{s \in \left(-\tau_0(\delta^*/\varepsilon)^2, (1-\tau_0)(\delta^*/\varepsilon)^2\right)} \left\{W\left(s/\sigma^2\right) - \frac{|s|}{2\sigma^2}\right\} \\
& \overset{d}{=} \sigma^2 \arg \max_{u \in \left(-\tau_0\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2, (1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2\right)} \left\{W(u) - \frac{|u|}{2}\right\},
\end{aligned}
$$

which gives the final result in Part (a) of Theorem 4.1 immediately. For a rigorous treatment of the continuous mapping theorem for the arg max function, see Kim and Pollard (1990).

(b) It takes three steps to derive the in-fill asymptotic distribution under the scheme that $h \to 0$ and $\varepsilon \to 0$ simultaneously and $(\delta^*/\varepsilon)\sqrt{h} \to 0$. The first step is to prove that $\widehat{\tau}_{LS,T} \xrightarrow{p} \tau_0$.

Note that when $k \leq k_0$,

$$
E(\Gamma(k)) = -(\frac{\delta^*}{\varepsilon})\sqrt{h}\sum_{t=1}^{k} E\left(Z_t - \mu\sqrt{h}\right) + \left(\frac{\delta^*}{\varepsilon}\right)^2 \frac{k}{2} h = \left(\frac{\delta^*}{\varepsilon}\right)^2 \frac{k}{2} h,
$$

and when $k > k_0$,

$$\begin{aligned}
E\left(\Gamma\left(k\right)\right) &= -(\frac{\delta^*}{\varepsilon})\sqrt{h}\sum_{t=1}^{k} E\left(Z_t - \mu\sqrt{h}\right) + \left(\frac{\delta^*}{\varepsilon}\right)^2 \frac{k}{2}h \\
&= -(\frac{\delta^*}{\varepsilon})\sqrt{h}\sum_{t=k_0+1}^{k} E\left(Z_t - \mu\sqrt{h}\right) + \left(\frac{\delta^*}{\varepsilon}\right)^2 \frac{k}{2}h \\
&= -(\frac{\delta^*}{\varepsilon})^2\left(k - k_0\right)h + \left(\frac{\delta^*}{\varepsilon}\right)^2 \frac{k}{2}h = \left(\frac{\delta^*}{\varepsilon}\right)^2 \frac{\left(2k_0 - k\right)}{2}h.
\end{aligned}$$

We then have,

$$E\left(\Gamma\left(k_0\right)\right) - E\left(\Gamma\left(k\right)\right) = \begin{cases} \left(\delta^*/\varepsilon\right)^2\left(k_0 - k\right)h/2 = \left(\delta^*/\varepsilon\right)^2\left(\tau_0 - \tau\right)/2 & \text{if } k \leq k_0 \\ \left(\delta^*/\varepsilon\right)^2\left(k - k_0\right)h/2 = \left(\delta^*/\varepsilon\right)^2\left(\tau - \tau_0\right)/2 & \text{if } k > k_0 \end{cases}$$

which leads to $E\left(\Gamma\left(k_0\right)\right) - E\left(\Gamma\left(k\right)\right) = \left(\delta^*/\varepsilon\right)^2\left|\tau - \tau_0\right|/2$ for any $1 \leq k < T$.

It is easy to see that for any k

$$\begin{aligned}
\Gamma\left(k\right) - \Gamma\left(k_0\right) &= \Gamma\left(k\right) - E\left(\Gamma\left(k\right)\right) + E\left(\Gamma\left(k\right)\right) - E\left(\Gamma\left(k_0\right)\right) - \Gamma\left(k_0\right) + E\left(\Gamma\left(k_0\right)\right) \\
&\leq \left|\Gamma\left(k\right) - E\left(\Gamma\left(k\right)\right)\right| + \left|\Gamma\left(k_0\right) - E\left(\Gamma\left(k_0\right)\right)\right| + E\left(\Gamma\left(k\right)\right) - E\left(\Gamma\left(k_0\right)\right).
\end{aligned}$$

As a result, $E\left(\Gamma\left(k_0\right)\right) - E\left(\Gamma\left(k\right)\right) \leq \left|\Gamma\left(k\right) - E\left(\Gamma\left(k\right)\right)\right| + \left|\Gamma\left(k_0\right) - E\left(\Gamma\left(k_0\right)\right)\right| - \left\{\Gamma\left(k\right) - \Gamma\left(k_0\right)\right\}$. Given that $\widehat{k}_{LS,T} = \arg\max\left\{\Gamma\left(k\right)\right\}$, we then have,

$$\left(\delta^*/\varepsilon\right)^2\left|\widehat{\tau}_{LS,T} - \tau_0\right|/2 \leq \left|\Gamma\left(\widehat{k}_{LS,T}\right) - E\left(\Gamma\left(\widehat{k}_{LS,T}\right)\right)\right| + \left|\Gamma\left(k_0\right) - E\left(\Gamma\left(k_0\right)\right)\right|.$$

Note that, for any $1 \leq k < T$, $\Gamma\left(k\right) - E\left(\Gamma\left(k\right)\right) = -(\frac{\delta^*}{\varepsilon}\sqrt{h})\sum_{t=1}^{k}\epsilon_t$ where $\epsilon_t \sim$i.i.d.$(0, \sigma^2)$. Because $Var\left(-(\frac{\delta^*}{\varepsilon}\sqrt{h})\sum_{t=1}^{k}\epsilon_t\right) = (\frac{\delta^*}{\varepsilon}\sqrt{h})^2 k\sigma^2 = (\delta^*/\varepsilon)^2\tau\sigma^2$ with $\tau = k/T \in (0,1)$, we have $\Gamma\left(k\right) - E\left(\Gamma\left(k\right)\right) = O_p\left(\delta^*/\varepsilon\right)$ as $\varepsilon \to 0$. Therefore, as $\varepsilon \to 0$,

$$\begin{aligned}
\left|\widehat{\tau}_{LS,T} - \tau_0\right| &\leq 2\left(\delta^*/\varepsilon\right)^{-2}\left\{\left|\Gamma\left(\widehat{k}_{LS,T}\right) - E\left(\Gamma\left(\widehat{k}_{LS,T}\right)\right)\right| + \left|\Gamma\left(k_0\right) - E\left(\Gamma\left(k_0\right)\right)\right|\right\} \\
&= 2\left(\delta^*/\varepsilon\right)^{-2}\left\{O_p\left(\delta^*/\varepsilon\right) + O_p\left(\delta^*/\varepsilon\right)\right\} = O_p\left(\varepsilon/\delta^*\right) \xrightarrow{p} 0.
\end{aligned}$$

The first step is done.

The second step is to prove that $\widehat{\tau}_{LS,T} - \tau_0 = O_p\left(\left(\sqrt{T}\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right)$. Choose a $\gamma > 0$ such that $\tau_0 \in (\gamma, 1-\gamma)$. Since $\widehat{\tau}_{LS,T}$ is consistent, for every $\Delta > 0$, $Pr\left\{\widehat{\tau}_{LS,T} \notin (\gamma, 1-\gamma)\right\} < \Delta$ when $h \to 0$ and $\varepsilon \to 0$ simultaneously with $\left(\delta^*/\varepsilon\right)\sqrt{h} \to 0$. Thus, we now only need to examine the behavior of $\Gamma\left(k\right)$ over those k for which $T\gamma < k < T\left(1-\gamma\right)$. To prove

$\widehat{\tau}_{LS,T}-\tau_0 = O_p\left(\left(\sqrt{T}\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right)$, we shall prove $Pr\left\{|\widehat{\tau}_{LS,T}-\tau_0| \geq M\left(\sqrt{T}\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\} \rightarrow 0$ when $M \rightarrow \infty$, $h \rightarrow 0$ and $\varepsilon \rightarrow 0$ simultaneously with $(\delta^*/\varepsilon)\sqrt{h} \rightarrow 0$.

For every $M > 0$, define $D_{T,M} = \left\{k \mid T\gamma < k < T(1-\gamma), |k-k_0| \geq M\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\}$. We then have

$$
\begin{aligned}
& Pr\left\{|\widehat{\tau}_{LS,T}-\tau_0| \geq M\left(\sqrt{T}\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\} \\
\leq\ & Pr\left\{\widehat{\tau}_{LS,T} \notin (\gamma, 1-\gamma)\right\} + Pr\left\{\widehat{\tau}_{LS,T} \in (\gamma, 1-\gamma), |\widehat{\tau}_{LS,T}-\tau_0| \geq M\left(\sqrt{T}\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\} \\
<\ & \Delta + Pr\left\{\sup_{k \in D_{T,M}} \{\Gamma(k)\} \geq \Gamma(k_0)\right\} \\
=\ & \Delta + P_1 \quad \text{with} \quad P_1 = \Pr\left\{\sup_{k \in D_{T,M}} \{\Gamma(k) - \Gamma(k_0)\} \geq 0\right\}.
\end{aligned}
$$

The event $\Gamma(k) - \Gamma(k_0) \geq 0$ implies

$$
\begin{aligned}
\Gamma(k) - E(\Gamma(k)) - \{\Gamma(k_0) - E(\Gamma(k_0))\} &\geq E(\Gamma(k_0)) - E(\Gamma(k)) \\
&= \left(\frac{\delta^*}{\varepsilon}\right)^2 \frac{|\tau-\tau_0|}{2} = \left(\frac{\delta^*}{\varepsilon}\right)^2 \frac{|k-k_0|}{2T}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
& \Gamma(k) - E(\Gamma(k)) - \{\Gamma(k_0) - E(\Gamma(k_0))\} \\
=\ & -\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)\sum_{t=1}^{k}\epsilon_t + \left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)\sum_{t=1}^{k_0}\epsilon_t = \begin{cases} \left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)\sum_{t=k+1}^{k_0}\epsilon_t & \text{when } k < k_0 \\ -\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)\sum_{t=k_0+1}^{k}\epsilon_t & \text{when } k > k_0 \end{cases}.
\end{aligned}
$$

Then

$$
\begin{aligned}
P_1 &\leq \Pr\left\{\sup_{k \in D_{T,M}} \frac{1}{|k-k_0|}\left(-\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)\sum_{t=1}^{k}\epsilon_t + \left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)\sum_{t=1}^{k_0}\epsilon_t\right) \geq \left(\frac{\delta^*}{\varepsilon}\right)^2 \frac{1}{2T}\right\} \\
&\leq P_1(k < k_0) + P_1(k > k_0)
\end{aligned}
$$

where $P_1(k<k_0) = \Pr\left\{\sup_{\{k<k_0 \text{ and } k \in D_{T,M}\}} \frac{1}{|k-k_0|}\left(\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)\sum_{t=k+1}^{k_0}\epsilon_t\right) \geq \left(\frac{\delta^*}{\varepsilon}\right)^2/2T\right\}$ and $P_1(k>k_0) = \Pr\left\{\sup_{\{k>k_0 \text{ and } k \in D_{T,M}\}} \frac{1}{|k-k_0|}\left(-\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)\sum_{t=k_0+1}^{k}\epsilon_t\right) \geq \left(\frac{\delta^*}{\varepsilon}\right)^2/2T\right\}$.

For the case of $k < k_0$ and $k \in D_{T,M}$, we have $T\gamma < k < T\tau_0 - M\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}$. Then

$$\begin{aligned}
P_1(k<k_0) &= \Pr\left\{\sup_{T\gamma<k<T\tau_0-M\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}} \frac{1}{|k-k_0|}\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\sum_{t=k+1}^{k_0}\epsilon_t\right) \geq \left(\frac{\delta^*}{\varepsilon}\right)^2 \frac{1}{2T}\right\} \\
&= \Pr\left\{\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-1} \sup_{T\gamma<k<T\tau_0-M\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}} \left(\frac{1}{|k-k_0|}\sum_{t=k+1}^{k_0}\epsilon_t\right) \geq \frac{1}{2}\right\} \\
&\leq \Pr\left\{\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-1} \sup_{|k-k_0|>M\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}} \left(\frac{1}{|k-k_0|}\sum_{t=k+1}^{k_0}\epsilon_t\right) \geq \frac{1}{2}\right\}.
\end{aligned}$$

From the Hájek and Rényi inequality as in Hájek and Rényi (1955), it is easy to get that, when $M \to \infty$ and $(\delta^*/\varepsilon)\sqrt{h} \to 0$,

$$\sup_{\left\{|k-k_0|>M\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\}} \left(\frac{1}{|k-k_0|}\sum_{t=k+1}^{k_0}\epsilon_t\right) = O_p\left(\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)/\sqrt{M}\right),$$

which leads to

$$\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-1} \sup_{\left\{|k-k_0|>M\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\}} \left(\frac{1}{|k-k_0|}\sum_{t=k+1}^{k_0}\epsilon_t\right) = O_p\left(1/\sqrt{M}\right) \to 0.$$

Therefore, $P_1(k<k_0) \to 0$.

Similar method can be used to prove $P_2(k<k_0) \to 0$. Then we get $P_1 \to 0$, and, therefore, $\widehat{\tau}_{LS,T} - \tau_0 = O_p\left(\left(\sqrt{T}\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right)$ when $h \to 0$ and $\varepsilon \to 0$ simultaneously with $(\delta^*/\varepsilon)\sqrt{h} \to 0$. The second step in done.

Given $\widehat{\tau}_{LS,T} - \tau_0 = O_p\left(\left(\sqrt{T}\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right)$, we have $\widehat{k}_{LS,T} - k_0 = O_p\left(\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right)$. Therefore, to derive the in-fill asymptotic distribution of $\widehat{k}_{LS,T}$, we only need to examine the behavior of $\Gamma(k) - \Gamma(k_0)$ for those k in the neighborhood of k_0 such that $k = \left\lfloor k_0 + s\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\rfloor$, where s varies in an arbitrary bounded interval. Then, for any $M > 0$ and $s = u\sigma^2 \in (-M, M)$, repeating the procedure in the proof of (a), which is counted as the third step of this proof, gives

$$\begin{aligned}
T\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^2 (\widehat{\tau}_{LS,T} - \tau_0) &\xrightarrow{d} \arg\max_{s\in(-M,M)} \left\{\sigma W(s) - \frac{|s|}{2}\right\} \\
&\overset{d}{=} \sigma^2 \arg\max_{u\in(-M/\sigma^2, M/\sigma^2)} \left\{W(u) - \frac{|u|}{2}\right\}.
\end{aligned}$$

As M can be chosen arbitrarily, the result in part (b) of Theorem 4.1 is proved.

Proof of Theorem 4.2: (a) From Model (10) we have $Z_t - \mu\sqrt{h} = \epsilon_t \sim$ i.i.d.$(0, \sigma^2)$ for $t \leq k_0$ and $Z_t - \mu\sqrt{h} - (\delta^*/\varepsilon)\sqrt{h} = \epsilon_t \sim$ i.i.d.$(0, \sigma^2)$ for $t > k_0$. Then, for $k \leq k_0$,

$$\begin{aligned}
& \overline{Z}_k - \overline{Z}_k^* \\
= \; & \frac{1}{k}\sum_{t=1}^{k} Z_t - \frac{1}{T-k}\sum_{t=k+1}^{T} Z_t = \frac{1}{k}\sum_{t=1}^{k} Z_t - \frac{1}{T-k}\left(\sum_{t=k+1}^{k_0} Z_t + \sum_{t=k_0+1}^{T} Z_t\right) \\
= \; & \frac{1}{k}\sum_{t=1}^{k} \epsilon_t + \mu\sqrt{h} - \frac{1}{T-k}\left((k_0 - k)\,\mu\sqrt{h} + (T - k_0)\left(\mu + \frac{\delta^*}{\varepsilon}\right)\sqrt{h} + \sum_{t=k+1}^{T} \epsilon_t\right) \\
= \; & \frac{1}{k}\sum_{t=1}^{k} \epsilon_t - \frac{1}{T-k}\sum_{t=k+1}^{T} \epsilon_t - \frac{T-k_0}{T-k}\frac{\delta^*}{\varepsilon}\sqrt{h}.
\end{aligned}$$

Similarly, for $k > k_0$ we have

$$\overline{Z}_k - \overline{Z}_k^* = \frac{1}{k}\sum_{t=1}^{k} \epsilon_t - \frac{1}{T-k}\sum_{t=k+1}^{T} \epsilon_t - \frac{k_0}{k}\frac{\delta^*}{\varepsilon}\sqrt{h}.$$

The LS estimator defined in (12) can be identically expressed as

$$\widehat{k}_{LS,T} = \arg\max_{k=1,\ldots,T-1}\left\{\left[\sqrt{T}V_k\left(Z_t\right)\right]^2\right\} \quad \text{with } \left[V_k\left(Z_t\right)\right]^2 = \frac{k\,(T-k)}{T^2}\left(\overline{Z}_k - \overline{Z}_k^*\right)^2.$$

When $h \to 0$ with a fixed ε, we have $\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^2\left(\widehat{k}_{LS,T} - k_0\right) = (\delta^*/\varepsilon)^2\left(\widehat{\tau}_{LS,T} - \tau_0\right) = O_p(1)$ taking values in the interval of $\left(-\tau_0\left(\delta^*/\varepsilon\right)^2, (1-\tau_0)\left(\delta^*/\varepsilon\right)^2\right)$. Therefore, to study the in-fill asymptotic distribution of $\widehat{k}_{LS,T}$ we only need to examine the behavior of $\left[\sqrt{T}V_k\left(Z_t\right)\right]^2$ for those k in the neighborhood of k_0 such that $k = \left\lfloor k_0 + s\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\rfloor$ with $s \in \left(-\tau_0\left(\delta^*/\varepsilon\right)^2, (1-\tau_0)\left(\delta^*/\varepsilon\right)^2\right)$. Then, for any fixed s, when $h \to 0$, it has $k \to \infty$ with $k/T \to \tau_0 + s\left(\frac{\delta^*}{\varepsilon}\right)^{-2} = \tau_0 + u$ and $T - k \to \infty$ with $(T-k)/T \to 1 - \tau_0 - s\left(\frac{\delta^*}{\varepsilon}\right)^{-2} = 1 - \tau_0 - u$, where $u = s\left(\frac{\delta^*}{\varepsilon}\right)^{-2} \in (-\tau_0, 1-\tau_0)$. Applying the FCLT to partial sums of the i.i.d. sequence of ϵ_t gives

$$\frac{\sqrt{T}}{k}\sum_{t=1}^{k} \epsilon_t = \frac{T}{k}\frac{1}{\sqrt{T}}\sum_{t=1}^{k} \epsilon_t \Rightarrow \frac{\sigma}{\tau_0 + u}B_1\left(\tau_0 + u\right),$$

and

$$\begin{aligned}
\frac{\sqrt{T}}{T-k}\sum_{t=k+1}^{T} \epsilon_t = \frac{T}{T-k}\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T} \epsilon_t \quad &\Rightarrow \quad \frac{\sigma}{1-\tau_0-u}\left[B_1(1) - B_1\left(\tau_0 + u\right)\right] \\
&\equiv \quad \frac{\sigma}{1-\tau_0-u}B_2\left(1 - \tau_0 - u\right),
\end{aligned}$$

where $B_1(s)$ is a standard Brownian motion and be independent of $B_2(1-s) \equiv B_1(1)-B_1(s)$ whenever s is fixed. Consequently, for $k \leq k_0$,

$$\begin{aligned}
& \left[\sqrt{T} V_k\left(Z_t\right)\right]^2 \\
= \quad & \frac{k(T-k)}{T^2}\left[\sqrt{T}\left(\overline{Z}_k-\overline{Z}_k^*\right)\right]^2 \\
= \quad & \frac{k(T-k)}{T^2}\left(\frac{\sqrt{T}}{k} \sum_{t=1}^k \epsilon_t-\frac{\sqrt{T}}{T-k} \sum_{t=k+1}^T \epsilon_t-\frac{T-k_0}{T-k} \frac{\delta^*}{\varepsilon}\right)^2 \\
\Rightarrow \quad & \left(\frac{\sigma \sqrt{1-\tau_0-u}}{\sqrt{\tau_0+u}} B_1\left(\tau_0+u\right)-\frac{\sigma \sqrt{\tau_0+u}}{\sqrt{1-\tau_0-u}} B_2\left(1-\tau_0-u\right)-\frac{\left(1-\tau_0\right) \sqrt{\tau_0+u}}{\sqrt{1-\tau_0-u}} \frac{\delta^*}{\varepsilon}\right)^2 \\
\stackrel{d}{=} \quad & \left(\sigma B_1\left(1-\tau_0-u\right)-\sigma B_2\left(\tau_0+u\right)-\frac{\left(1-\tau_0\right) \sqrt{\tau_0+u}}{\sqrt{1-\tau_0-u}} \frac{\delta^*}{\varepsilon}\right)^2 .
\end{aligned}$$

Similarly, for $k > k_0$,

$$\left[\sqrt{T} V_k\left(Z_t\right)\right]^2 \Rightarrow\left(\sigma B_1\left(1-\tau_0-u\right)-\sigma B_2\left(\tau_0+u\right)-\frac{\tau_0 \sqrt{1-\tau_0-u}}{\sqrt{\tau_0+u}} \frac{\delta^*}{\varepsilon}\right)^2 .$$

Therefore, with $\widetilde{B}(\cdot)$ defined as in Part (a) of Theorem 4.2, we have,

$$T\left(\frac{\delta^*}{\varepsilon} \sqrt{h}\right)^2\left(\widehat{\tau}_{LS,T}-\tau_0\right) \xrightarrow{d}\left(\frac{\delta^*}{\varepsilon}\right)^2 \arg \max_{u \in\left(-\tau_0, 1-\tau_0\right)}\left[\widetilde{B}(u)\right]^2,$$

which leads to the result in Part (a) of Theorem 4.2 immediately.

(b) We first prove that, when $\varepsilon \rightarrow 0$, $\widehat{\tau}_{LS,T} \xrightarrow{p} \tau_0$. Let

$$V_k\left(Z_t\right)=\sqrt{\frac{k(T-k)}{T^2}}\left(\overline{Z}_k^*-\overline{Z}_k\right)=\sqrt{\frac{k(T-k)}{T^2}}\left(\frac{1}{T-k} \sum_{t=k+1}^T Z_t-\frac{1}{k} \sum_{t=1}^k Z_t\right) .$$

In the following we only consider the case $k \leq k_0$ because of the symmetry. We assume without loss of generality that $\delta^*/\varepsilon>0$ (otherwise consider the series $-Z_t$). We then have

$$\begin{aligned}
E\left[V_k\left(Z_t\right)\right] & =\sqrt{\tau(1-\tau)}\left(\frac{T-k_0}{T-k}\left(\mu+\frac{\delta^*}{\varepsilon}\right) \sqrt{h}+\frac{k_0-k}{T-k} \mu \sqrt{h}-\mu \sqrt{h}\right) \\
& =\sqrt{\tau(1-\tau)} \frac{T-k_0}{T-k} \frac{\delta^*}{\varepsilon} \sqrt{h}=\sqrt{\tau(1-\tau)} \frac{1-\tau_0}{1-\tau} \frac{\delta^*}{\varepsilon} \sqrt{h}>0,
\end{aligned}$$

where $\tau = k/T$. Hence,

$$\begin{aligned}
& E\left[V_{k_0}\left(Z_t\right)\right] - E\left[V_k\left(Z_t\right)\right] \\
= \; & \sqrt{\tau_0\left(1-\tau_0\right)}\frac{\delta^*}{\varepsilon}\sqrt{h} - \sqrt{\tau\left(1-\tau\right)}\frac{1-\tau_0}{1-\tau}\frac{\delta^*}{\varepsilon}\sqrt{h} \\
= \; & \left(1-\tau_0\right)\frac{\delta^*}{\varepsilon}\sqrt{h}\left(\frac{\sqrt{\tau_0}}{\sqrt{1-\tau_0}} - \frac{\sqrt{\tau}}{\sqrt{1-\tau}}\right) \\
= \; & \left(1-\tau_0\right)\frac{\delta^*}{\varepsilon}\sqrt{h}\left(\frac{\tau_0}{1-\tau_0} - \frac{\tau}{1-\tau}\right)\left(\frac{\sqrt{\tau_0}}{\sqrt{1-\tau_0}} + \frac{\sqrt{\tau}}{\sqrt{1-\tau}}\right)^{-1} \\
= \; & \frac{\tau_0-\tau}{1-\tau}\frac{\delta^*}{\varepsilon}\sqrt{h}\left(\frac{\sqrt{\tau_0}}{\sqrt{1-\tau_0}} + \frac{\sqrt{\tau}}{\sqrt{1-\tau}}\right)^{-1} \\
\geq \; & \left|\tau-\tau_0\right|\frac{\delta^*}{\varepsilon}\sqrt{h}\left(2\frac{\sqrt{\tau_0}}{\sqrt{1-\tau_0}}\right)^{-1},
\end{aligned}$$

where the last inequality comes from the fact that $1-\tau < 1$, and $\tau/\left(1-\tau\right)$ is an increasing function over the interval of $(0, \tau_0)$. Note that

$$\begin{aligned}
& \left|V_k\left(Z_t\right)\right| - \left|V_{k_0}\left(Z_t\right)\right| \\
= \; & \left|V_k\left(Z_t\right) - E\left[V_k\left(Z_t\right)\right] + E\left[V_k\left(Z_t\right)\right]\right| - \left|V_{k_0}\left(Z_t\right) - E\left[V_{k_0}\left(Z_t\right)\right] + E\left[V_{k_0}\left(Z_t\right)\right]\right| \\
\leq \; & \left|V_k\left(Z_t\right) - E\left[V_k\left(Z_t\right)\right]\right| + \left|E\left[V_k\left(Z_t\right)\right]\right| - \left\{\left|V_{k_0}\left(Z_t\right) - E\left[V_{k_0}\left(Z_t\right)\right]\right| - \left|E\left[V_{k_0}\left(Z_t\right)\right]\right|\right\} \\
= \; & \left|V_k\left(Z_t\right) - E\left[V_k\left(Z_t\right)\right]\right| - \left|V_{k_0}\left(Z_t\right) - E\left[V_{k_0}\left(Z_t\right)\right]\right| + E\left[V_k\left(Z_t\right)\right] - E\left[V_{k_0}\left(Z_t\right)\right].
\end{aligned}$$

We then have

$$\begin{aligned}
& \left|\widehat{\tau}_{LS,T}-\tau_0\right|\frac{\delta^*}{\varepsilon}\sqrt{h}\left(2\frac{\sqrt{\tau_0}}{\sqrt{1-\tau_0}}\right)^{-1} \\
\leq \; & \left|V_{\widehat{k}_{LS,T}}\left(Z_t\right) - E\left[V_{\widehat{k}_{LS,T}}\left(Z_t\right)\right]\right| - \left|V_{k_0}\left(Z_t\right) - E\left[V_{k_0}\left(Z_t\right)\right]\right| - \left\{\left|V_{\widehat{k}_{LS,T}}\left(Z_t\right)\right| - \left|V_{k_0}\left(Z_t\right)\right|\right\} \\
\leq \; & \left|V_{\widehat{k}_{LS,T}}\left(Z_t\right) - E\left[V_{\widehat{k}_{LS,T}}\left(Z_t\right)\right]\right| - \left|V_{k_0}\left(Z_t\right) - E\left[V_{k_0}\left(Z_t\right)\right]\right| = O_p\left(1/\sqrt{T}\right),
\end{aligned}$$

where the second inequality is due to $\widehat{k}_{LS,T} = \arg\max\left\{\left[V_k\left(Z_t\right)\right]^2\right\}$, and the third equality comes from the fact that for any $1 \leq k < T$,

$$\begin{aligned}
V_k\left(Z_t\right) - E\left[V_k\left(Z_t\right)\right] &= \sqrt{\frac{k\left(T-k\right)}{T^2}}\left(\frac{1}{T-k}\sum_{t=k+1}^{T}\epsilon_t - \frac{1}{k}\sum_{t=1}^{k}\epsilon_t\right) \\
&= \frac{1}{\sqrt{T}}\left(\sqrt{\frac{k}{T}}\frac{1}{\sqrt{T-k}}\sum_{t=k+1}^{T}\epsilon_t - \sqrt{\frac{T-k}{T}}\frac{1}{\sqrt{k}}\sum_{t=1}^{k}\epsilon_t\right) \\
&= \frac{1}{\sqrt{T}}O_p\left(1\right).
\end{aligned}$$

Therefore, when $\varepsilon \to 0$,

$$|\widehat{\tau}_{LS,T} - \tau_0| \leq 2\frac{\sqrt{\tau_0}}{\sqrt{1-\tau_0}} \left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-1} O_p\left(\frac{1}{\sqrt{T}}\right) = 2\frac{\sqrt{\tau_0}}{\sqrt{1-\tau_0}} \frac{\varepsilon}{\delta^*} O_p(1) \to 0.$$

Then, following the procedure in the proof of Proposition 3 in Bai (1994), it can be proved that $\widehat{\tau}_{LS,T} - \tau_0 = O_p\left(\sqrt{T}\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}$, when $h \to 0$ and $\varepsilon \to 0$ simultaneously with the condition of $(\delta^*/\varepsilon)\sqrt{h} \to 0$. Finally, following the procedure in the proof of Theorem 1 in Bai (1994), the limiting distribution in Part (b) of Theorem 4.2 is obtained. The details of these two steps are omitted for simplicity.

References

[1] Andrews, D. W. K., 1993, Exactly Median-unbiased Estimation of First Order Autoregressive/unit Root Models. Econometrica 61, 139–166.

[2] Arvanitis, S. and Demos, A., 2014, On the Validity of Edgeworth Expansions and Moment Approximations for Three Indirect Inference Estimators, Working Paper, Athens University of Economics and Business.

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

[4] Bai, J., 1995. Least absolute deviation estimation of a shift. Econometric Theory 11, 403-436.

[5] Bai, J., 1997a. Estimating multiple breaks one at a time. Econometric Theory 13, 315-352.

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

[7] Bai, J., and Perron, P., 1998. Estimating and testing linear models with multiple structural breaks. Econometrica 66, 47-78.

[8] Bai, J., Lumsdaine, R. L., and Stock, J. H., 1998. Testing for and dating common breaks in multivariate time series. The Review of Economic Studies 65, 395-432.

[9] Barndorff-Nielsen, O.E., and N. Shephard, 2004. Econometric analysis of realised covariation: High frequency based covariance, regression and correlation in financial economics. Econometrica 72, 885-925

[10] Bhattacharya, P. K., 1987. Maximum likelihood estimation of a change-point in the distribution of independent random variables: General multiparameter case. Journal of Multivariate Analysis 23, 183-208.

[11] Efron, B., 1982, *The Jackknife, the Bootstrap and Other Resampling Method*, SIAM, Philadephia.

[12] Forneron, J. J., and Ng, S., 2015. The ABC of Simulation Estimation with Auxiliary Statistics, Working Paper, Columbia University.

[13] Gallant, A.R., and Tauchen, G., 1996. Which moments to match? Econometric Theory 12, 657–681.

[14] Gouriéroux, C., Monfort, A., and Renault, E., 1993. Indirect estimation. Journal of Applied Econometrics 8, 85-118.

[15] Gouriéroux, C., Renault, E., Touzi, N., 2000. Calibration by simulation for small sample bias correction. In: Mariano, R.S., Schuermann, T., Weeks, M. (Eds.), Simulation-Based Inference in Econometrics: Methods and Applications. Cambridge University Press, pp. 328–358.

[16] Gouriéroux, C., Phillips, P. C. B., and Yu, J., 2010. Indirect estimation for dynamic panel models. Journal of Econometrics 157, 68-77.

[17] Hall, P. 1992, *The Bootstrap and Edgeworth Expansion*, Springer-Verlag.

[18] Hawkins, D. L., Gallant, A. R., and Fuller, W., 1986. A simple least squares method for estimating a change in mean. Communications in Statistics-Simulation and Computation 15, 523-530.

[19] Hájek, J. and Rényi, A., 1955. Generalization of an inequality of Kolmogorov. Acta Mathematica Hungarica, 6(3-4), pp.281-283.

[20] Hinkley, D. V., 1970. Inference about the change-point in a sequence of random variables. Biometrika 57, 1-17.

[21] Ibragimov, I. A., and Has'minskii, R. Z., 1981. Statistical estimation. Springer.

[22] Kendall, M. G., 1954. Note on bias in the estimation of autocorrelation. Biometrika 41, 403-404.

[23] Kim, J. and Pollard, D., 1990. Cube root asymptotics. The Annals of Statistics 18,191-219.

[24] MacKinnon, J. G., and Smith Jr, A. A., 1998. Approximate bias correction in econometrics. Journal of Econometrics 85, 205-230.

[25] Perron, P., 1989. The great crash, the oil price shock, and the unit root hypothesis. Econometrica 57, 1361-1401.

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

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

[28] Phillips, P. C. B., 2012. Folklore theorems, implicit maps, and indirect estimation. Econometrica 80, 425-454.

[29] Phillips, P. C. B., and Yu, J., 2009a. Simulation-based estimation of contingent-claims prices. Review of Financial Studies 22, 3669-3705.

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

[31] Phillips, P. C. B., and Yu, J., 2009c. Maximum likelihood and Gaussian estimation of continuous time models in finance. In: Handbook of Financial Time Series, 707-742.

[32] Smith, A. A., 1993. Estimating nonlinear time series models using simulated vector autoregressions. Journal of Applied Econometrics 8, 63-84.

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

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

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

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