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

January 28, 2016

Abstract

Based on the Girsanov theorem, this paper first 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 seriously biased. These two properties are also found in the finite sample distribution of the least squares estimator of structural break point in the discrete time model. The paper then builds a continuous time approximation to the discrete time model and develops an in-fill asymptotic theory for the least squares estimator. The obtained in-fill asymptotic distribution is asymmetric and tri-modal and delivers good approximations to the finite sample distribution. In order to reduce the bias in the estimation of both the continuous time model and the discrete time model, a simulation-based method based on the indirect estimation approach is proposed. Monte Carlo studies show that the indirect estimation method achieves substantial bias reductions. However, since the binding function has a slope less than one, the variance of the indirect estimator is larger than that of the original estimator.

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 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 2 grant with the MOE's official grant number MOE2011-T2-2-096. 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. Bhattacharya (1994) provides a review of the statistics literature on the problem while Perron (2006) gives a review of the econometrics literature on the same problem. There are also several books devoted to this topic of research, including Csörgő and Horváth (1997), Chen and Gupta (2011). Both strands of the literature have addressed the problem in many aspects, from estimation, testing to computation, from frequentist's methods to Bayesian methods, from one structural break to multiple structural breaks, from univariate settings to multivariate settings. In addition to its statistical implications, the economic and financial implications of structural break problem have also been extensively studied; see, for example, Hansen (2001) and Andreou and Ghysels (2009) for excellent reviews.

The literature has developed 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). The asymptotic theory has been obtained by assuming that the time span of data goes 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 $(-\infty, +\infty)$, and has an analytical expression for the probability density function (pdf). It is symmetric with the origin 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. Is the lack of study on the quality of approximation of the asymptotic distribution to the finite sample distribution due to the good performance of the asymptotic distribution? Or is the lack of attention due to the difficulty in studying the finite sample theory? Is there any substantial bias in the commonly used estimators of the structural break point in finite sample?

This paper systematically investigates the exact distributional properties and the bias problem in the estimation of structural break points. To the best of our knowledge, our study is the first systematic analysis of the exact distribution theory in the literature. We first develop the exact distribution of the maximum likelihood (ML) estimator of structural break point in a continuous time model, assuming that a continuous record over a finite time span is available. We document the asymmetry and the trimodality in the exact distribution. As a result, the exact distribution suggests that the ML estimator is biased whenever the true break point is not in the middle of the sample. Aiming to retain the properties of asymmetry and trimodality in the discrete time model, we study the exact discretization of a continuous time model with an unknown structural break point and develop an in-fill asymptotic theory for the least squares (LS) estimator of break point. To reduce the bias in the estimation of break point both in the continuous time model and in the discrete time model, an indirect estimation procedure is proposed.

Our study makes several contributions to the literature. First, we develop a novel approach to obtain the exact distribution of the ML estimator of break point. Since the likelihood function and the sum of squared residuals are not differentiable with respect to the break point in discrete time models, the traditional approaches to obtain the exact distribution are not feasible. By using the Girsanov theorem, we obtain the likelihood function in a continuous time model with a structural break and then the exact distribution of the ML estimator.

Second, we show 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 seriously biased. It is also found that the further the fractional structural break point away from 50%, the larger the bias. When the fractional structural break point is smaller (larger) than 50%, the bias is positive (negative).

Third, we find that the properties of asymmetry and trimodality are shared by the finite sample distribution of the LS estimator of break point in the discrete time model, suggesting substantial bias in the LS estimation especially when the signal-to-noise ratio is not very large. To better approximate the finite sample distribution in the discrete time model, 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 in-fill asymptotic distribution retains the properties of asymmetry and trimodality found in the finite sample distribution, and, hence, provides better approximations than the long-span asymptotic distribution.

Finally, we propose to do bias reduction by using the indirect estimation procedure. 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), Nickell (1981), Tang and Chen (2009), Yu (2012) for various types of autoregressive models. However, it is difficult to use the analytical approach in this context as the bias formula is difficult to obtain analytically. The primary advantage of the indirect estimation procedure 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 the indirect estimation procedure, without knowing the analytical form of the bias, achieves substantial bias reduction. Since it is easy to simulate the model and estimate the break point parameter, the indirect estimation is a convenient method for reducing the bias in the estimation of structural break points. However, since the binding function has a slope less than one, the variance of the indirect estimator is larger than that of the original estimator.

Our in-fill asymptotic treatment has a spirit similar to Phillips (1987), Perron (1991), and Barndorff-Nielsen and Shephard (2004). The comparison of the in-fill asymptotic distribution and the long-span asymptotic distribution in the autoregressive process was recently considered in Yu (2014) and Zhou and Yu (2015). It was also found that the in-fill asymptotic distribution provides better approximations to the finite sample distribution than the long-span asymptotic distribution when the process is highly persistent.

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 the 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 estimators under different settings. The indirect estimation procedure and its applications in the continuous time model and the discrete time model with structural break are introduced in Section 5. In Section 6, we provide simulation results and compare the finite sample performance of the indirect estimation method with that of the traditional estimation methods and 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 in statistics include Hinkley (1970), Ibragimov and Has'minskii (1981), Hawkins et al. (1986), and Yao (1987). In econometrics, Jushan Bai and Pierre Perron have made many contributions to the literature through their individual works as well as their collaborative works; see, for example, Perron (1989), Bai (1994, 1995, 1997a, 1997b, 2010), 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

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

where $t = 1, \ldots, T$ with T being the number of observations of Y_t, ϵ_t is a sequence of independent and identically distributed (i.i.d.) random variables with $E(\epsilon_t) = 0$ and $Var(\epsilon_t) = \sigma^2$, and 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 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 $\widehat{\tau}_{ML,T} = \widehat{k}_{ML,T}/T$. Hinkley (1970) showed that $\widehat{k}_{ML,T} - k_0$ converges in distribution as the sample sizes before and after the break point tend to infinity. He also pointed out that the distribution of $\widehat{k}_{ML,\infty} - k_0$, where $\widehat{k}_{ML,\infty}$ denotes $\widehat{k}_{ML,T}$ when $T \to \infty$, has no closed-form expression, and suggested a numerical method to compute the distribution. However, the suggested numerical scheme is difficult to use for small δ since the distribution becomes rather dispersive when δ is small. This difficulty motivates Yao (1987) to develop a limit theory as $\delta \to 0$.

Letting $\delta \to 0$, Yao (1987) derived a long-span limiting distribution as

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

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. So the limiting distribution is the location of the extremum of **a** two-sided Brownian motion with triangular drift over the interval $(-\infty, \infty)$. Given that $I(\mu)$ depends on the error distribution, there is no invariance principle in the limit theory. Yao (1987) also derived the pdf of the long-span limiting distribution as

$$g(x) = 1.5 e^{|x|} \Phi\left(-1.5|x|^{0.5}\right) - 0.5 \Phi\left(-0.5|x|^{0.5}\right),$$

and the cumulative distribution function (cdf) as

$$G(x) = 1 + \sqrt{x/2\pi} e^{-x/8} - (x + 5) \Phi\left(-0.5\sqrt{x}\right)/2 + 1.5 e^{x} \Phi\left(-1.5\sqrt{x}\right) \quad \text{for } x > 0,$$

and $G(x) = 1 - G(-x)$ for $x \leq 0$, where $\Phi(x)$ is the cdf of 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 unknown break point. The LS estimator of the break point k takes the form of

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

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 $\delta_T \to 0$ 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 $I(\mu)$ 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, in Bai (1994) no assumption is made about the error distribution, and, hence, an invariance principle applies.

When the error term in model (1) becomes a weakly stationary process with a long-run variance $\left[a(1) \right]^2$, Bai (1994) showed that

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

A continuous time model with a structural break in the drift function was studied in Ibragimov and Has'minskii (1981, hereafter IH). The model takes the form of

$$dX(t) = \frac{1}{\varepsilon} S(t - \tau_0) dt + dB(t), \tag{6}$$

where $t \in [0, 1]$, $S(t - \tau_0)$ is a non-stochastic drift term with discontinuity at time τ_0 and $\lim_{x \to 0+} S(x) - \lim_{x \to 0-} S(x) = \delta^*$, ε is a small parameter and $B(t)$ represents a

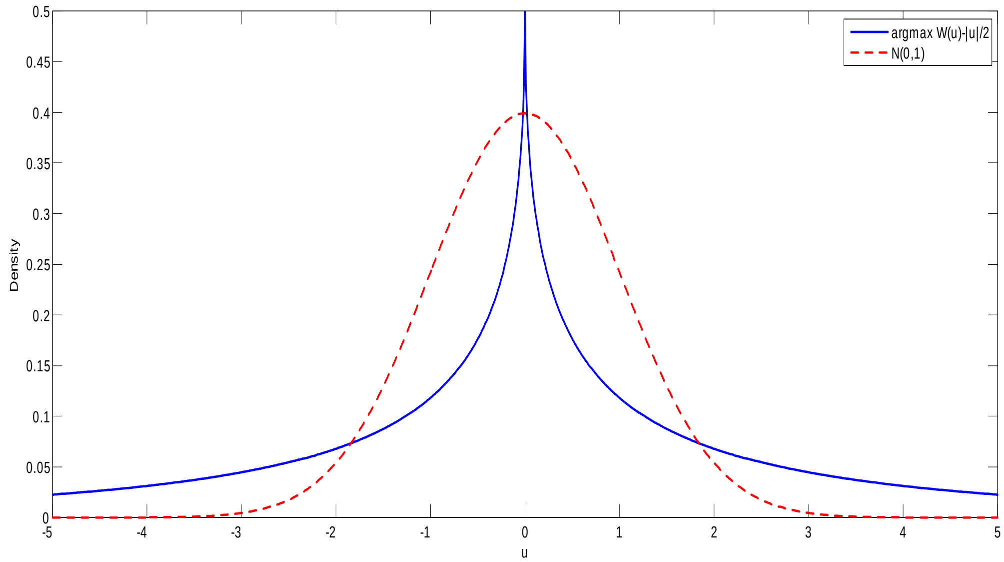

Figure 1: The pdfs of $\arg \max_{u \in (-\infty, \infty)} \left\{ W(u) - \frac{1}{2}|u| \right\}$ and a standard normal distribution.

standard Brownian motion. An important feature in (6) is that the break size in the drift function is δ^*/ε, which goes to infinity as $\varepsilon \to 0$. IH assumed that a continuous record is available. Following the development of the local asymptotic theory of Le Cam (1960), IH examined the behavior of the normalized likelihood ratio in the small neighborhood of the true break point τ_0 such that $\tau = \tau_0 + \varepsilon^2 u$ with $u = O_p(1)$, and showed that as $\varepsilon \to 0$,

$$\left(\frac{\delta^*}{\varepsilon}\right)^2 (\hat{\tau}_{ML} - \tau_0) \xrightarrow{d} \arg \max_{u \in (-\infty, \infty)} \left\{ W(u) - \frac{1}{2}|u| \right\}. \tag{7}$$

The limiting distributions obtained in Yao (1987), Bai (1994) and IH listed in (3), (5) and (7), respectively, are exactly the same, which is the distribution of the location of the extremum of a two-sided Brownian motion with triangular drift over the interval $(-\infty, \infty)$. Figure 1 plots the pdf of it. For the purpose of comparison, the pdf of a standard normal distribution is also plotted. It 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 an unique mode at the origin and is symmetric about it, suggesting that all the estimators studied in the literature have no bias in the limiting distribution no matter what the true value of the structural break point is.

Unfortunately, the asymptotic distribution of the ML estimator and the LS estima-

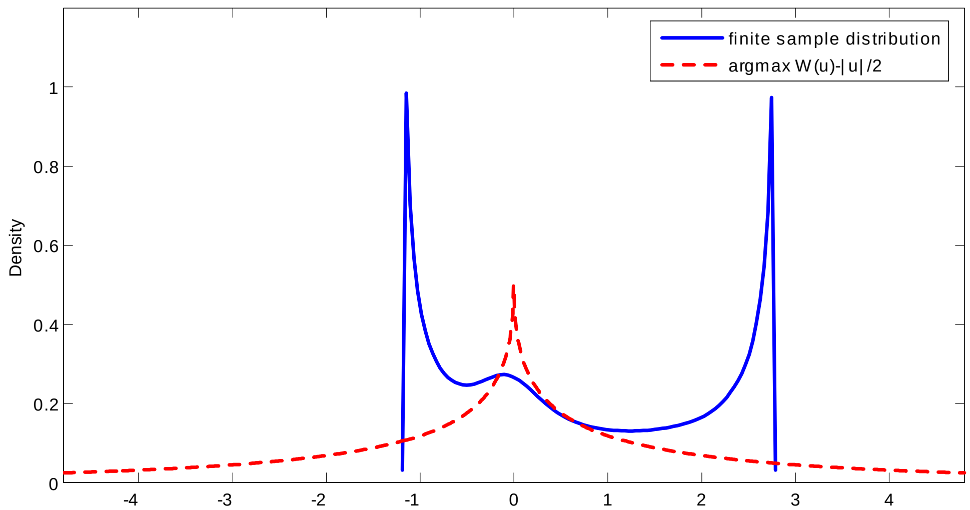

Figure 2: The pdf of the finite sample distribution of $T\left(\frac{\delta_T}{\sigma}\right)^2(\widehat{\tau}_{LS,T}-\tau_0)$ when $T=100$, $\delta_T=0.2$, $\sigma=1$ and $\tau_0=0.3$ in Model (1) and the pdf of $\arg\max_{u\in(-\infty,\infty)}\left\{W(u)-\frac{1}{2}|u|\right\}$.

tor derived 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 limiting distribution in (5) and the finite sample distribution of $T\left(\frac{\delta}{\sigma}\right)^2(\widehat{\tau}_{LS,T}-\tau_0)$ with $\widehat{\tau}_{LS,T}$ defined in (4) 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. There are three striking distinctions between the two distributions. First, the finite sample distribution is asymmetric, whereas the asymptotic distribution is symmetric. Second, the finite sample distribution displays trimodality while the asymptotic distribution has a unique mode. Third, the finite sample distribution indicates that the LS estimator $\widehat{\tau}_{LS,T}$ is seriously biased. The bias is 0.1704, which is about 57% of the true value. In contrast, there is no bias suggested by the asymptotic distribution. It is this inadequacy of the asymptotic distribution for approximating the finite sample distribution that motivates us to develop an alternative distribution theory for the estimation of 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

$$dX(t) = \widetilde{S}(t-\tau_0)dt + \sigma dB(t), \tag{8}$$

where $t \in [0,1]$,

$$\widetilde{S}(t-\tau_0) = \begin{cases} \mu & if\ t \le \tau_0 \\ \mu + \delta^*/\varepsilon & if\ t > \tau_0 \end{cases},$$

μ, δ^*, ε and τ_0 are all constants, 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 make sure that one break happens during the time interval $(0,1)$. δ^*/ε is the break size. The continuous time model is a natural choice for capturing the different amount of sample information before and after the break point. As long as $\tau_0 \neq 1/2$, the amount of sample information contained by observations over the time interval $[0,\tau_0]$ is different from that over the time interval $[\tau_0, 1]$.[1] An alternative representation of Model (8) is

$$dX(t) = \left[\left(\mu + \left(\frac{\delta^*}{\varepsilon}\right) 1_{[t>\tau_0]}\right)\right] dt + \sigma dB(t), \tag{9}$$

where $1_{[t>\tau_0]}$ is an indicator function. We assume that a continuous record is available and all parameters are known except for τ. 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.

Following IH, for any $\tau \in (0,1)$ we obtain the exact log-likelihood ratio of Model (9) via the Girsanov Theorem[2]

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

which leads to the ML estimator of τ as

$$\widehat{\tau}_{ML} = \arg\max_{\tau \in (0,1)} \log\left(\frac{dP_\tau}{dP_{\tau_0}}\right). \tag{10}$$

[1] In IH's continuous time model stated in (6), a symmetric distribution as (7) was obtained because they assumed that the break size goes to infinity and applied the local asymptotic approach developed in Le Cam (1960).

[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

$$W(u) = \begin{cases} W_1(-u) = B(\tau_0) - B(\tau_0 - (-u)) & \text{if } u \leq 0 \\ W_2(u) = B(\tau_0) - B(\tau_0 + u) & \text{if } u > 0 \end{cases}, \tag{11}$$

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.

It can be seen that, when $\mu = 0$, Model (8) becomes the one studied in IH with the signal-to-noise ratio δ^*/ε being replaced by $\delta^*/(\sigma\varepsilon)$. Therefore, when $\varepsilon \to 0$, the asymptotic distribution of $\widehat{\tau}_{ML}$ defined in (10) is the same as the one given in IH with δ^*/ε being replaced by $\delta^*/(\sigma\varepsilon)$. Since $\widehat{\tau}_{ML}$ is independent of μ, the asymptotic distribution of $\widehat{\tau}_{ML}$ applies to any value of μ. However, when ε is fixed, we will show that the distribution of $\widehat{\tau}_{ML}$ is asymmetric when $\tau_0 \neq 1/2$ and has trimodality. We report these results in the following theorem.

Theorem 3.1 *Consider Model (8) with a continuous record being available. For the ML estimator $\widehat{\tau}_{ML}$ defined in (10),*
(a) when ε is a constant, we have the exact distribution as

$$\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2 (\widehat{\tau}_{ML} - \tau_0) \stackrel{d}{=} \arg \max_{u \in \left(-\tau_0\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2, (1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2\right)} \left\{W(u) - \frac{|u|}{2}\right\}; \tag{12}$$

(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 (11), and $\stackrel{d}{=}$ denotes equivalence in distribution.

The distribution in (12) is exact. It is different from the limiting distribution developed in the literature as given 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 $(-\infty, \infty)$. Since the interval is symmetric about zero, the limiting distribution is symmetric. However, the exact distribution in (12) 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 τ_0 is 1/2, which means that the true break point is exactly in the middle of the sample, the interval $\left(-\tau_0\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2, (1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2\right)$ becomes

$\left(-\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2/2, \left(\frac{\delta^*}{\sigma\varepsilon}\right)^2/2\right)$ which is symmetric about the origin. In this case the exact distribution is symmetric. However, if τ_0 is not $1/2$ (either smaller or larger than $1/2$), 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 (12) 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. Second, the interval over which to find the extremum of $W(u) - \frac{1}{2}|u|$ is unbounded for the limiting distribution in (5). Whereas, the interval is always bounded for the exact distribution. Such a difference has an implication for the modality of the distribution, as explained below.

Because of this change in the interval to locate the extremum, we cannot obtain the pdf or cdf of the exact distribution in closed-form. As a result, we obtain the pdf by simulations as for the case of the Dickey-Fuller distribution. Figure 3 plots the density of $\widehat{\tau}_{ML} - \tau_0$ when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively) when $\varepsilon = 1$ and the signal-to-noise ratio $(\frac{\delta^*}{\sigma\varepsilon})$ is 1. Figures 4-7 plot the density of $\widehat{\tau}_{ML} - \tau_0$ when the signal-to-noise ratio is 2, 4, 6 and 8, respectively. There are several interesting observations from these plots. First and most importantly, when $\tau_0 = 50\%$, the density of $\widehat{\tau}_{ML} - \tau_0$ is symmetric about the origin, no matter what the signal-to-noise ratio is. As a result, there is no bias in $\widehat{\tau}_{ML}$. However, when τ_0 is not 50%, the density is not symmetric any more. In particular, if τ_0 is less (more) than 50%, the density is positively (negatively) skewed, indicating an upward (downward) bias in $\widehat{\tau}_{ML}$. The smaller the signal-to-noise ratio, the larger the bias. The further τ_0 away from 50%, the larger the bias, a feature that becomes more apparent in our simulation study reported later.[3]

Second, the exact distribution displays trimodality, a feature being more apparent when the signal-to-noise ratio is smaller. 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 $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$. The closer the boundary point to the origin, the bigger the mode at the boundary point.

That the origin is a mode is well expected because the drift term in $W(u) - \frac{1}{2}|u|$ is $-\frac{1}{2}|u|$ which is negative and the random term is $W(u) \sim (0, |u|) = O_p(\sqrt{|u|})$. When $|u|$ is large, the negative drift term dominates the random term in $W(u) - \frac{1}{2}|u|$. As a result, when there is no bound in the interval, the probability for $W(u) - \frac{1}{2}|u|$ to reach the maximum at a large value of $|u|$ is small, and decreases as $|u|$ becomes larger. However, because of the randomness in $W(u)$, it is possible for $W(u) - \frac{1}{2}|u|$ to reach the maximum at a large value of $|u|$. This also gives a reasonable explanation for the

[3]Detailed results about the bias will be reported in Section 6.

shape of the long-span limiting distribution in (5) as apparent in Figure 1.

When $\frac{\delta^*}{\sigma\varepsilon}$ gets smaller, $\frac{1}{2}|u|$ only takes smaller values at the boundary points. This mean that the negative drift become less dominant, and, hence, it is more likely for $W(u) - \frac{1}{2}|u|$ to reach the maximum at the neighborhoods of the two boundary points.

To explain why the other two modes are 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 this small left neighborhood, the reason why it is more likely for $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$ to be the arg max of $W(u) - \frac{1}{2}|u|$ than any interior point is that, for $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$ $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$ to be the arg max, the value of $W(u) - \frac{1}{2}|u|$ at $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$ has to 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. Similar arguments apply to the other boundary point, $-\tau_0\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$.

The arguments in the above two paragraphs help explain why the two modes at the boundary points become more pronounced when $\frac{\delta^*}{\sigma\varepsilon}$ decreases. 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 the highest mode when the signal-to-noise ratio is very small, as apparent in Figures 3.

To explain why the mode on the left boundary point is larger (smaller) than that on the right boundary point when τ_0 is less (greater) than 50%, note that $\frac{1}{2}|u|$ takes a smaller (larger) value at $-\tau_0\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$ $\left((1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2\right)$. Hence, it is more (less) likely for $W(u) - \frac{1}{2}|u|$ to reach the maximum in the neighborhood of $-\tau_0\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$ than that of $(1-\tau_0)\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2$.

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 build a continuous-time approximation to the discrete time structural break model widely studied in the literature, aiming to obtain a better approximation to the finite sample distribution of the break point estimation in the discrete time model.

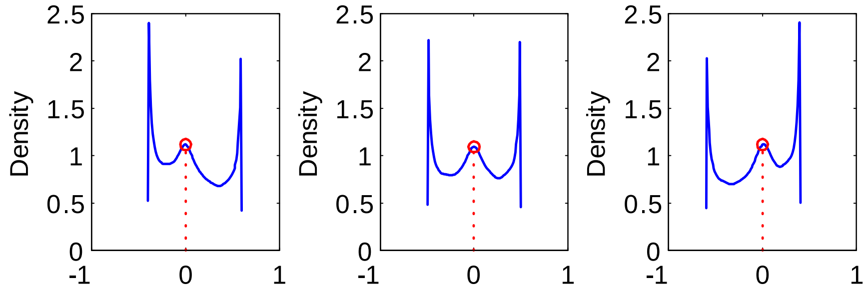

Figure 3: The density of $\widehat{\tau}_{ML} - \tau_0$ given in Equation (12) when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively) and the signal-to-noise ratio $(\frac{\delta^*}{\sigma\epsilon})$ is 1.

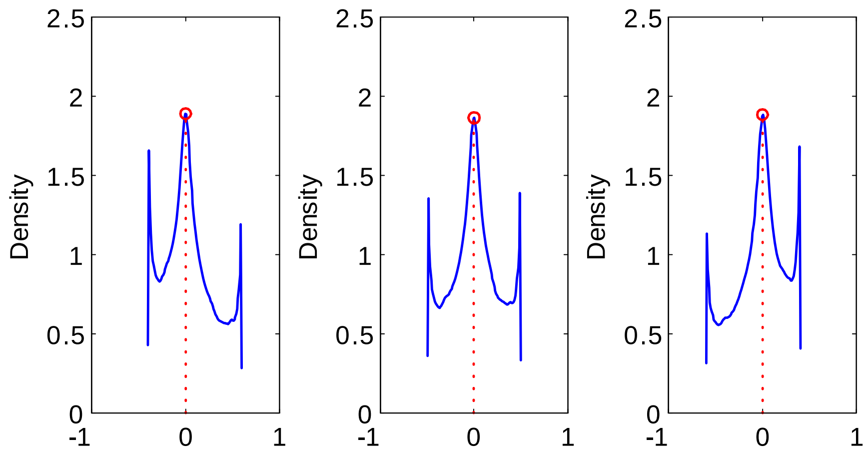

Figure 4: The density of $\widehat{\tau}_{ML} - \tau_0$ given in Equation (12) when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively) and the signal-to-noise ratio $\frac{\delta^*}{\sigma\epsilon}$ is 2.

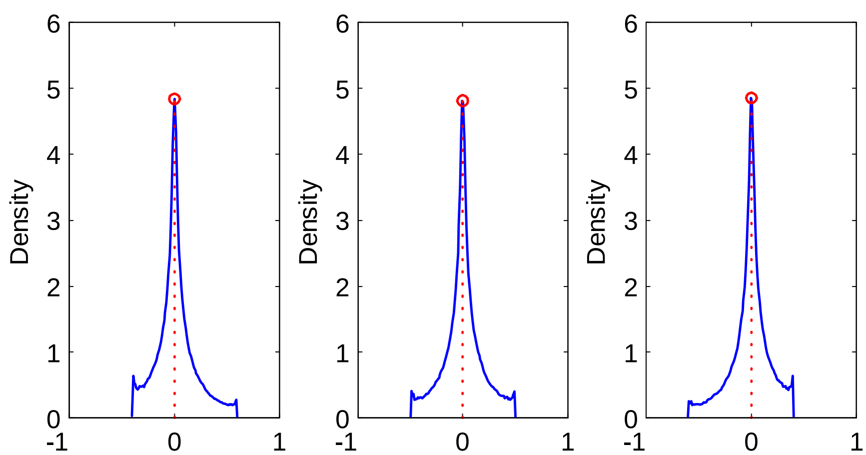

Figure 5: The density of $\widehat{\tau}_{ML} - \tau_0$ given in Equation (12) when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively) and the signal-to-noise ratio $\frac{\delta^*}{\sigma\epsilon}$ is 4.

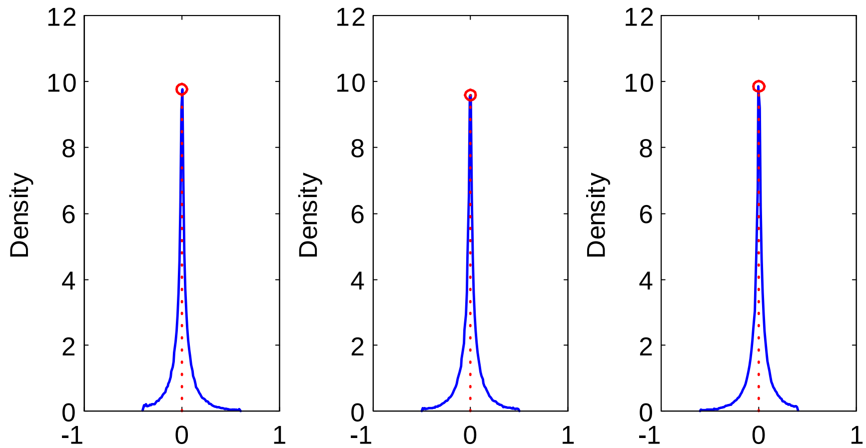

Figure 6: The density of $\widehat{\tau}_{ML} - \tau_0$ given in Equation (12) when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively) and the signal-to-noise ratio $\frac{\delta^*}{\sigma\epsilon}$ is 6.

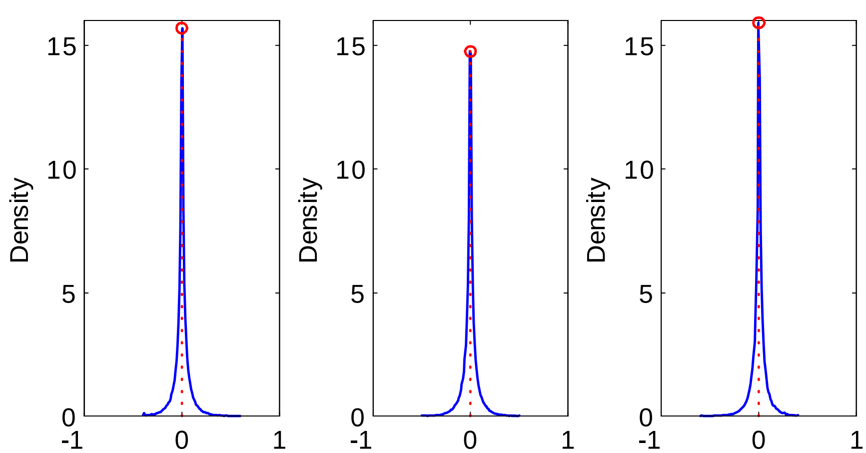

Figure 7: The density of $\widehat{\tau}_{ML} - \tau_0$ given in Equation (12) when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively) and the signal-to-noise ratio $\frac{\delta^*}{\sigma\epsilon}$ is 8.

In particular, we approximate the discrete time model studied in Hinkley (1970), Yao (1987) and Bai (1994) by using the exact discretization of a continuous time model. Based on the continuous time approximation, we then develop the in-fill asymptotic theory for the LS estimator of the break point under two settings.

Assume observations are 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. The in-fill asymptotics correspond to the case when $h \to 0$. It is assumed that Th is fixed, say at 1. Clearly, if $h \to 0$, the sample size $T \to \infty$. In the limit of $h \to 0$, a continuous record is available. For simplicity, we let τ_0/h be an integer, denoted by k_0. The notation X_{th} is used to represent a discrete time process. The exact discretization of the continuous time process $X(\cdot)$ defined in (9) takes the form of

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

where $\epsilon_{th} \sim \text{i.i.d.}N(0, \sigma^2)$. As ϵ_{th} is independent of h, we simply write it as ϵ_t. Letting $Z_t = \left(X_{th} - X_{(t-1)h}\right)/\sqrt{h}$, we have

$$Z_t = \begin{cases} \mu\sqrt{h} + \epsilon_t & \text{if } \ t \leq k_0, \\ (\mu + \delta^*/\varepsilon)\sqrt{h} + \epsilon_t & \text{if } \ t > k_0. \end{cases} \tag{13}$$

Whenever h is fixed, the discrete time model in Equation (13) is the same as the one studied in Hinkley (1970), Yao (1987) and Bai (1994) given in Equation (1) with ϵ_t being normally distributed and the shift in mean being $\delta = (\delta^*/\varepsilon)\sqrt{h}$.

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 is different from the asymptotic schemes employed in the literature. For example, Bai (1994) allowed the break size shrinks to zero but at a rate slower than $\sqrt{\log T}/\sqrt{T}$ as $T \to \infty$. The slower convergence rate of the break size may explain why the limiting distribution in (5) cannot approximate well the finite sample distribution for the model with a small break, as demonstrated below in simulations.

We now develop the in-fill asymptotic theory for the LS estimator of τ with ε fixed. 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 when $\tau_0 \neq 1/2$, and has trimodality. In both cases, the in-fill asymptotic distribution provides better approximations to the finite sample distribution than the long-span limiting distribution developed in the literature as given in (5). The superiority of the in-fill asymptotic distribution over the long-span asymptotic distribution was recently documented in Yu (2014) and Zhou and Yu (2015) in the context of autoregressive processes.

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 $1/\sqrt{T}$. We show that the in-fill asymptotic distribution is the same as the limiting 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 the break point is defined as

$$\begin{aligned}
\widehat{k}_{LS,T} &= \arg\min_{k=1,\ldots,T-1} \left\{ \sum_{t=1}^{k} \left(Z_t - \mu\sqrt{h} \right)^2 + \sum_{t=k+1}^{T} \left(Z_t - (\mu + \delta^*/\varepsilon)\sqrt{h} \right)^2 \right\} \\
&= \arg\min_{k=1,\ldots,T-1} \left\{ 2(\delta^*/\varepsilon)\sqrt{h} \sum_{t=1}^{k} \left(Z_t - \mu\sqrt{h} \right) - (\delta^*/\varepsilon)^2 hk \right\} \\
&= \arg\max_{k=1,\ldots,T-1} \left\{ -(\delta^*/\varepsilon)\sqrt{h} \sum_{t=1}^{k} \left(Z_t - \mu\sqrt{h} \right) + (\delta^*/\varepsilon)^2 hk/2 \right\}. \qquad (14)
\end{aligned}$$

The corresponding estimator of the fractional break point is $\widehat{\tau}_{LS,T} = \widehat{k}_{LS,T}/T$. When the error distribution in Model (13) is Gaussian, the LS estimators of k and τ are also the ML estimators, as defined in Yao (1987). Comparing to Yao's long-span asymptotic distribution, the in-fill asymptotic distribution given in Part (a) of the following theorem provides an alternative asymptotic approximation to the finite sample distribution of $\widehat{\tau}_{LS,T}$. Part (b) of the following theorem connects our in-fill asymptotics to Yao's long-span asymptotics.

Theorem 4.1 *Consider Model (13) 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 (14). 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

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

Table 1: The bias in finite sample obtained from the simulated data, the bias calculated from the in-fill asymptotic distribution, and the bias calculated from the long-span asymptotic distribution in Yao (1987). The number of replications is set at 100,000.

Case		Bias		
$\frac{\delta^*}{\sigma\varepsilon}$	τ_0	Finite sample	In-fill asymptotics	Long-span asymptotics
2	0.3	0.0909	0.0911	0
2	0.7	-0.0921	-0.0903	0
4	0.3	0.0307	0.0299	0
4	0.7	-0.0305	-0.0302	0
6	0.3	0.0078	0.0073	0
6	0.7	-0.0080	-0.0072	0

where $W(u)$ is the two-sided Brownian motion defined in (11).

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, according to Part (a) of Theorem 4.1, $\widehat{\tau}_{LS,T}$ 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, according to Part (b) of Theorem 4.1, $\widehat{\tau}_{LS,T}$ is consistent but $\widehat{k}_{LS,T} - k_0$ diverges at a rate slower than T.*

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. The proof of Theorem 4.1 can be easily extended to the case where the errors in Model (13) 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 8 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 (13) 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. Table 1 reports the finite sample bias of the LS estimator $\widehat{\tau}_{LS,T}$, the bias

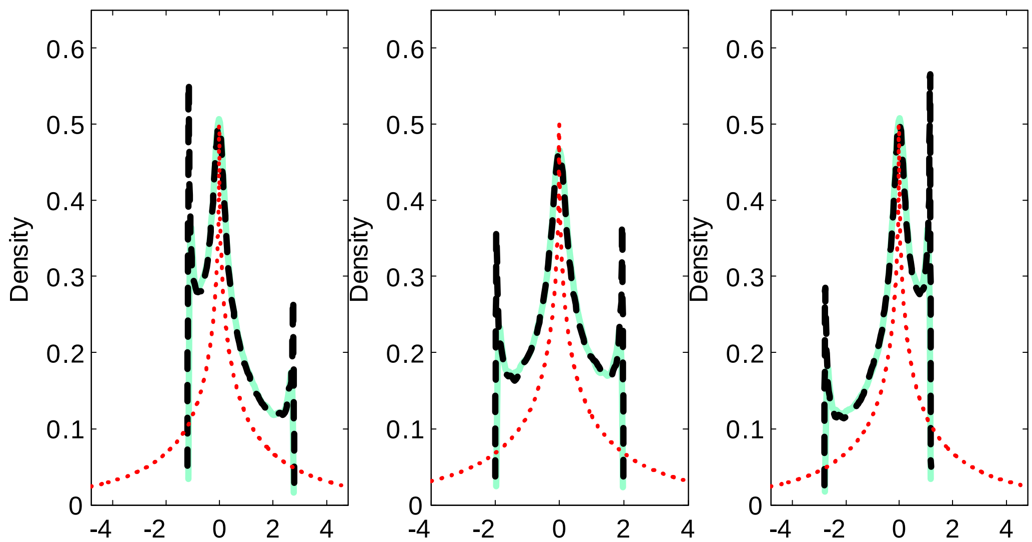

Figure 8: The pdf of $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS,T}-\tau_0)$ when $\tau_0=0.3,0.5,0.7$ (the left, middle and right panel respectively) and $\frac{\delta^*}{\sigma\varepsilon}=2$. The blue solid line is the finite sample distribution when $T=100$; the 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, and the bias implied by the long-span limiting distribution.

Several features are apparent in Figure 8 and 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. Not surprisingly, it provides much better approximations to the finite sample distribution than the long-span asymptotic distribution. Fourth, the in-fill asymptotic distribution captures the finite sample bias very well.

Figures 9-10 plot the finite sample density of $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS,T}-\tau_0)$ when the break size is 0.4 and 0.6, respectively, as well as the corresponding density of the in-fill asymptotic distribution given in Part (a) of Theorem 4.1 and the density of the long-

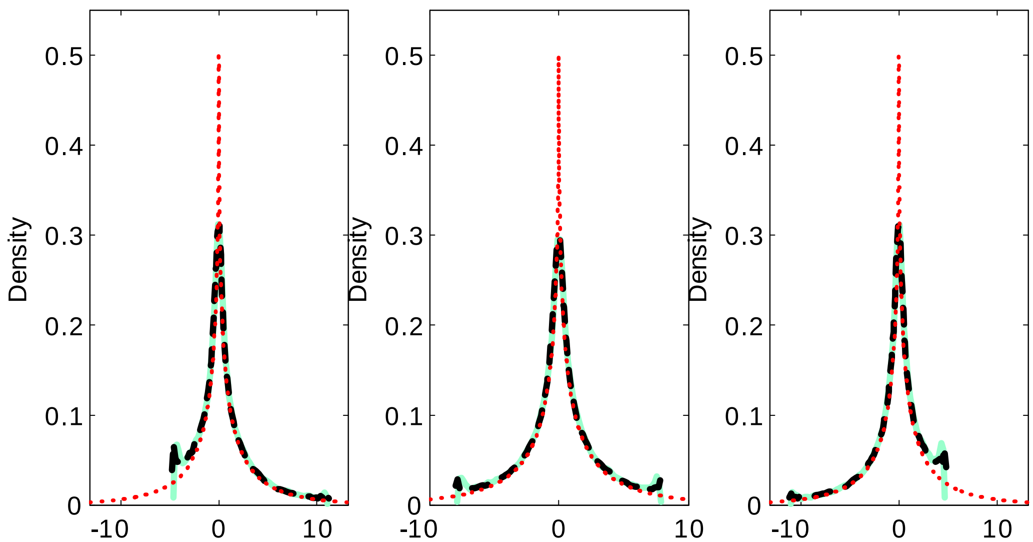

Figure 9: The pdf of $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS,T}-\tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\frac{\delta^*}{\sigma\varepsilon} = 4$. The blue solid line is the finite sample distribution when $T = 100$; the black broken line is the density given in Part (a) of Theorem 4.1; and the red dotted line is the long-span limiting distribution in Yao (1987).

span limiting distribution given in Yao (1987). Qualitatively, similar conclusions can be drawn from Figures 9-10 to those from Figure 8. Comparing Figures 9-10 with Figure 8, we can see that, as the break size increases, the trimodality becomes less pronounced and the degree of asymmetry reduces. As a result, the magnitude of bias becomes smaller. Moreover, as the break size gets larger, the long-span asymptotic distribution given in Yao (1987) can better approximate the finite sample distribution. However, the finite sample distribution is less concentrated around zero and less peaked than the long-span asymptotic distribution. In all cases, the in-fill asymptotic distribution given in Part (a) of Theorem 4.1 always provides better approximations to the finite sample distribution than the long-span asymptotic distribution.

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. In this case, following Bai (1994), the LS estimator of the break point takes

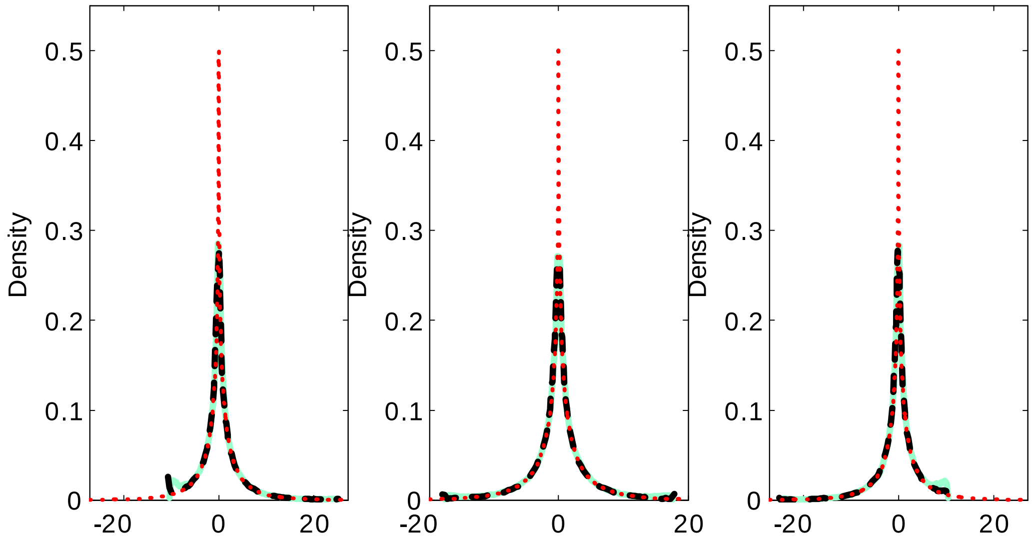

Figure 10: The pdf of $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS,T}-\tau_0)$ when $\tau_0=0.3,0.5,0.7$ (the left, middle and right panel respectively) and $\frac{\delta^*}{\sigma\varepsilon}=6$. The blue solid line is the finite sample distribution when $T=100$; the 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).

the form of

$$\begin{aligned}\widehat{k}_{LS,T} &= \arg\min_{k=1,\dots,T-1}\left\{\sum_{t=1}^{k}\left(Z_t-\overline{Z}_k\right)^2+\sum_{t=k+1}^{T}\left(Z_t-\overline{Z}_k^*\right)^2\right\}\\ &= \arg\max_{k=1,\dots,T-1}\left\{\left[V_k\left(Z_t\right)\right]^2\right\},\end{aligned}\tag{15}$$

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 (13) 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 (15),*
(a) when $h\to 0$ with a fixed ε, we have the following in-fill asymptotic distribution

$$T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS,T}-\tau_0)\xrightarrow{d}\left(\frac{\delta^*}{\sigma\varepsilon}\right)^2\arg\max_{u\in(-\tau_0,1-\tau_0)}\left[\tilde{B}\left(\mu\right)\right]^2,\tag{16}$$

with

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

$B_1(\cdot)$ and $B_2(\cdot)$ being two independent standard Brownian motions;
(b) when $h \to 0$ and $\varepsilon \to 0$ simultaneously with $(\delta^*/\varepsilon)\sqrt{h} \to 0$, we have the following small-ε in-fill asymptotic distribution

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

where $W(u)$ is the two-sided Brownian motion defined in (11).

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 in Part (a) of Theorem 4.2 is asymmetric. When $\tau_0 = 1/2$, the interval becomes symmetric, and we have*

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

which becomes symmetrically distributed about zero. As a result, the distribution in Part (a) of Theorem 4.2 is symmetric about zero when $\tau_0 = 1/2$.

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 (13) follow a weakly stationary process with a long-run variance $[a(1)]^2$. In this case, the results in Theorem 4.2 still applies with σ^2 being replaced by $[a(1)]^2$.*

Figure 11 plots 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 4.1 and the density of the long-span limiting distribution given in Bai (1994). The data are simulated from Model (13) 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. The finite sample bias of the LS estimator $\widehat{\tau}_{LS,T}$, the bias implied by the in-fill asymptotic distribution, and the bias implied by the long-span limiting distribution are reported in Table 2.

Several features are apparent in Figure 11 and Table 2. 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 Table 2 to Table 1, it can be seen that when the means are unknown and have to be estimated, the bias in $\widehat{\tau}_{LS,T}$ increases. In spite of a larger bias in $\widehat{\tau}_{LS,T}$, it can be seen from Table 2 that the in-fill asymptotic distribution captures the finite sample bias reasonably well.

Figures 12-14 plot the finite sample density of $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2 (\widehat{\tau}_{LS,T} - \tau_0)$ when the break size is 0.4, 0.6 and 0.8, respectively, as well as the corresponding 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 Bai (1994). Qualitatively, similar conclusions can be drawn from Figures 12-14 to those from Figure 11. Comparing Figures 12-14 with Figure 11, we can see that, as the break size increases, the trimodality becomes less pronounced and the degree of asymmetry reduces. As a result, the magnitude of bias becomes smaller. Moreover, as the break size gets larger, the long-span asymptotic distribution given in Bai (1994) can better approximate the finite sample distribution. However, the finite sample distribution is less concentrated around zero and less peaked than the long-span asymptotic distribution. In all cases, the in-fill asymptotic distribution given in Part (a) of Theorem 4.1 always provides better approximations to the finite sample distribution than the long-span asymptotic distribution.

5 Bias Correction via Indirect Estimation

The indirect estimation 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 the moments and likelihood function of the model are 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 data simulated from the true model. Indirect inference chooses the parameters of the

Table 2: The bias in finite sample obtained from simulated data, the bias calculated from the in-fill asymptotic distribution, and the bias calculated from the long-span asymptotic distribution in Bai (1994). The number of replications is set at 100,000.

Case		Bias		
$\frac{\delta^*}{\sigma\varepsilon}$	τ_0	Finite sample	In-fill asymptotics	Long-span asymptotics
2	0.3	0.1704	0.1619	0
2	0.7	-0.1717	-0.1611	0
4	0.3	0.1068	0.0885	0
4	0.7	-0.1062	-0.0874	0
6	0.3	0.0511	0.0363	0
6	0.7	-0.0495	-0.0362	0
8	0.3	0.0202	0.0123	0
8	0.7	-0.0199	-0.0122	0

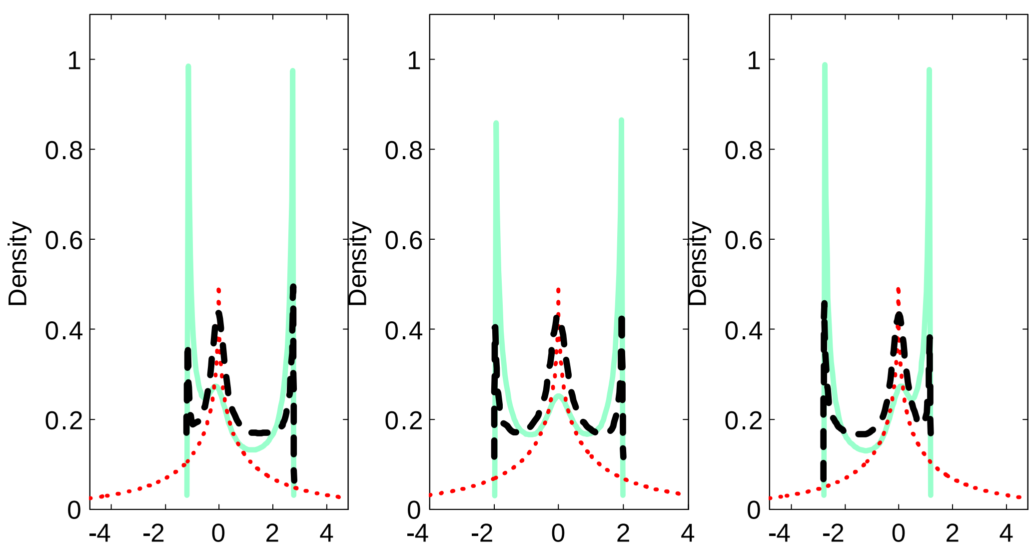

Figure 11: The pdf of $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS}-\tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\frac{\delta^*}{\sigma\varepsilon} = 2$. The blue solid line is the finite sample distribution when $T = 100$; 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).

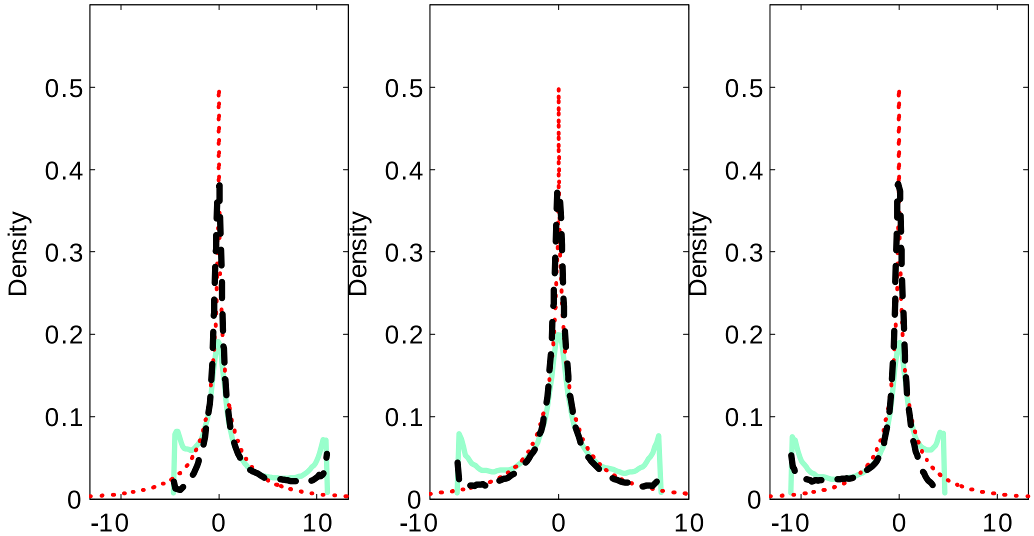

Figure 12: The pdf of $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS}-\tau_0)$ when $\tau_0=0.3,0.5,0.7$ (the left, middle and right panel respectively) and $\frac{\delta^*}{\sigma\varepsilon}=4$. The blue solid line is the finite sample distribution when $T=100$; the 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).

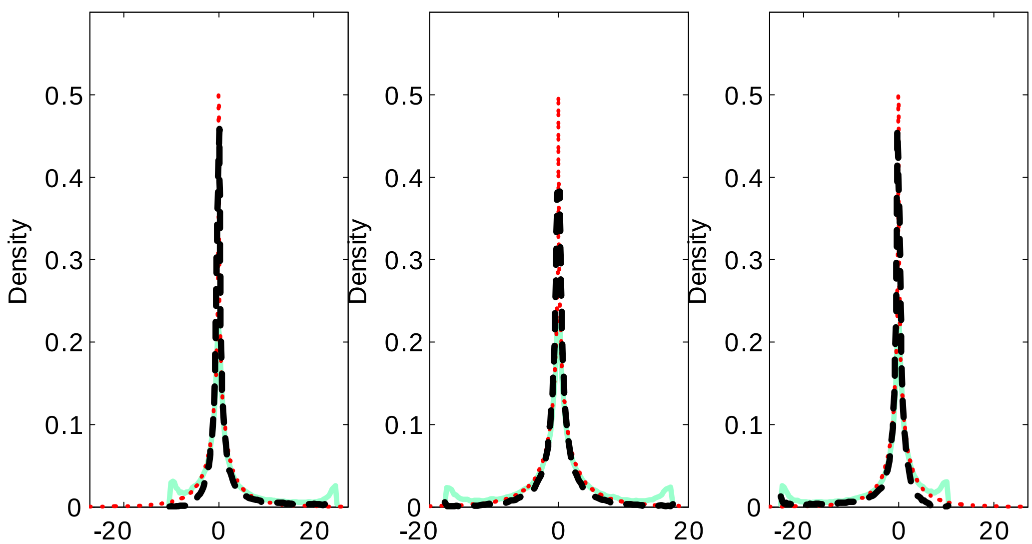

Figure 13: The pdf of $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS}-\tau_0)$ when $\tau_0=0.3,0.5,0.7$ (the left, middle and right panel respectively) and $\frac{\delta^*}{\sigma\varepsilon}=6$. The blue solid line is the finite sample distribution when $T=100$; the 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).

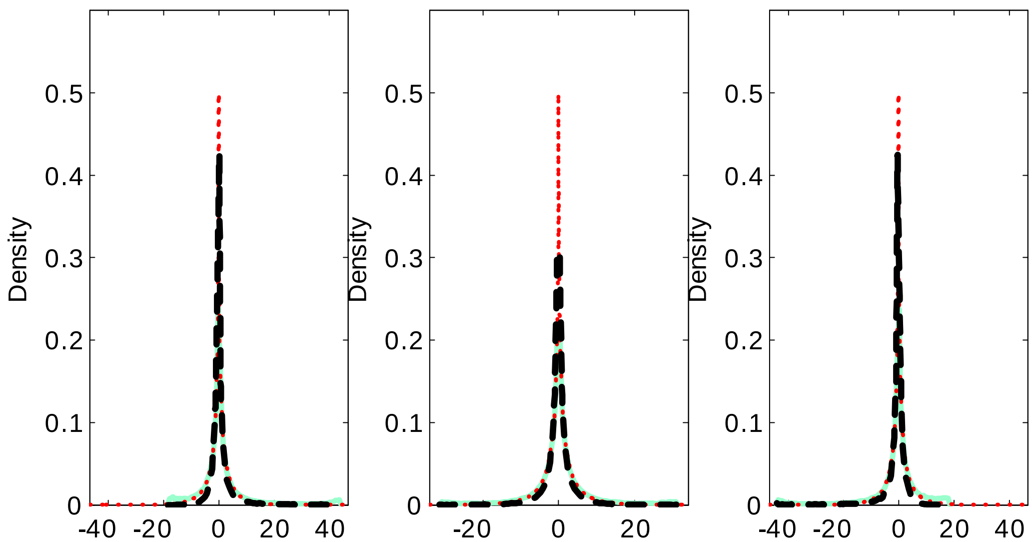

Figure 14: The pdf of $T\left(\frac{\delta^*}{\sigma\varepsilon}\sqrt{h}\right)^2(\widehat{\tau}_{LS}-\tau_0)$ when $\tau_0 = 0.3, 0.5, 0.7$ (the left, middle and right panel respectively) and $\frac{\delta^*}{\sigma\varepsilon} = 8$. The blue solid line is the finite sample distribution when $T = 100$; 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).

true model so that these two sets of parameter estimates of the auxiliary model are as close as possible. Typically, one chooses the auxiliary model that is amenable to estimate and approximate the true model well at the same time.

Gouriéroux et al. (1993) and Gallant and Tauchen (1996) established the asymptotic properties of the indirect estimator, including consistency, asymptotic normality, and asymptotic efficiency. McKinnon and Smith (1998) and Gouriéroux et al. (2000) developed a particular indirect estimation procedure, where the auxiliary model is chosen to be the true model in order to improve finite sample properties of the original estimator. Arvanitis and Demos (2014) established sufficient conditions for second order bias correction of the general indirect estimator. Moreover, they give primitive conditions for finite sample properties of the general indirect estimator and also introduced an iterative procedure to further improve the performance of the indirect estimator.

When the auxiliary model is identical to the true model, the indirect estimation obtains the bias function by simulating from the true model and hence also the auxiliary model. In this section, we apply this indirect estimation procedure to do bias correction in estimating τ and k, the fractional and the absolute structural break point. It is important to obtain the bias function via simulations because, from Equations (12), we know that the bias formula and the bias expansion are too difficult to deal with explicitly. The same idea was also used to estimate continuous time models in Phillips and Yu (2009a,c) and dynamic panel models in Gouriéroux et al. (2010).

The application of the indirect estimation procedure for estimating structural break proceeds as follows. Given a parameter θ (say τ), we simulate data $\tilde{\boldsymbol{y}}(\theta) = \{\tilde{y}_0^h, \tilde{y}_1^h, \ldots, \tilde{y}_T^h\}$ from the true model, such as, Equation (9) or (13), where $h = 1, ..., H$, with H being 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.

The indirect estimation method matches the estimator from the actual observations with the one estimated from the simulated data to obtain the indirect estimator. To be specific, let $Q_T(\theta; \boldsymbol{y})$ be the objective function of the original (biased) estimation method applied to actual data $(\boldsymbol{y})$ for estimating the parameter θ. The corresponding extremum estimator $\hat{\theta}$ obtained is then denoted as

$$\hat{\theta}_T = \arg\max_{\theta\in\Theta} Q_T(\theta; \boldsymbol{y}),$$

and the corresponding estimator based on the hth simulated path for some fixed θ is

$$\tilde{\theta}_T^h(\theta) = \arg\max_{\theta\in\Theta} Q_T(\theta; \boldsymbol{y}(\theta)),$$

where Θ is a compact parameter space.

The indirect estimator is then defined as

$$\hat{\theta}_{IE,T,H} = \arg\min_{\theta\in\Theta} \left\| \hat{\theta}_T - \frac{1}{H}\sum_{h=1}^{H} \tilde{\theta}_T^h(\theta) \right\|, \tag{17}$$

for some distance measure $\|\cdot\|$. When H goes to infinity, it is expected that $\frac{1}{H}\sum_{h=1}^{H} \tilde{\theta}_T^h(\theta) \xrightarrow{p} E(\tilde{\theta}_T^h(\theta))$. Then the indirect estimator becomes

$$\hat{\theta}_{IE,T} = \arg\min_{\theta\in\Theta} \left\| \hat{\theta}_T - b_T(\theta) \right\|$$

where $b_T(\theta) = E(\tilde{\theta}_T^h(\theta))$ is the binding (or bias) function.

To apply the indirect estimation to the observed data, we assume that the true model is given either by the continuous time model given by (9) or the discrete time model given by (13). At first, we employ the ML method and the LS method to the actual data in order to obtain $\widehat{\tau}_{ML}$ and $\widehat{\tau}_{LS,T}$. Then the corresponding estimator for the hth simulated path is $\widetilde{\tau}_T^h(\tau)$ and the indirect estimator is

$$\widehat{\tau}_{IE,T} = \arg\min_{k\in\Theta} \left\| \widehat{\tau}_T - b_T(\tau) \right\|, \tag{18}$$

where $\widehat{\tau}_T$ is the original ML estimator or LS estimator of τ from the actual data, $b_T(\tau)$ is the binding function with the form

$$b_T(\tau) = E(\widetilde{\tau}_T^h(\tau)),$$

which, in practice, can be effectively replaced by $\frac{1}{H}\sum_{h=1}^{H}\widetilde{\tau}_T^h(\tau)$ since H can be chosen arbitrarily large.

Based on $\widehat{\tau}_{IE,T}$, we can define the indirect estimator of the absolute break point as $\hat{k}_{IE,T} = \widehat{\tau}_{IE,T} \times T$. Let the corresponding binding function be $b_T(k) = b_T(\tau) \times T$. If the binding function is invertible, then "b_T-mean-unbiasedness" can be defined as $b_T^{-1}\left(E\left(b_T\left(\widehat{\tau}_{IE,T}\right)\right)\right) = \tau_0$. Gouriéroux et al (2000) gives the conditions under which the indirect estimator is b_T-mean-unbiased for τ. By the same reason, $\hat{k}_{IE,T}$ is "b_T-mean-unbiased" if $b_T^{-1}\left(E\left(b_T\left(\widehat{k}_{IE,T}\right)\right)\right) = k_0$. Moreover, Gouriéroux et al (2000) shows that if $b_T(\cdot)$ is an affine function, the indirect estimator is exactly mean unbiased. One may deduce that when the binding function is close to be affine, the indirect estimator is close to be exactly mean-unbiased. Gouriéroux et al (2000) gives conditions for the second order bias correction by the indirect estimator when the auxiliary model is identical to the true model and $\widehat{\tau}_T$ is consistent. Arvanitis and Demos (2014) gives sufficient conditions for the second order bias correction by the general indirect estimator. In particular, they show that the indirect estimator is second order unbiased if the binding function $b_T(\cdot)$ is asymptotically linear.

Since $\widehat{\tau}_{ML}$ and $\widehat{\tau}_{LS,T}$ are consistent when $\varepsilon \to 0$, we can establish the second order bias correction by the indirect estimator. To derive the asymptotic distribution of the indirect 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 $\widehat{\tau}_{LS,T}$ when $\varepsilon \to 0$ and the asymptotic theory (including the rate of convergence and the limiting distribution) should be the same as the original estimator.

When ε is fixed, if the binding function is invertible, that is, $\widehat{\tau}_{IE,T} = b^{-1}(\widehat{\tau}_T)$, one may informally apply the Delta method to study the efficiency of the indirect estimator as

$$\mathrm{Var}(\widehat{\tau}_{IE,T}) \approx \left(\frac{\partial b_T(\tau_0)}{\partial \tau}\right)^{-2} \mathrm{Var}(\widehat{\tau}_T). \tag{19}$$

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 $\left|\frac{\partial b_T(\tau_0)}{\partial \tau}\right| > 1$, $\widehat{\tau}_T^{IE}$ will have a smaller variance than $\widehat{\tau}_T$. If the finite sample distribution developed in Section 3 suggests that τ is over estimated when $\tau_0 < 50\%$ and is under estimated when $\tau_0 > 50\%$, the binding function is expected to be flatter than the 45 degrees line. As a result, we expect some efficiency loss from the indirect estimation as the variance of the indirect estimator will be larger than that of the original estimator.

Alternative bias correction methods may be applied to correct the bias in the original estimator. For example, if the median is chosen to be the binding function, then the median unbiased estimator of Andrews (1993) has the form of

$$\widehat{\tau}_{MU,T} = \arg\min_{k\in\Theta} \left\| \widehat{\tau}_T - \hat{\rho}_{0.5}\left(\widetilde{\tau}_T^1(\tau), ..., \widetilde{\tau}_T^H(\tau)\right) \right\|, \tag{20}$$

where $\hat{\rho}_{0.5}$ is the median obtained from $\left\{\widetilde{\tau}_T^h(\tau)\right\}_{h=1}^H$. If $\lim_{H\to\infty} \hat{\rho}_{0.5}\left(\widetilde{\tau}_T^1(\tau), ..., \widetilde{\tau}_T^H(\tau)\right)$ exists and is invertible and monotonic, then $\widehat{\tau}_{MU,T}$ is exactly median unbiased. The motivation for using the median to capture the location is because the finite sample distribution of $\widehat{\tau}_T$ is asymmetric in which case the median is a better measure of the location than the mean.

For another example, one may use the bootstrap method of Efron (1979) to reduce the bias. The parametric bootstrap was shown to be an effective method for bias correction (Hall 1992). The performance of the parametric bootstrap was recently illustrated in the parameter estimation in the context of continuous time models in Tang and Chen (2009). 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 estimator from each bootstrap sample path by applying the same estimation procedure (call them $\tau_{*T}^h(\widehat{\tau}_T)$, $h = 1, ..., H$). Let $\overline{\tau}_{*T}(\widehat{\tau}_T) = \frac{1}{H}\sum_{h=1}^H \tau_{*T}^h(\widehat{\tau}_T)$. The bootstrap estimator of the bias is $\overline{\tau}_{*T}^h(\widehat{\tau}_T) - \widehat{\tau}_T$ and, hence, the bootstrap bias-corrected estimator of τ is $\widehat{\tau}_T - (\overline{\tau}_{*T}(\widehat{\tau}_T) - \widehat{\tau}_T) = 2\widehat{\tau}_T - \overline{\tau}_{*T}(\widehat{\tau}_T)$. Gouriéroux et al. (2000) compares the higher order properties of the indirect estimator and the bootstrap method based on the Edgeworth expansions. More simulation based methods can be used. Forneron and Ng (2015) discusses a variety of simulation based methods.

6 Monte Carlo Results

In this section, we design two Monte Carlo experiments to examine the bias of ML estimator of τ in the continuous time model (9) and in the LS estimator of k in the discrete time model (1). We also compare the finite sample performance of the indirect estimator with that of the original estimator, the median unbaised estimator and the parametric bootstrap estimator. Following the suggestions from a referee, we do not invert the binding function to obtain the indirect estimator as the inversion may lead to an estimate outside of the interval $[0, 1]$. Instead we obtain the indirect estimator based on the definition (17).

In the first experiment, data are generated from Model (9), with $\sigma = 1$, $\varepsilon = 1$, $\delta^* = 2, 4, 6$, $\tau_0 = 30\%, 50\%, 70\%$, $dB(t)$ ~i.i.d.$N(0, h)$, where $h = \frac{1}{10000}$. For each combination of δ^* and τ_0, we obtain the ML estimate of τ from Equation (12) and the indirect estimator with $H = 10,000$.[4] Our focus is to examine the finite sample properties of $\hat{\tau}$, so it is assumed that the structural break size δ^*/ε and the standard deviation σ are known during the simulation.

Table 3 reports the bias, the standard error, and the root mean squared errors (RMSE) of ML estimate and the indirect estimate of τ, obtained from 100,000 replications. Some observations can be obtained from the table. Firstly, when $\tau_0 = 50\%$, the ML estimate does not have any noticeable bias in all cases. However, when $\tau_0 \neq 50\%$, ML suffers from a bias problem. For example, when $\tau_0 = 30\%$ and $\frac{\delta^*}{\sigma\varepsilon} = 2$, the bias is 0.0912 and about 30% of the true value. This is very substantial. In general, the bias becomes larger when τ_0 is further away from 50%, 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. Secondly, in all cases when $\tau_0 \neq 50\%$, the indirect estimate substantially reduces the bias. For example, when $\frac{\delta^*}{\sigma\varepsilon} = 2$ and $\tau_0 = 70\%$, the indirect estimation method removes about two thirds of the bias in ML. Finally, the bias reduction by the indirect estimation method comes with a cost of a higher variance, which causes the RMSE of the indirect estimate slightly higher than its ML counterpart.

Table 3 also reports the statistics of the median unbiased estimator and the bootstrap estimator of τ, also obtained from 100,000 replications. Compared with the indirect estimation, the median unbiased estimation is less effective than for bias reduction but is more efficient in terms of variance. In terms of RMSE, the median unbiased estimation performs better. This finding is consistent with what was reported in Tables 7-8 of Phillips and Yu (2009a) in the continuous time model. However, compared with the indirect estimation, the bootstrap method performs similarly in terms of bias reduction but increases the variance more than the indirect estimation in almost all cases.

In the second experiment, data are generated from Model (1), with $\sigma = 1$, $\varepsilon = 1$, $\delta^* = 0.2, 0.4$ and 0.6, $\tau_0 = 0.3, 0.5, 0.7$, ϵ_t ~i.i.d.$N(0, 1)$, where we choose $T =$

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

Table 3: Monte Carlo comparison of the bias and RMSE of ML estimates, median unbiased estimates, parametric bootstrap estimates and indirect estimates. The number of simulated path is set to be 10,000 for the median unbaised method, the parametric bootstrap method and indirect estimation method. 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

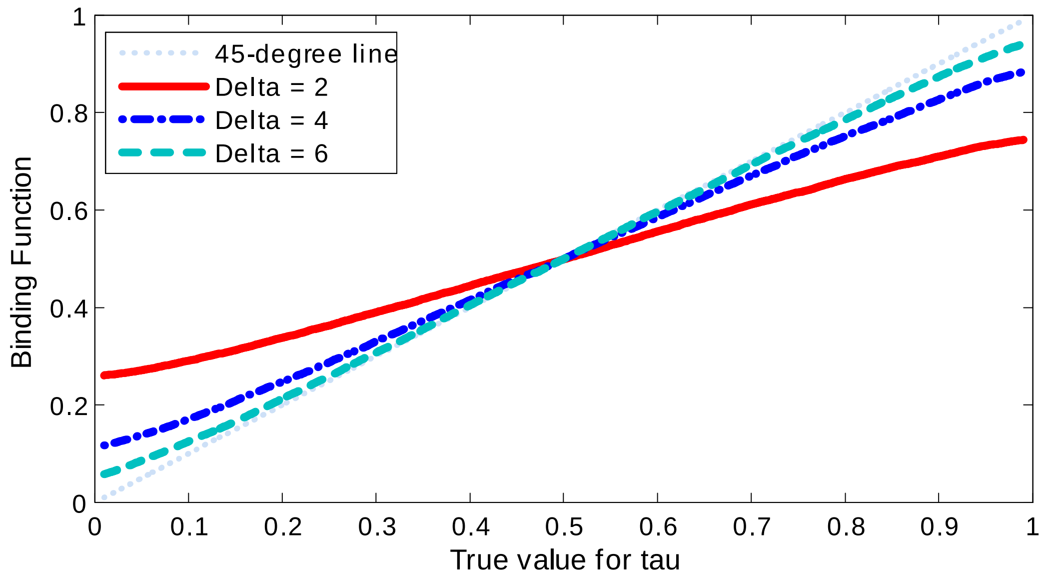

Figure 15: Binding function of ML for the continuous time model when $h = 0.0001$.

$80, 100, 120$. For each combination of δ^*, τ_0 and T, we obtain the LS estimate of k based on Equation (4) and the indirect estimate for each replication. As in the continuous time model, it is assumed that the structural break size δ^*/ε and the standard deviation σ are known. The reason why we focus on k is that k is a practically important parameter.

Table 4 reports the bias, the standard error, and the RMSE of the LS estimator and the indirect estimator of k, obtained from 100,000 replications. We may draw the following conclusions from Table 4. Firstly, when $\tau_0 = 50\%$, the LS estimate does not have any noticeable bias in all cases. However, when $\tau_0 \neq 50\%$, LS suffers from a bias problem. For example, when $T = 80, \tau_0 = 30\%$ and $\frac{\delta^*}{\sigma\varepsilon} = 0.2$, the bias is 8.1045 while the true value of k is 24. The bias is about 34% of the true value, which 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. Secondly, in all cases when $\tau_0 \neq 50\%$, the indirect estimate substantially reduces the bias. For example, when $T = 100$, $\frac{\delta^*}{\sigma\varepsilon} = 0.2$ and $\tau_0 = 30\%$, the indirect estimation method removes more than half of the bias in LS. Finally, the bias reduction by the indirect estimation method comes with a cost of a higher variance, which causes the RMSE of the indirect estimate slightly higher than its LS counterpart.

Table 4 also reports the bias, the standard error, and the RMSE of the median unbiased estimator and the bootstrap estimator of k, also obtained from 100,000 repli-

Table 4: Monte Carlo comparison of the bias and RMSE of LS, Median Unbiased, Parametric Bootstrap and Indirect Estimates. The number of simulated path is set to be 10,000 for median unbaised, parametric bootstrap and indirect estimation. The number of replications is set at 100,000.

Case				Bias				Standard error				RMSE			
T	$\frac{\delta}{\sigma}$	τ_0	k_0	LS	MU	PB	IE	LS	MU	PB	IE	LS	MU	PB	IE
80	0.2	0.3	24	8.1045	6.5020	4.2129	4.2479	22.5150	25.9251	34.5476	30.4505	23.9293	26.7280	34.8035	30.7454
80	0.2	0.5	40	0.0099	-0.0762	0.4524	0.3123	22.2311	25.5596	34.1048	30.0445	22.2311	25.5597	34.1078	30.0461
80	0.2	0.7	56	-8.0730	-6.6366	-3.4944	-3.4891	22.5175	25.9119	34.4718	30.4846	23.9209	26.7483	34.6484	30.6836
80	0.4	0.3	24	3.1931	2.8585	1.1935	0.2573	16.1841	16.7072	20.1255	19.1355	16.4961	16.9500	20.1609	19.1372
80	0.4	0.5	40	-0.0033	-0.0204	0.8996	-0.0436	16.3752	16.8505	20.1440	19.4841	16.3752	16.8505	20.1641	19.4842
80	0.4	0.7	56	-3.1703	-2.9039	0.4772	-0.2703	16.1693	16.6757	19.8349	19.2144	16.4772	16.9266	19.8406	19.2163
80	0.6	0.3	24	0.9818	0.9683	0.8462	-0.1242	10.7869	10.8179	11.8190	11.6133	10.8315	10.8611	11.8492	11.6140
80	0.6	0.5	40	0.0070	0.0066	0.9317	0.0042	11.2217	11.2440	12.2760	12.0716	11.2217	11.2440	12.3113	12.0716
80	0.6	0.7	56	-0.9402	-0.9269	1.0206	0.2092	10.7894	10.8218	11.9522	11.7125	10.8303	10.8615	11.9957	11.7143
100	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
100	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
100	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
100	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
100	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
100	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
100	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
100	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
100	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
120	0.2	0.3	36	9.8999	7.8775	4.2082	3.7117	31.7253	35.1662	45.9043	41.7342	33.2340	36.0377	46.0968	41.8989
120	0.2	0.5	60	-0.0430	-0.2779	0.2209	0.1827	31.5139	34.8725	45.5344	41.5039	31.5139	34.8736	45.5350	41.5043
120	0.2	0.7	84	-10.0532	-8.5837	-3.1223	-3.3778	31.8510	35.2401	46.0509	41.9541	33.3999	36.2704	46.1567	42.0899
120	0.4	0.3	36	2.8649	2.6831	0.8284	-0.0508	20.3247	20.6738	23.1646	22.9143	20.5256	20.8472	23.1794	22.9143
120	0.4	0.5	60	-0.0707	-0.0761	1.0478	-0.1194	20.8886	21.1797	23.8508	23.5692	20.8886	21.1799	23.8738	23.5692
120	0.4	0.7	84	-2.8582	-2.6931	0.8727	0.0773	20.3374	20.6780	23.3805	23.0592	20.5372	20.8526	23.3968	23.0593
120	0.6	0.3	36	0.5883	0.5834	0.8692	-0.1961	12.2823	12.2969	13.0589	12.8577	12.2964	12.3107	13.0878	12.8592
120	0.6	0.5	60	0.0038	0.0038	0.9825	-0.0010	12.7882	12.7979	13.3917	13.1536	12.7882	12.7979	13.4277	13.1536
120	0.6	0.7	84	-0.5510	-0.5462	1.1217	0.1203	12.2804	12.2946	12.7169	12.9038	12.2928	12.3067	12.7663	12.9044

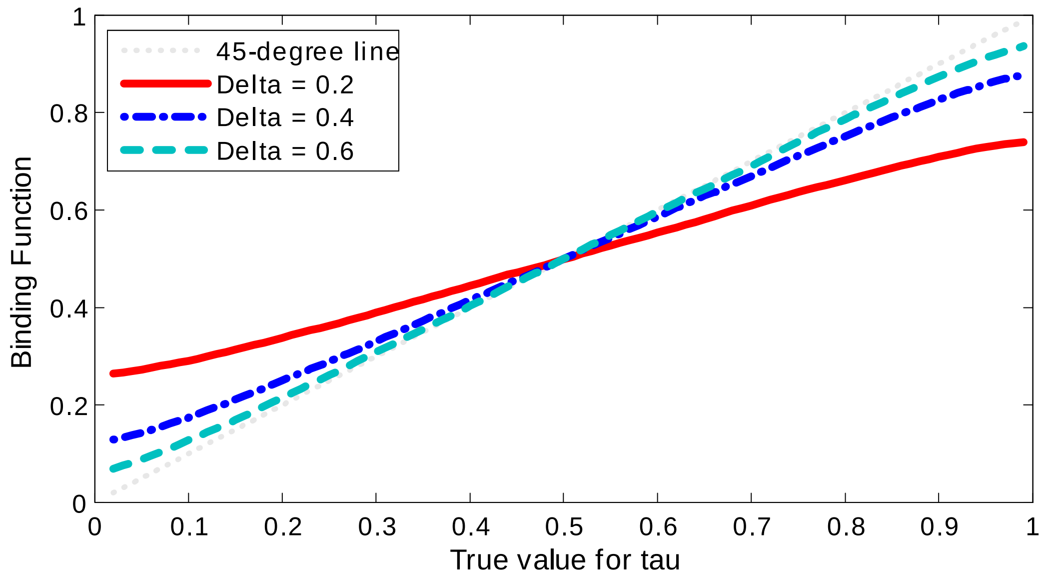

Figure 16: Binding function of LS for the discrete time model when $T = 100$

cations. Compared with the indirect estimation, the median unbiased estimation is less effective for bias reduction but is more efficient in terms of variance. In terms of RMSE, the median unbiased estimation performs better than the indirect estimation. However, compared with the indirect estimation, the bootstrap method performs similarly in terms of bias reduction but increases the variance more than the indirect estimation. So the bootstrap method is dominated by the indirect estimation method in terms of the finite sample property.

To understand why the indirect estimation increases the variance relative to the original estimator, we plot the binding functions in these two models in Figure 15 and Figure 16, where we also plot the 45 degrees line for the purpose of comparison. Figure 15 corresponds to the continuous time model with $\frac{\delta^*}{\sigma\varepsilon} = 2, 4, 6$ and Figure 16 to the discrete time model with $T = 100$, $\frac{\delta^*}{\sigma\varepsilon} = 0.2, 0.4$ and 0.6. Several conclusions can be made. First, the binding functions always pass through the 45 degrees line at the middle point of τ, suggesting no bias when $\tau = 50\%$ and that the bias becomes smaller when the true break point gets close to the middle. Second, the binding functions are flatter than the 45 degrees line in all cases, explaining why the variance of the indirect estimate is larger than that of ML estimate. The smaller the signal-to-noise ratio, the flatter the binding function and hence the bigger loss in efficiency. Third, the binding function is not exactly a straight line. It is easy to see the nonlinearity near the two boundary points. This explains why the indirect estimator is not exactly mean unbiased.

7 Conclusions

This paper is concerned about the large sample 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 developed long-span limiting distribution is the distribution of the location of the extremum of a two-sided Brownian motion with triangular drift over the interval $(-\infty, +\infty)$, which is symmetric and has the origin 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. The exact distribution is the distribution of the location of the extremum of a two-sided Brownian motion with triangular drift over a finite interval, the two boundary points of which depend on the location of the true break point. 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. In all cases, the in-fill asymptotic distribution approximates the finite sample distribution better 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 indirect estimation 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 indirect estimation procedure is effective in reducing the bias in the commonly used break point estimators while the variance of the estimation is increased.

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. How to extend the indirect estimation technique to a multiple parameters setting is also wide-open.

Appendix

Proof of Theorem 3.1: (a) 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 (11). 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 Theorem 3.1 immediately.

(b) It is a straightforward result of Part (a). A formal proof for a general model can be found in IH (1981), Theorem 2.2.

Proof of Theorem 4.1: (a) Let $\Gamma(k) = -(\frac{\delta^*}{\varepsilon})\sqrt{h}\sum_{t=1}^{k}\left(Z_t - \mu\sqrt{h}\right) + \left(\frac{\delta^*}{\varepsilon}\right)^2 hk/2$. Then, the LS estimator $\widehat{k}_{LS,T}$ defined in (14) can be expressed as

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

As $T = 1/h$, $\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^2 \left(\widehat{k}_{LS,T} - k_0\right) = (\delta^*/\varepsilon)^2 \left(\widehat{\tau}_{LS,T} - \tau_0\right) = O_p(1)$ 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 $t \leq k_0$, 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 (11). 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 $k > k_0$,

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

We then have,

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

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

It is easy to see that for any k

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

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

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

Note that, for any $1 \leq k < T$, $\Gamma(k) - E(\Gamma(k)) = -(\frac{\delta^*}{\varepsilon}\sqrt{h})\sum_{t=1}^{k}\epsilon_t$ 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 $\Gamma(k) - E(\Gamma(k)) = O_p(\delta^*/\varepsilon)$ as $\varepsilon \to 0$. Therefore,

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

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 $\Delta > 0$, $Pr\{\widehat{\tau}_{LS,T} \notin (\gamma, 1-\gamma)\} < \Delta$ when $h \to 0$ and $\varepsilon \to 0$ simultaneously with $(\delta^*/\varepsilon)\sqrt{h} \to 0$. Thus, we now only need to examine the behavior of $\Gamma(k)$ over those k for which $T\gamma < k < T(1-\gamma)$. 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 $Pr\left\{|\widehat{\tau}_{LS,T} - \tau_0| \geq M\left(\sqrt{T}\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^{-2}\right\} \to 0$ when $M \to \infty$, $h \to 0$ and $\varepsilon \to 0$ simultaneously with $(\delta^*/\varepsilon)\sqrt{h} \to 0$.

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

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

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

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

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

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

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

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 (13) we have $Z_t - \mu\sqrt{h} = \epsilon_t \sim$ i.i.d.$(0, \sigma^2)$ for $t \le k_0$ 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 $k \le k_0$,

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

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 (15) can be identically expressed as

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

When $h \to 0$ with a fixed ε, we have $\left(\frac{\delta^*}{\varepsilon}\sqrt{h}\right)^2\left(\widehat{k}_{LS,T} - k_0\right) = (\delta^*/\varepsilon)^2(\widehat{\tau}_{LS,T} - \tau_0) = O_p(1)$ taking 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 $\left[\sqrt{T}V_k(Z_t)\right]^2$ 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 $s \in \left(-\tau_0(\delta^*/\varepsilon)^2, (1-\tau_0)(\delta^*/\varepsilon)^2\right)$. 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

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

and

$$\frac{\sqrt{T}}{T-k}\sum_{t=k+1}^{T}\epsilon_t = \frac{T}{T-k}\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T}\epsilon_t \Rightarrow \frac{\sigma}{1-\tau_0-u}B_2(1-\tau_0-u),$$

where $B_1(\cdot)$ and $B_2(\cdot)$ are two independent standard Brownian motions determined by the errors before and after k respectively. As a result, for $k \leq k_0$,

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

Similarly, for $k > k_0$,

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

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

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

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

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

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

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

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

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 $1 \leq k < T$,

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

Therefore, when $\varepsilon \to 0$,

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

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] Andreou, E., and Ghysels, E., 2009. Structural breaks in financial time series. In: Handbook of financial time series. Springer Berlin Heidelberg, pp. 839-870

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

[3] 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.

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

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

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

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

[8] Bai, J., 2010. Common breaks in means and variances for panel data. Journal of Econometrics 157, 78-92.

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

[10] 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.

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

[12] 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.

[13] Chen, J. and Gupta, A.K., 2011, *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*, Birkhäuser.

[14] Csörgő M. and Horváth, L., 1997, *Limit theorems in change-point analysis*, Wiley.

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

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

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

[18] Gouriéroux, C., Monfort, A., and Renault, E., 1993. Indirect estimation. Journal of applied econometrics 8, 85-118.

[19] 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.

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

[21] Hansen, B. E., 2001. The new econometrics of structural break: Dating breaks in US labor productivity. Journal of Economic perspectives, 117-128.

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

[23] 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.

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

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

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

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

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

[29] Le Cam, L., 1960. Locally asymptotically normal families of distributions. University of California Publications in Statistics 3: 37–98.

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

[31] Nickell, S., Biases in dynamic models with fixed effects. Econometrica 49, 1417–1426.

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

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

[34] Perron, P., 2006. Dealing with structural breaks. Palgrave Handbook of Econometrics 1, pp. 278-352.

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

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

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

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

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

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

[41] Tang, C. Y., and S. X. Chen, 2009, Parameter Estimation and Bias Correction for Diffusion Processes, Journal of Econometrics 149, 65-81.

[42] 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, 1321-1328.

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

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

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