12-2014

On bias in the estimation of structural break points

Liang JIANG
Singapore Management University, liang.jiang.2011@phdecons.smu.edu.sg

Xiaohu WANG
Chinese University of Hong Kong

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

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

SMU
SINGAPORE MANAGEMENT UNIVERSITY

On Bias in the Estimation of Structural Break Points

Liang Jiang, Xiaohu Wang and Jun Yu

December 2014

Paper No. 22-2014

On Bias in the Estimation of Structural Break Points*

Liang Jiang
Singapore Management University

Xiaohu Wang
The Chinese University of Hong Kong

Jun Yu
Singapore Management University

December 27, 2014

Abstract

Based on the Girsanov theorem, this paper obtains the exact finite sample distribution of the maximum likelihood estimator of structural break points in a continuous time model. The exact finite sample theory suggests that, in empirically realistic situations, there is a strong finite sample bias in the estimator of structural break points. This property is shared by least squares estimator of both the absolute structural break point and the fractional structural break point in discrete time models. A simulation-based method based on the indirect estimation approach is proposed to reduce the bias both in continuous time and discrete time models. 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 change, Bias reduction, Indirect estimation, Break point

*Yu acknowledges support from the Singapore Ministry of Education for Academic Research Fund under 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 in both econometrics and 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.

In terms of estimating structure break points, the literature has developed asymptotic theory for estimating the (fractional) structural break point (i.e. the (absolute) structural break point divided by the total sample size), including consistency, rates of convergence, and limit distributions; see, for example, Yao (1987) and Bai (1994). Interestingly and rather surprisingly, the finite sample theory for estimating structure break points seems to have received little attention in the literature. Is this lack of attention due to the good approximation of the asymptotic distribution to the finite sample distribution in empirically realistic cases and hence there is no need to study the finite sample theory? In particular, is there any bias in the traditional estimator of structural break points? Simulations provided in Yao (1987) seem to suggest that the asymptotic distribution is not necessarily close to the finite sample distribution while simulations provided in Bai (1994) seem to suggest there is little bias in the traditional estimator when the true break point is in the middle of the sample. Or is the lack of attention due to the difficulty in studying the finite sample theory and in approximating the bias, even in the first order?

This paper systematically investigates the finite sample 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 finite sample issues in the literature. We develop the finite sample distribution of the maximum likelihood (ML) estimator of the structural break point in a continuous time model and relate the continuous time model to the discrete time models studied in the literature. We also document the bias both in the continuous time and the discrete time models, and propose an indirect estimation procedure to alleviate the bias via simulations.

Our study makes several contributions to the literature on structural breaks. First, we obtain the finite sample distribution of the ML and least squares (LS) estimators in some simple models and then obtain the bias from the finite sample distribution. It is shown that the bias can be substantial in the ML/LS estimators of the fractional structural break point and the absolute structural break point. The further the fractional structural break point away from 50%, the more the bias. When the fractional structural break point is smaller (bigger) than 50%, the bias is positive (negative).

Second, we develop a novel approach to obtaining the finite sample distribution. Since the likelihood function and the sum of squared residuals are not differentiable with respect to break point in the discrete time models, the traditional approaches of obtaining the finite sample theory are not feasible. By using the Girsanov theorem, we obtain the likelihood function in a continuous time model with a structural break and then obtain the finite sample distribution of the ML estimator.

Third, we propose to do bias reduction 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 analytic approach as in Kendall (1954), Nickell (1981), Yu (2012) for various types of autoregressive models. However, it is difficult to use the analytic approach in this context as the bias formula is difficult to obtain. It is shown that the indirect estimation procedure, without knowing the analytical form to approximate the bias, achieves substantial bias reduction. 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. 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 analytic expression for the bias function. 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 the structural break points.

The rest of the paper is organized as follows. In Section 2, we first briefly review the literature and then develop a continuous time model with a structural break and discuss the finite sample properties of the ML estimator of the structural break point. Section 3 connects the continuous time model to the discrete time models previously considered in the literature. Section 4 introduces the indirect estimation technique and applies it to both the continuous time and the discrete time models with structural break points. In Section 5, Monte Carlo experiments are designed to obtain the bias of traditional estimators in models with structural breaks. We also compare the finite sample performance of the indirect estimation estimate with that of the traditional estimation methods. Sections 6 concludes.

2 Bias in a Continuous Time Model

2.1 A literature review and motivations

The literature on estimating structural break points is extensive. A partial list of contributions in statistics include Chernoff and Zacks (1964), Hinkley (1969, 1970), Bhattacharya and Brockwell (1976), Ibragimov and Has'minskii (1981), Hawkins et al. (1986), Bhattacharya (1987), and Yao (1987). A key reference is Hinkley (1970) that develops not only the ML method for estimating the absolute break point but also its distributional behavior as the sample sizes before and after the change-point tend to infinity. In econometrics, Jushan Bai and Pierre Perron have made many contributions to the literature through their individual work as well as their collaborative work; see for example, Perron (1989), Bai (1994, 1995, 1997a, 1997b, 2010), Bai and Perron (1998) and Bai et al. (1998). For example, Bai (1994) extends the earlier literature by proposing the least squares (LS) method to estimate the break point in linear processes and develop its large sample theory. Bai and Perron (1998) uses the LS method to estimate linear models with multiple structural breaks.

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 $t = 1, \ldots, T$, ϵ_t is a sequence of independent and identically distributed (i.i.d.) continuous random variables with zero mean, k_0 is the true value of the absolute structural break point k, constant μ measures the mean of Y_t before break and δ is the size of structural break. Let the probability density function (pdf) of Y_t be $f(Y_t, \mu)$ for $t \leq k_0$ and $f(Y_t, \mu + \delta)$ for $t > k_0$. And denote τ_0 the true value of the fractional structural break point τ, i.e., $\tau_0 = k_0/T$. Under the assumption that the form of function f and parameters μ and δ are all known and at least one observation comes from each distribution, the ML estimator of k_0 is defined as

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

The corresponding estimator of τ is $\widehat{\tau}_{ML} = \widehat{k}_{ML}/T$. Hinkley (1970) showed that $\widehat{k}_{ML} - k_0$ converges in distribution as the sample sizes before and after the break point tend to infinity. He also pointed out that the distribution of $\widehat{k}_\infty - k_0$, where $\widehat{k}_\infty$ denotes $\widehat{k}_{ML}$ with infinite sample, has no closed-form expression, and gave a numerical method to compute the distribution. However, this numerical scheme is difficult to handle 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 sequential limit distribution as

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

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. Since $I(\mu)$ depends on the error's distribution, no invariance principle applies to the sequential limit distribution. Yao (1987) also derived the pdf of the sequential limit distribution as

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

and its cdf as

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

$G(x) = 1 - G(-x)$ if $x \leq 0$, where $\Phi(x)$ is the cdf of the standard normal distribution.

For the same model as in Equation (1), Hawkins et al. (1986) and Bai (1994) studied the LS estimators of k and τ with unknown μ and δ. The LS estimator of k takes the form of

$$\widehat{k}_{LS} = \arg\min_{k=1,\ldots,T-1} S_k^2 = \arg\max_{k=1,\ldots,T-1} V_k^2, \tag{4}$$

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 $V_k^2 = \frac{T(T-k)}{T^2}\left(\overline{Y}_k^* - \overline{Y}_k\right)^2$. The corresponding estimator of τ is $\widehat{\tau}_{LS} = \widehat{k}_{LS}/T$. Hawkins et al. (1986) showed that $T^\alpha\left(\widehat{\tau}_{LS} - \tau_0\right) \xrightarrow{p} 0$ for any $\alpha < 1/2$. Bai (1994) improved the rate of convergence by showing that $\widehat{\tau}_{LS} - \tau_0 = O_p\left(\frac{1}{T\delta^2}\right)$. This convergence rate also applies to $\widehat{\tau}_{ML}$ when ϵ_t is an i.i.d. Gaussian sequence. Because, in the case where $\epsilon_t \sim$ i.i.d.$N(0,\sigma^2)$, the LS estimator are equivalent to the ML estimator with unknown μ and δ, whose limit theory, as argued in Hinkley (1970), is the same as that of the ML estimator when μ and δ are known as long as μ and δ can be consistently estimated. While $\widehat{\tau}_{LS}$ is consistent, $\widehat{k}_{LS}$ is inconsistent since $\widehat{k}_{LS} - k_0 = O_p\left(\frac{1}{\delta^2}\right)$.

To develop the limit distribution with an invariance principle, δ has to go to zero as $T \to \infty$, as shown in Bai (1994). This kind of limit theory is particularly useful in constructing conference interval when the size of the break is small. Let δ_T be the size of break that depends on T. Bai showed that if $\epsilon_t \sim$i.i.d.$(0, \sigma^2)$, $\delta_T \to 0$ and $\frac{\sqrt{T}\delta_T}{\sqrt{\log T}} \to \infty$ as $T \to \infty$,

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

When ϵ_t is normally distributed, the Fisher information $I(\mu)$ turns out to be σ^{-2}. Therefore, the simultaneous asymptotic distribution in Bai (1994) is the same as the sequential asymptotic distribution in Yao (1987). Bai (1994) also derived the limit distribution when ϵ_t is a short memory ARMA process, which is the same as shown in Equation (5) by replacing σ^2 with the long-run variance of ϵ_t. To obtain the limit distribution, Bai (1994) examined the behavior of normalized objective function in the small neighborhood of the true break point k_0 such that $k = [k_0 + v\left(\delta_T\right)^{-2}]$ where v varies in a bounded interval. This is equivalent to the local asymptotic theory of Le Cam (1960).

A study which is closest to ours is Ibragimov and Has'minskii (1981). Ibragimov and Has'minskii analyzed a simple continuous time model

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

where $t \in [0, 1]$, $S(t - \tau_0)$ is a non-stochastic drift term with discontinuity at time τ_0 (i.e. τ_0 is the structural break point), and ε is a small parameter. Let $\lim_{x \to 0+} S(x) - \lim_{x \to 0-} S(x) = \delta$ denote the size of the break. Following the development of the local asymptotic theory of Le Cam, Ibragimov and Has'minskii (1981), under the assumption that a continuous record is available, examined the behavior of the normalized likelihood ratio in the small neighborhood of the true break point τ_0 such that $\tau = \tau_0 + \varepsilon^2 u$ and showed that as $\varepsilon \to 0$,

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

Figure 1 plots the pdf of the limit distribution given in Yao (1987), Bai (1994), and Ibragimov and Has'minskii (1981). For the purpose of comparison, we also plot the pdf of the standard normal distribution. It can be seen that both distributions are symmetric, suggesting no bias in the limit distribution when estimating the fractional break

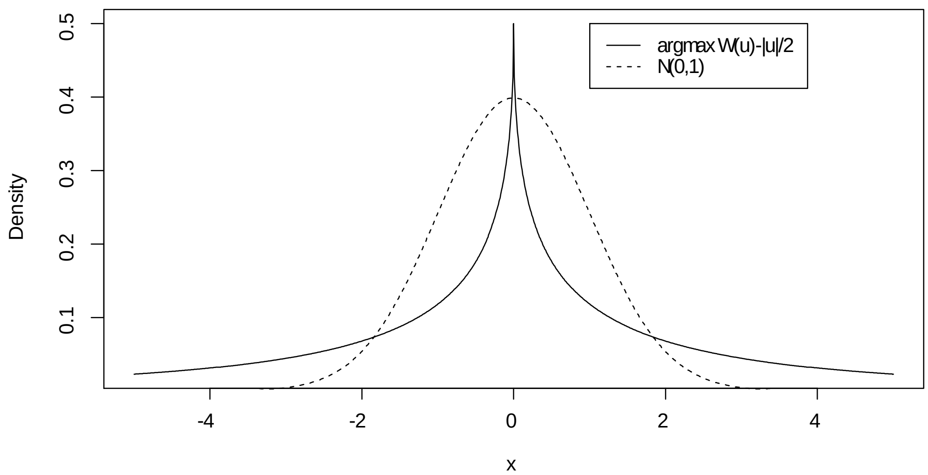

Figure 1: The pdfs of $\arg\max_{u\in(-\infty,\infty)}\left\{W(u)-\frac{1}{2}|u|\right\}$ and a standard norm distribution

point using ML/LS. However, relative to the standard normal distribution, the limit distribution has much fatter tails and a much higher peak. The symmetric property is a result of using the local asymptotic approach to develop the limit distribution in all cases. This property does not help us to understand the finite sample bias in estimating the break points.

The asymptotic arguments above do not take account of asymmetry in the sample before and after the break. To capture the influence of asymmetric information before and after the break, a continuous time model is a natural choice. As long as $\tau_0 \neq 1/2$, the information contained by observations over the time interval $[0, \tau_0]$ and those over the time interval $[\tau_0, 1]$ are different, even if the continuous records are available and there are infinite number of observations before and after the break. This is because in continuous time models the time span also conveys useful information.

There is another motivation to consider a continuous time model. To study the finite sample bias for traditional estimators, a typical approach is to consider the first order condition to an extremum problem that defines the associated estimator; see for example, Rilstone et al. (1996) and Bao and Ullah (2007). While this approach covers many popular models, it is not applicable to the problem of estimating the structural break point in discrete time models, regardless if ML or LS is used. This is because the objective functions in (2) and (4) are not differentiable and hence no first order condition is available for developing high order expansions. Using continuous time models we can avoid this difficulty.

2.2 A continuous time model

The continuous time model considered in the paper is

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

with $t \in [0,1]$, where $S(t-\tau_0)$ is a non-stochastic drift term with discontinuity at time τ_0. Let $\lim_{x\to 0+} S(x) - \lim_{x\to 0-} S(x) = \delta$ denote the size of the break. Different from Model (6) studied in Ibragimov and Has'minskii (1981), we let $\varepsilon = 1$, not $\varepsilon \to 0$. In addition, we have σ in the diffusion function, capturing the noise level. Hence the signal-to-noise ratio in our model is δ/σ, which is a constant, unlike what was assumed in Ibragimov and Has'minskii (1981).

Furthermore, in order to establish a link to the discrete time model in Yao (1987) and Bai (1994), we consider the case in which $S(t-\tau_0)$ only takes two values such that

$$S(t-\tau_0) = \begin{cases} \mu & if\ t \le \tau_0 \\ \mu+\delta & if\ t > \tau_0 \end{cases}, \tag{9}$$

with $t \in [0,1]$, where τ_0 is the unknown true break point, and $\tau_0 \in [\alpha, \beta]$ with $0 < \alpha < \beta < 1$. Consequently Model (8) can be rewritten as

$$dX(t) = (\mu + \delta 1_{[t>\tau_0]})dt + \sigma dB(t). \tag{10}$$

Following Le Cam (1960) and Ibragimov and Has'minskii (1981), we obtain the exact log-likelihood ratio of Model (10) via the Girsanov Theorem[1]

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

The ML estimator of τ_0 is

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

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

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

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

When $\tau > \tau_0$, we have

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

Thus, we can write the exact log-likelihood ratio as

$$\log\left(\frac{dP_\tau}{dP_{\tau_0}}\right) = \frac{\delta}{\sigma}\left(B(\tau_0) - B(\tau)\right) - \frac{\delta^2}{2\sigma^2}|\tau - \tau_0|. \tag{11}$$

This implies that the ML estimator of τ_0 is

$$\widehat{\tau}_{ML} = \arg\max_{\tau\in(0,1)}\left\{\frac{\delta}{\sigma}\left(B(\tau_0) - B(\tau)\right) - \frac{\delta^2}{2\sigma^2}|\tau - \tau_0|\right\}, \tag{12}$$

which leads to

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

We now define a two-sided Brownian motion as

$$W(u) = \begin{cases} W_1(-u) = B(\tau_0) - B(\tau_0 - (-u)) & \text{if } u \leq 0 \\ W_2(u) = B(\tau_0) - B(\tau_0 + u) & \text{if } u > 0 \end{cases}, \tag{13}$$

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 as they are composed by increments of the Brownian motion $B(\cdot)$ before and after the time τ_0 respectively with $W_1(0) = W_2(0) = 0$.

We then have

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

where $\overset{d}{=}$ denotes equivalence in distribution. Consequently, we obtain

$$\left(\frac{\delta}{\sigma}\right)^2(\widehat{\tau}_{ML} - \tau_0) \overset{d}{=} \arg\max_{u\in\left(-\tau_0\left(\frac{\delta}{\sigma}\right)^2,(1-\tau_0)\left(\frac{\delta}{\sigma}\right)^2\right)}\left\{W(u) - \frac{1}{2}|u|\right\}, \tag{14}$$

the exact distribution of the ML estimator $\widehat{\tau}_{ML}$ with a continuous record being available, which is also called in this paper the exact finite sample distribution of $\widehat{\tau}_{ML}$ in the sense that it is obtained with a finite time span before and after the break, which is $(0, \tau_0]$ and $[\tau_0, 1)$ respectively.

It seems that the finite sample distribution given in Equation (14) is similar to the limit distributions given in Yao (1987), Bai (1994) and Ibragimov and Has'minskii (1981) listed in (3), (5) and (7), respectively. However, there is one critical difference between them. The limit distributions in (3), (5) and (7) correspond 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 limit distribution is symmetric about zero. However, the finite sample distribution in (14) corresponds to the location of the extremum of $W(u) - \frac{1}{2}|u|$ over the interval of $\left[-\tau_0 \left(\frac{\delta}{\sigma}\right)^2, (1-\tau_0)\left(\frac{\delta}{\sigma}\right)^2\right]$, therefore depends on the true value of break point τ_0. Only when τ_0 is 50%, that is the true break point is exactly at the middle, $\left[-\tau_0 \left(\frac{\delta}{\sigma}\right)^2, (1-\tau_0)\left(\frac{\delta}{\sigma}\right)^2\right]$ becomes $\left(\frac{\delta}{\sigma}\right)^2 [-50\%, 50\%]$ and symmetric about zero. In this case the finite sample distribution will be symmetric about zero. If τ_0 is not 50% (either smaller or bigger than 50%), the interval and hence the finite sample distribution will be asymmetric. It is easy to see that the finite sample distribution in (14) 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.

Because of this difference in the interval to locate the extremum, we cannot obtain the pdf or cdf of the finite sample distribution in closed-form. As a result, we obtain the pdf by simulations as for the case of the Dickey-Fuller distributions.

Figure 2 plots the densities of $\widehat{\tau}_{ML}$ given in Equation (14) when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively) and the signal-to-noise ratio (δ/σ) is 1. Figure 3-6 plots the densities of $\widehat{\tau}_{ML} - \tau_0$ when the signal-to-noise ratio is 2, 4, 6, 8. There are several interesting observations from these plots. First and most importantly, when $\tau_0 = 50\%$, the densities of $\widehat{\tau}_{ML} - \tau_0$ is always symmetric about zero, no matter what value the signal-to-noise ratio takes. As a result, there is no finite sample bias in this case. However, when τ_0 is not 50%, the density is not symmetric any more. In particular, if τ_0 is less (larger) than 50%, the density is positively (negatively) skewed and there is a 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, although this feature does not show up in the graphs.

Second, there are tri-modality in the finite sample distribution when the signal-to-noise ratio is low (for example when $\delta/\sigma = 1, 2, 4$). The true value is one of the three modes while the two boundary points (0 and 1) are the other two modes. For very small

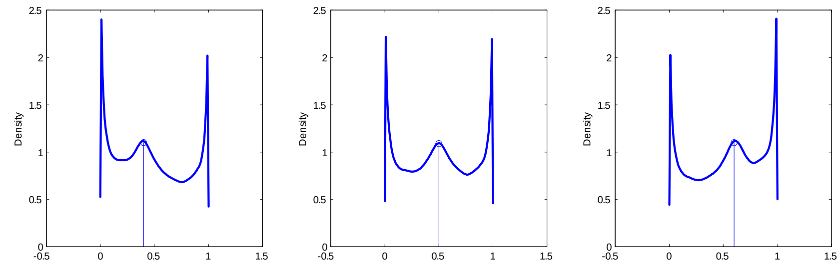

Figure 2: The density of $\widehat{\tau}_{ML}$ given in Equation (14) when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively) and the signal-to-noise ratio (δ/σ) is 1. In each panel, the verticle line represents the true value of τ_0.

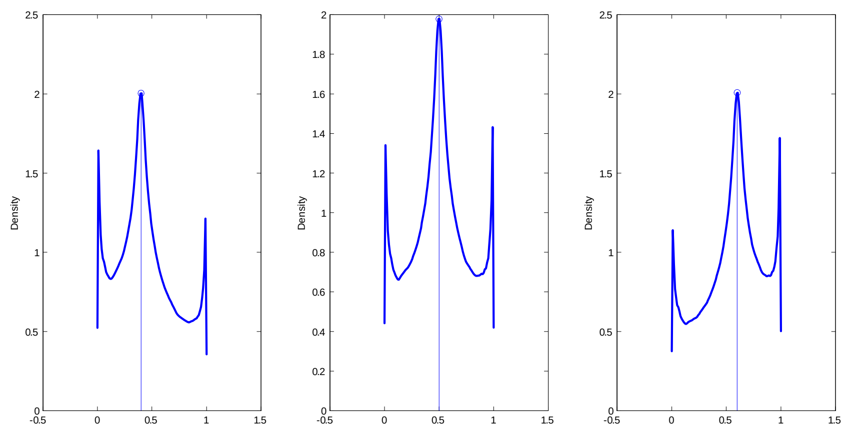

Figure 3: The density of $\widehat{\tau}_{ML}$ given in Equation (14) when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively) and the signal-to-noise ratio (δ/σ) is 2. In each panel, the verticle line represents the true value of τ_0.

signal-to-noise ratio, for example $\delta/\sigma = 1$, the highest mode is not the true value, but the two boundary points when $\tau_0 = 50\%$; it becomes the left (right) boundary point if τ_0 is smaller (larger) than 50%. However, the highest mode moves to the true value when the signal-to-noise ratio increases in all cases with $\delta/\sigma \geq 2$. It is also found that, the mode on the left boundary point is always larger (smaller) than that on the right boundary point when τ_0 is smaller (larger) than 50%. When the signal-to-noise ratio is large enough, tri-modaility becomes unique modality. In this case, the shape of the distribution is similar to that in Figure 1 but is more peaked at the mode.

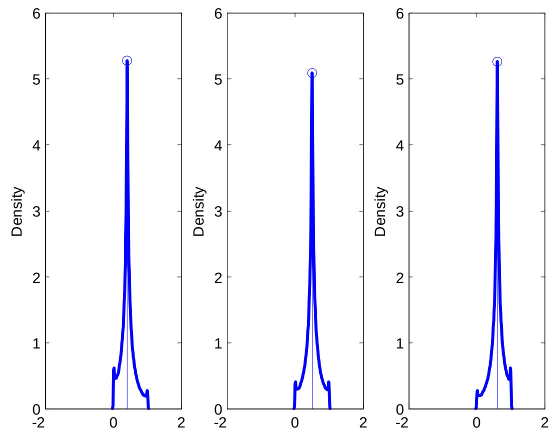

Figure 4: The density of $\widehat{\tau}_{ML}$ given in Equation (14) when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively) and the signal-to-noise ratio (δ/σ) is 4. In each panel, the verticle line represents the true value of τ_0.

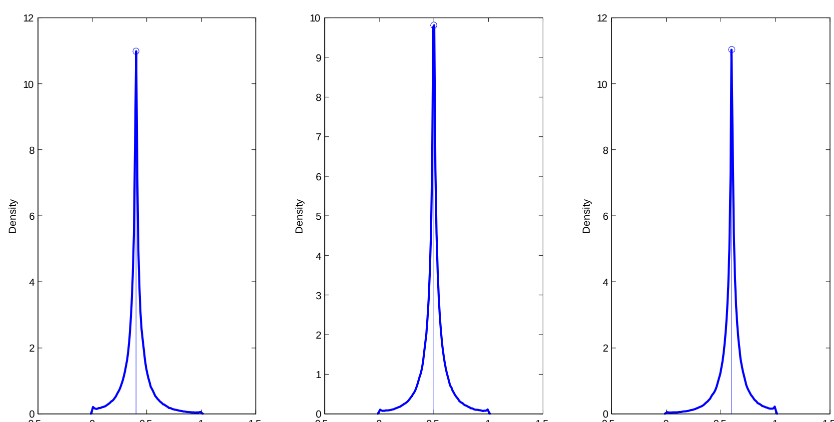

Figure 5: The density of $\widehat{\tau}_{ML}$ given in Equation (14) when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively) and the signal-to-noise ratio (δ/σ) is 6. In each panel, the verticle line represents the true value of τ_0.

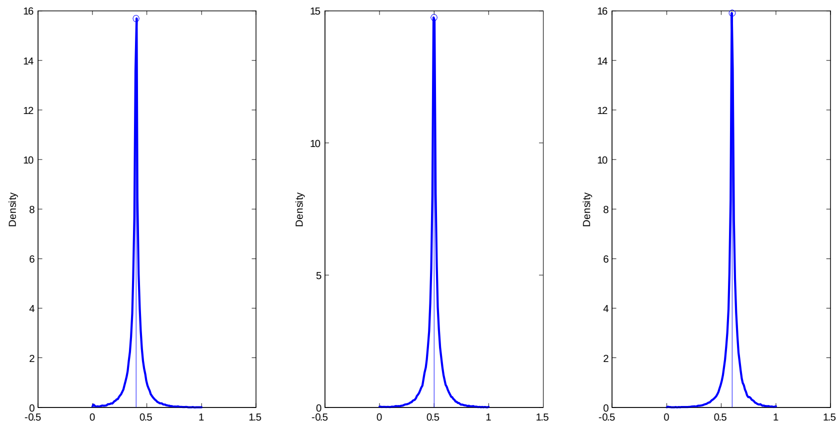

Figure 6: The density of $\widehat{\tau}_{ML}$ given in Equation (14) when $\tau_0 = 0.4, 0.5, 0.6$ (the left, middle and right panel respectively) and the signal-to-noise ratio (δ/σ) is 8. In each panel, the verticle line represents the true value of τ_0.

3 Bias in a Discrete Time Model

As reviewed in Section 2, Hinkley (1970), Yao (1987) and Bai (1994) examined the change-in-mean model in the discrete time context.[2] Since the objective functions are not differentiable with respect to k, it is very difficult to obtain the finite sample distribution in the discrete time model. Yao (1987) and Bai (1994) developed the large sample properties under the additional assumptions about the size of the structural break δ_T. In this Section, we will study the finite sample properties and the bias of $\widehat{\tau}_{ML}$ and $\widehat{k}_{ML}$ in a discrete time model.

Let us start with the continuous time model specified in Equation (10). Splitting the interval $[0, 1]$ into $1/h$ subintervals so that each interval has a size of h, we then get $T = 1/h$ observations of the stochastic process $X(\cdot)$ at T equally spaced points $\{th\}_{t=1}^{T}$, and have the following exact discrete time representation:

$$X_{th} - X_{(t-1)h} = \begin{cases} \mu h + \sqrt{h}\epsilon_{th} & \text{if } \ t = 1, \cdots, \lfloor \tau_0/h \rfloor, \\ (\mu + \delta)h + \sqrt{h}\epsilon_{th} & \text{if } \ t = \lfloor \tau_0/h \rfloor + 1, \cdots, T, \end{cases} \tag{15}$$

where ϵ_{th} ~i.i.d.$N(0, \sigma^2)$, $\lfloor \cdot \rfloor$ is the integer-valued function. Considering that ϵ_{th} is independent of h, we now write it as ϵ_t.

[2] In Bai (1994), ϵ_t can be a linear process satisfying the summability condition. So Bai's model is more general than Hinkley (1970) and Yao (1987)

Letting $Z_t = \left(X_{th} - X_{(t-1)h}\right)/\sqrt{h}$, we obtain

$$Z_t = \begin{cases} \mu\sqrt{h} + \epsilon_t & \text{if } \ t \leq \lfloor \tau_0/h \rfloor, \\ (\mu+\delta)\sqrt{h} + \epsilon_t & \text{if } \ t > \lfloor \tau_0/h \rfloor. \end{cases} \tag{16}$$

Whenever h is fixed, the model in equation (16) is the same as the one in equation (1) with ϵ_t being assumed to follow $N(0, \sigma^2)$, $k_0 = \lfloor \tau_0/h \rfloor$ being the absolute break point.

For the sequential limit distribution of Yao (1987) to be able to provide a good approximation to the finite sample distribution, it is required that the sample size T goes to infinity at a faster rate than that at which the squared structural break size goes to zero. However, in Model (16), when $h \to 0$, the structural break size $\delta\sqrt{h}$ goes to zero at the rate of $1/\sqrt{T}$. Hence, the sample size T does not go to infinity at a faster rate than that at which the squared structural break size goes to zero. As a result, Yao's limit distribution may not well approximate the finite sample distribution in (16) when h is small.

The simultaneously double asymptotic distribution given in Bai (1994) is essentially the same as the sequential limit distribution in Yao (1987). To derive the double asymptotic distribution, Bai (1994) assumed that the magnitude of break size goes to zero at a rate smaller than $\sqrt{\log T}/\sqrt{T}$. However in Model (16), when $h \to 0$, $\delta\sqrt{h} = \delta/\sqrt{T} = O_p\left(1/\sqrt{T}\right)$. This may explain why Bai's limit distribution may not well approximate the finite sample distribution in (16) when h is small.

On the other hand, our exact finite sample distribution in Equation (14) can be regarded as a good approximation to the finite sample distribution of the ML estimator of the break point in model (15) when h is small. It is easy to find that the ML estimator of the break point in model (15) is the same as the one in Model (16). Therefore, the finite sample distribution in Equation (14) could well approximate the finite sample distribution of $\widehat{\tau}_{ML}$ in the discrete time model (16) when h is small. In particular, we expect there is no bias in $\widehat{\tau}_{ML}$ in the discrete time model (16) when $\tau_0 = 50\%$. However, we expect a upward (downward) bias in $\widehat{\tau}_{ML}$ in the discrete time model (16) when τ_0 is smaller (larger) than 50%. Since $\hat{k}_{ML} = \lfloor \widehat{\tau}_{ML} T \rfloor$, we expect the traditional estimator of the absolute break point k in the discrete time model (16) is also asymmetric and has the bias in finite sample. The bias in $\widehat{\tau}_{ML}$ and $\hat{k}_{ML}$ in the discrete time model will be discussed in detail in the next section.

Consider the special case when $\tau_0 = 50\%$. Notice that the fraction of the sample before and after the break is the same in this case. Also note that Equation (14) can be written as

$$\left(\frac{\delta}{\sigma}\right)^2 (\hat{\tau} - \tau_0) \stackrel{d}{=} \arg \max_{u \in \left[-\frac{1}{2}\left(\frac{\delta}{\sigma}\right)^2, \frac{1}{2}\left(\frac{\delta}{\sigma}\right)^2\right]} W(u) - \frac{1}{2}|u|. \tag{17}$$

This result is similar to the limit theory given by Equation (5). Given that δ should be replaced by $\delta\sqrt{h}$ and T should be replaced by $1/h$ in (5), the left hand sight in the two equations are identical. The only difference is on the right hand side. The finite sample theory in the continuous time model is the location of the extremum over a finite interval which depends on the signal-to-noise ratio. The limit distribution in the discrete time model is the location of the extremum over an infinite interval. As a result, we expect the finite sample distribution be closer to the limit distribution when the signal-to-noise ratio is large. This expectation can be confirmed by the middle panels in Figures 2-6.

4 Bias Correction via Indirect Estimation

The indirect estimation is a simulation-based method, first introduced by Smith (1993), Gourieroux 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 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.

Gourieroux 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 (1997) and Gourieroux 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 primitive conditions for finite sample properties of the indirect estimator and also introduced an iterative procedure to further improve the performance of the indirect estimator. The indirect estimation obtains the bias function by simulating from the true model and hence the auxiliary model. In this section, we apply the 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 (14) and (17), we know that the bias formula and the bias expansion are too difficult to deal with explicitly. The same idea was used to estimate continuous time models in Phillips and Yu (2009) and dynamic panel models in Gourieroux et al. (2010).

The application of the indirect estimation procedure for estimating structural break proceeds as follows. Given a parameter θ (either τ or k), 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 (10) or (1), 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

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

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

$$\hat{\theta}_T^{IE} = \arg\max_{\theta\in\Theta} \left\| \hat{\theta}_T - b_T(\theta) \right\|$$

where $b_T(\theta) = E(\tilde{\theta}_T^h(\theta)$ is the binding (or bias) function. If $b_T(\theta)$ is invertible, then the indirect estimator can be directly written as

$$\hat{\theta}_T^{IE} = b_T^{-1}(\hat{\theta}_T).$$

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 (10) or the discrete time model given by (1). At first, we employ the LS method of Bai (1994) or the ML method to the actual data in order to obtain $\hat{k}_T$. Then the corresponding estimator for the hth simulated path is $\tilde{k}_T^h(k)$ and the indirect estimation estimator is

$$\hat{k}_T^{IE} = \arg\max_{k\in\Theta} \left\| \hat{k}_T - b_T(k) \right\|,$$

where $\hat{k}_T$ is the original estimator of k from the actual data that has T observations, $b_T(k)$ is the binding function with the form

$$b_T(k) = E(\tilde{k}_T^h(k)),$$

which, in practice, can be effectively replaced by $\frac{1}{H}\sum_{h=1}^{H} \tilde{k}_T^h(k)$ since H can be chosen arbitrarily large. If the binding function is invertible, then

$$\hat{k}_T^{IE} = b_T^{-1}\left(\hat{k}_T\right). \tag{18}$$

Based on $\hat{k}_T^{IE}$, we can define the indirect estimator of the fractional break point as $\widehat{\tau}_T^{IE} = \hat{k}_T^{IE}/T$. Let the corresponding binding function be $b_T(\tau) = b_T(k)/T$. If $b_T(k)$ is invertible, $b_T(\tau)$ is also invertible. Hence,

$$\widehat{\tau}_T^{IE} = b_T^{-1}\left(\widehat{\tau}_T\right), \tag{19}$$

where $\widehat{\tau}_T$ is the original estimator of τ from the actual data.

Following the discussion of the finite sample property in Gourieroux et al. (2000) and Phillips (2012), we impose the following assumption.

Assumption 1. The binding function $b_T(\tau)$, mapping from $(0, 1)$ to $b_T(0, 1)$, is uniformly continuous and one-to-one.

Under Assumption 1, the binding function $b_T(\cdot)$ is invertible. We have $\widehat{\tau}_T^{IE}$ is "b_T-mean-unbiased", since

$$E\left(b_T\left(\widehat{\tau}_T^{IE}\right)\right) = E\left(\widehat{\tau}_T\right) = E(\widetilde{\tau}_T^h(\tau_0)) = b_T(\tau_0),$$

and

$$b_T^{-1}\left(E\left(b_T\left(\widehat{\tau}_T^{IE}\right)\right)\right) = \tau_0. \tag{20}$$

By the same reason, $\hat{k}_T^{IE}$ is also "b_T-mean-unbiased", i.e., $b_T^{-1}\left(E\left(b_T\left(\widehat{k}_T^{IE}\right)\right)\right) = k_0$.

Moreover, when $b_T(\cdot)$ is linear, the indirect estimator of τ and k is exactly mean-unbiased since, in (20), we have

$$b_T^{-1}\left(E\left(b_T(\widehat{\tau}_T^{IE})\right)\right) = E\left(b_T^{-1}\left(b_T(\widehat{\tau}_T^{IE})\right)\right) = E\left(\widehat{\tau}_T^{IE}\right) = \tau_0,$$

which is a especially appealing property in the practice when the binding function is close to linear.

It is important to point out the indirect estimator shares the same consistency property as the original estimator. Since only $\widehat{\tau}_T$ is consistent, hence we can only ensure the consistency of $\widehat{\tau}_T^{IE}$ but not $\widehat{k}_T^{IE}$.

Regarding the efficiency, from Equation (19) and by the Delta method, we have

$$\mathrm{Var}(\widehat{\tau}_T^{IE}) \approx \left(\frac{\partial b_T(\tau_0)}{\partial \tau}\right)^{-2} \mathrm{Var}(\widehat{\tau}_T). \tag{21}$$

Hence, the efficiency loss (or gain) is measured by $\frac{\partial b_T(\tau_0)}{\partial \tau}$. If $\left|\frac{\partial b_T(\tau_0)}{\partial \tau}\right| < 1$, $\widehat{\tau}_T^{IE}$ has a bigger variance than $\widehat{\tau}_T$. However, if $\left|\frac{\partial b_T(\phi_0)}{\partial \phi}\right| > 1$, $\widehat{\tau}_T^{IE}$ will have a small variance than $\widehat{\tau}_T$. If the finite sample distribution developed in Section 2 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 estimation will be larger than that of the original estimator.

5 Monte Carlo Results

In this section, we design two Monte Carlo experiments to examine the bias in the LS estimator of k in the discrete time model (1) and the ML estimator of τ in the continuous time model (10), and to compare the finite sample performance of the indirect estimator and the original estimators. When inverting the binding function, following Phillips and Yu (2009), we choose a set of grid points for τ, namely, $\tau = [0.1, 0.11, ..., 0.89, 0.9]$ and calculate $b_T(\tau)$ for each τ via simulations. We then use the standard linear interpolation and extrapolation methods to obtain the binding functions in the domain $[0, 1]$.[3]

In the first experiment, data are generated from Model (10), with $\sigma = 1$, $\delta = 2, 4, 6$, $\tau_0 = 30\%, 50\%, 70\%$, $dB(t) \sim iid\ N(0, h)$, where $h = \frac{1}{1000}$. For each combination of δ and τ_0, we obtain the ML estimator of τ from Equation (12) and the indirect estimator. Our focus is to examine the finite sample properties of $\hat{\tau}$, so it is assumed that the structural break size δ and the standard deviation σ are known during the simulation.

Table 1 reports the bias, the standard error, and the root mean squared errors (RMSE) of the ML estimate and the indirect estimate of τ, obtained from 10,000

[3] However, if the indirect estimator of τ takes a value outside of the interval $[0, 1]$ for one particular replication, such a replication is discarded for both ML and the indirect estimation.

Table 1: Monte Carlo comparison of the bias and RMSE of ML and Indirect Estimates. The number of simulated path is set to be 10,000 for indirect estimation. The number of replications is set at 10,000.

Case		Bias		Standard Error		RMSE	
$\frac{\delta}{\sigma}$	τ_0	ML	IE	ML	IE	ML	IE
2	0.3	0.1337	0.0736	0.1408	0.2688	0.1942	0.2787
2	0.5	-0.0016	-0.0025	0.1268	0.2407	0.1268	0.2407
2	0.7	-0.1323	-0.0712	0.1400	0.2669	0.1926	0.2762
4	0.3	0.0518	0.0222	0.1543	0.1870	0.1628	0.1883
4	0.5	0.0021	0.0029	0.1511	0.1820	0.1511	0.1820
4	0.7	-0.0435	-0.0137	0.1479	0.1787	0.1542	0.1792
6	0.3	0.0118	0.0037	0.1100	0.1163	0.1106	0.1164
6	0.5	0.0004	-0.0003	0.1172	0.1228	0.1172	0.1228
6	0.7	-0.0104	-0.0027	0.1092	0.1156	0.1097	0.1156

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 $\delta/\sigma = 2$, the bias is 0.1337 and about 45% 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 $\frac{\delta}{\sigma} = 2$ and $\tau_0 = 70\%$, the indirect estimation method removes about half 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.

In the second experiment, data are generated from Model (1), with $\sigma = 1$, $\delta = 0.5, 1$, $\tau_0 = 0.3, 0.5, 0.7$, $\epsilon_t \sim iid\ N(0,1)$, where we choose $T = 50, 80, 100, 120$. 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 and the standard deviation σ are known. The reason we focus on k is because k is a practically important parameter to estimate.

Table 2 reports the bias, the standard error, and the root mean squared errors (RMSE) of the ML estimate and the indirect estimator of k, obtained from 10,000 replications. We may draw the following conclusions from Table 2. First, 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 $T = 50, \tau_0 = 30\%$ and $\delta/\sigma = 0.5$, the bias is nearly 9 while the true value of k is 15. The bias is about

Table 2: Monte Carlo comparison of the bias and RMSE of LS and Indirect Estimates. The number of simulated path is set to be 10,000 for indirect estimation. The number of replications is set at 10,000.

Case				Bias		Standard error		RMSE	
T	$\frac{\delta}{\sigma}$	τ_0	k_0	LS	IE	LS	IE	LS	IE
50	0.5	0.3	15	8.9750	6.8050	3.7450	11.6250	9.7250	13.4703
50	0.5	0.5	25	0.0250	-0.0300	3.0950	9.3150	3.0951	9.3150
50	0.5	0.7	35	-8.8650	-6.4750	3.7400	12.0500	9.6216	13.6795
50	1	0.3	15	1.4150	-0.8200	5.0550	6.8500	5.2493	6.8989
50	1	0.5	25	-0.1050	-0.1500	4.5900	5.8700	4.5912	5.8719
50	1	0.7	35	-1.6450	0.4500	5.0950	6.9350	5.3540	6.9496
80	0.5	0.3	24	11.728	5.472	7.544	17.88	13.9448	18.6986
80	0.5	0.5	40	-0.016	-0.632	5.912	12.832	5.9120	12.8476
80	0.5	0.7	56	-12.088	-7.592	5.4432	18.256	13.2570	19.7717
80	1	0.3	24	0.936	-0.352	6.752	7.68	6.8166	7.6881
80	1	0.5	40	-0.008	-0.024	6.2	6.792	6.2000	6.7920
80	1	0.7	56	-0.944	0.208	6.976	7.976	7.0396	7.9787
100	0.5	0.3	30	12.83	4.36	10.66	23.20	16.6807	23.6061
100	0.5	0.5	50	0.35	0.26	8.02	15.13	8.0276	15.1322
100	0.5	0.7	70	-9.22	2.01	10.21	22.02	13.7569	22.1115
100	1	0.3	30	0.72	-0.11	7.28	7.79	7.3155	7.7908
100	1	0.5	50	0.06	0.02	6.49	6.80	6.4903	6.8000
100	1	0.7	70	-0.82	0.09	7.53	8.11	7.5745	8.1105
120	0.5	0.3	36	6.636	-4.724	14.724	24.3	16.1503	24.7549
120	0.5	0.5	60	-0.096	0.252	12.792	20.388	12.7924	20.3896
120	0.5	0.7	84	-6.936	3.816	14.82	24.66	16.3628	24.9535
120	1	0.3	36	0.588	-0.096	7.308	7.656	7.3316	7.6566
120	1	0.5	60	0	-0.024	6.756	6.984	6.7560	6.9840
120	1	0.7	84	-0.504	0.108	7.176	7.524	7.1937	7.5248

60% 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 $T = 80$, $\frac{\delta}{\sigma} = 0.5$ and $\tau_0 = 30\%$, the indirect estimation method removes more than half 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.

To understand why the indirect estimation increases the variance, we plot the binding functions in these two models in Figure 7 and Figure 8, where we also plot the 45 degrees line for the purpose of comparison. Figure 7 corresponds to the continuous

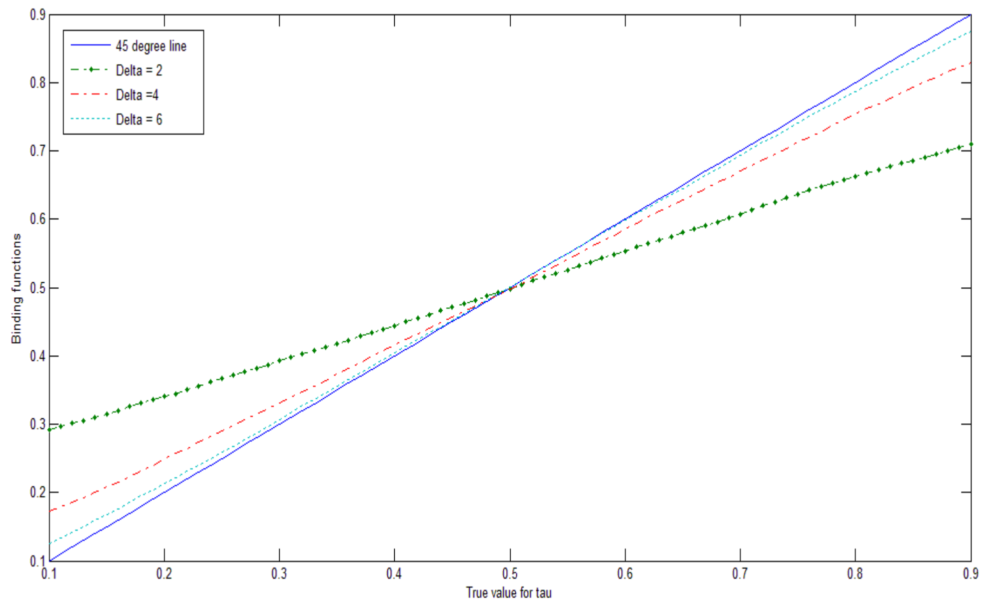

Figure 7: Binding function of ML for the continuous time model when $h = 0.001$

time model with $\delta = 2, 4, 6$ and Figure 8 to the discrete time model with $T = 100$, $\delta = 0.5, 1$. Several conclusions can be made. Firstly, 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 monotonically increase as τ or k increases, suggesting that the binding functions are invertible. However, in all cases, the binding functions are flatter than the 45 degrees line, explaining why the variance of the indirect estimate is larger than that of the 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 boundaries when $\delta = 0.5$ in the discrete time model. Due to the presence of nonlinearity, the indirect estimation procedure cannot completely remove the bias, although it is "b_T-mean-unbiased".

6 Conclusions

This paper is concerned about the finite sample properties in the estimation of structural break points. We find that the finite sample bias is substantial in many practically relevant situations. While the literature on structural break has focused the a great deal

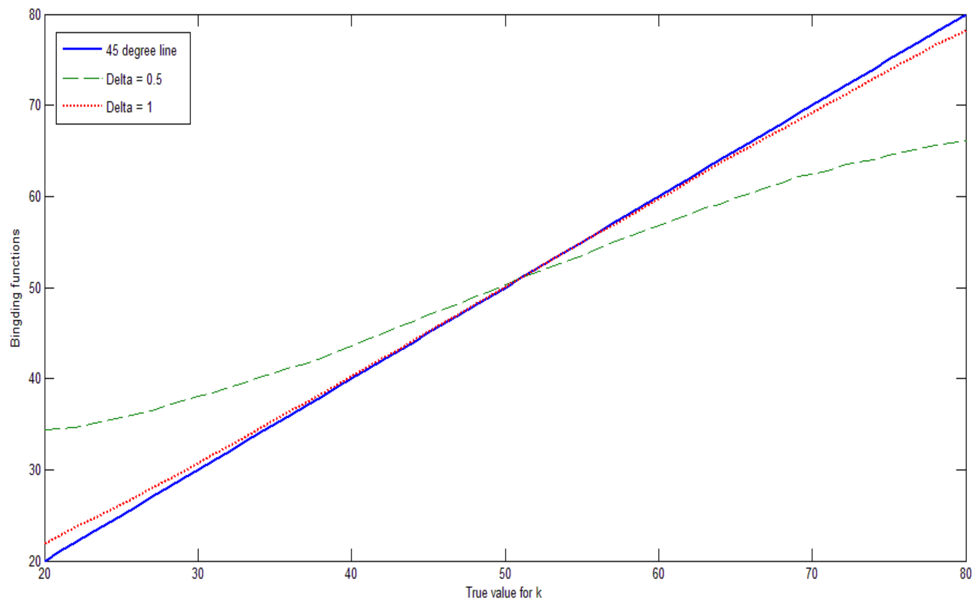

Figure 8: Binding function of LS for the discrete time model when $T = 100$

of attention to develop asymptotic properties, the finite sample problem has received no attention in this literature, to the best of our knowledge. We hope to fill up this important gap in the literature.

In this paper we address the finite sample problem in several aspects. First we derive the finite sample distribution of the structural break estimator in the continuous time model. We then establish its connection to the discrete time models considered in the literature. It is shown that when the true break point is at the middle of the sample, the finite sample distribution is symmetric but can have tri-modality. However, when the true break point occurs earlier than the middle of the sample, the finite sample distribution is skewed to the right and there is a positive bias. When the true break point occurs later than the middle of the sample, the finite sample distribution is skewed to the left and there is a negative bias.

To reduce the bias in finite sample, we obtain the binding functions via simulations and then use the indirect estimation technique to estimate the break parameter. Indirect estimation essentially inverts the binding function at the original estimator obtained from the actual data. It 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 of the traditional break point estimators.

The models considered in this paper are very simply 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 the finite sample distribution for more realistic models and real time estimators are needed. How to extend the indirect estimation technique in a multiple parameter settings are also useful.

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

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

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

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

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

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

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

[9] 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.

[10] Bao, Y., and A. Ullah, 2007, The second-order bias and mean squared error of estimators in time-series models, Journal of Econometrics 140(2), 650-669.

[11] 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.

[12] Bhattacharya, P. K., 1994. Some aspects of change-point analysis. Lecture Notes-Monograph Series, 28-56.

[13] Bhattacharya, P. K, .and Brockwell, P. J., 1976. The minimum of an additive process with applications to signal estimation and storage theory. Z. Wahrsch. verw. Gebiete 37, 51-75.

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

[15] Chernoff, H., and S. Zacks, 1964, Estimating the Current Mean of a Normal Distribution which is Subjected to Changes in Time Series, The Annals of Mathematical Statistic, 35(3), 999-1018.

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

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

[18] Gourieroux, 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] 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.

[23] Hinkley, D. V., 1969. Inference about intersection in the two-phase regression problem. Biometrika 56, 495-504.

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

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

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

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

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

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

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

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

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

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

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

[35] Rilstone, P., V.K. Srivastava and A. Ullah, 1996, The second order bias and mean squared error of nonlinear estimators, Journal of Econometrics 72(2), 369–395.

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

[37] 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.

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