Journal of Econometrics 205 (2018) 156–176

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

New distribution theory for the estimation of structural break point in mean[☆]

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

[a] Fanhai International School of Finance and School of Economics, Fudan University, China
[b] Chinese University of Hong Kong, Hong Kong
[c] School of Economics and Lee Kong Chian School of Business, Singapore Management University, Singapore

ARTICLE INFO

Article history:
Available online 26 March 2018

JEL classification:
C11
C46

Keywords:
Structural break
Bias reduction
Indirect estimation
Exact distribution
In-fill asymptotics

ABSTRACT

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

1. Introduction

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

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

☆ An earlier draft of this paper was circulated under the title "On Bias in the Estimation of Structural Break Points". We would like to thank the editors, three referees, Peter Phillips, Graham Elliott, Fa Wang, and participants of many conferences and seminars, including Princeton-QUT-SJTU-SMU workshop on Frontiers in Econometrics held at SMU on April 18–19, 2015 and the SETA conference held at Hitotsubashi University on May 30–31, 2015, 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.
E-mail address: yujun@smu.edu.sg (J. Yu).

https://doi.org/10.1016/j.jeconom.2018.03.009

This paper makes several contributions to the literature. First, by using the Girsanov theorem, we develop the exact distribution of the maximum likelihood (ML) estimator of the structural break point in a continuous time model, assuming that a continuous record over a finite time span is available. It is shown that the exact distribution is asymmetric when the true break point is not in the middle of the sample. Moreover, the exact distribution has trimodality when the signal-to-noise ratio (the break size over the standard deviation of the error term) is not too large, regardless of the location of the true break point. Asymmetry together with trimodality makes the ML estimator biased and suggests that the long-span asymptotic distribution does not conform to the exact distribution. It is also found that upward (downward) bias is obtained when the fractional structural break point is smaller (larger) than 1/2, and the further the fractional structural break point away from 1/2, the larger the bias.

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

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

The in-fill asymptotic treatment is not new in the literature.[1] Recently, Yu (2014) and Zhou and Yu (2015) demonstrated that the in-fill asymptotic distribution provides better approximations to the finite sample distribution than the long-span asymptotic distribution in persistent autoregressive models. In the same spirit, we show in the paper that the in-fill asymptotic distribution for the LS estimator of the structural break point conforms better to the finite sample distribution than the long-span counterpart. Interestingly, our in-fill asymptotic distribution is the same as that of Elliott and Müller (2007) although Elliott and Müller mainly focused their attention on the problem of constructing confidence sets for the break point.

The rest of the paper is organized as follows. Section 2 gives a brief review of the literature and provides the motivations of the paper. Section 3 develops the exact distribution of the ML estimator of structural break point in a continuous time model. Section 4 establishes a continuous time approximation to the discrete time model previously considered in the literature and develops the in-fill asymptotic theory for the LS estimator under different settings. The IE procedure and its applications in the continuous time and the discrete time models with a structural break are introduced in Section 5. In Section 6, we provide simulation results and compare the finite sample performance of IE with that of the traditional estimation methods and of other simulation-based methods. Section 7 concludes. All proofs are contained in the Appendix.

2. Literature review and motivations

The literature on estimating structural break points is too extensive to review. A partial list of contributions include Hinkley (1970), Hawkins et al. (1986), Yao (1987), Bai (1994, 1995, 1997a, b), 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

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

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

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

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