# Structural change in non-stationary AR(1) models

Chong, Terence Tai Leung and Pang, Tianxiao and Zhang, Danna and Liang, Yanling

The Chinese University of Hong Kong and Nanjing University, Zhejiang University, University of Chicago, Zhejiang University

2017

# Structural change in non-stationary AR(1) models*

April 25, 2017

Tianxiao Pang[1,†], Terence Tai-Leung Chong[2,3], Danna Zhang[4] and Yanling Liang[1],

[1]School of Mathematical Sciences, Yuquan Campus, Zhejiang University, Hangzhou 310027, P.R.China

[2]Department of Economics and Lau Chor Tak Institute of Global Economics and Finance,

The Chinese University of Hong Kong, Hong Kong

[3]Department of International Economics and Trade, Nanjing University, Nanjing 210093, P.R.China

[4]Department of Statistics, University of Chicago, 5734 S. University Ave., Chicago, Illinois 60637, USA

**Abstract.** This paper revisits the asymptotic inference for non-stationary AR(1) models of Phillips and Magdalinos (2007a) by incorporating a structural change in the AR parameter at an unknown time $k_0$. Consider the model $y_t = \beta_1 y_{t-1} I\{t \leq k_0\} + \beta_2 y_{t-1} I\{t > k_0\} + \varepsilon_t,\ t = 1, 2, \cdots, T$, where $I\{\cdot\}$ denotes the indicator function, one of $\beta_1$ and $\beta_2$ depends on the sample size $T$, and the other is equal to one. We examine four cases: Case (I): $\beta_1 = \beta_{1T} = 1 - c/k_T$, $\beta_2 = 1$; (II): $\beta_1 = 1$, $\beta_2 = \beta_{2T} = 1 - c/k_T$; (III): $\beta_1 = 1$, $\beta_2 = \beta_{2T} = 1 + c/k_T$; and case (IV): $\beta_1 = \beta_{1T} = 1 + c/k_T$, $\beta_2 = 1$, where $c$ is a fixed positive constant, and $k_T$ is a sequence of positive constants increasing to $\infty$ such that $k_T = o(T)$. We derive the limiting distributions of the $t$-ratios of $\beta_1$ and $\beta_2$ and the least squares estimator of the change point for the cases above under some mild conditions. Monte Carlo simulations are conducted to examine the finite-sample properties of the estimators. Our theoretical findings are supported by the Monte Carlo simulations.

**Keywords:** AR(1) model, Least squares estimator, Limiting distribution, Mildly explosive, Mildly integrated, Structural change, Unit root.

---

*Tianxiao Pang and Yanling Liang's research was supported by the Department of Education of Zhejiang Province in China (N20140202).

E-mail addresses: txpang@zju.edu.cn, chong2064@cuhk.edu.hk, zhangdanna0507@gmail.com and 1040857986@qq.com.

†Corresponding author.

JEL Classification: C22

# 1 Introduction and Main Results

The change-point problem in time series regression models has received considerable attention in the literature over the past decades. Many economic time series data are characterized by single or multiple structural changes (Stock and Watson (1996, 1999), Hansen (2001)). Bai and Perron (1998) provided the estimation and test procedures for linear models with multiple structural changes. Leybourne et al. (2003), Harvey et al. (2006), Halunga and Osborn (2012) and Kejriwal et al. (2013) investigated structural changes in persistence. Chong (2003), Pitarakis (2004) and Bai et al. (2008) studied the estimation and tests of the change point under model misspecification. Qu (2008) tested for the structural change in regression quantiles. Recent development in this area includes that of Lee et al. (2016), who investigated the change-point problem in high-dimensional regression models.

An important strand of literature on structural change focuses on autoregressive models. Structural changes in autoregressive models are of interest as the time series properties of the model, such as stationarity, may be different before and after the change. As a result, the rates of convergence and the asymptotic distributions of the estimators are difficult to derive. Chong (2001) investigated the statistical inference for the change point in various AR(1) models. Berkes et al. (2011) studied the structural change from an AR(1) model to a threshold AR(1) model. An important application of the AR(1) change-point model was given by Mankiw and Miron (1986) and Mankiw et al. (1987), who found that the short-term interest rate has changed from a stationary process to a near random walk since the Federal Reserve System was founded at the end of 1914. Other applications can be found in Barsky (1987) and Burdekin and Siklos (1999) for inflation rate series, in Hakkio and Rush (1991) for government budget deficits, in Phillips et al. (2011) for 1990's NASDAQ stock prices and in Phillips and Yu (2011), Phillips et al. (2015a, 2015b, 2015c) and Phillips and Shi (2017) for financial bubbles and collapses.

This paper revisits and generalizes the model of Chong (2001). Consider the following AR(1) model with a change in the AR parameter at an unknown time $k_0$,

$$y_t = \beta_1 y_{t-1} I\{t \le k_0\} + \beta_2 y_{t-1} I\{t > k_0\} + \varepsilon_t, \qquad t = 1, 2, \cdots, T, \tag{1.1}$$

where $I\{\cdot\}$ denotes the indicator function, and $\{\varepsilon_t, t \ge 1\}$ is a sequence of i.i.d. random variables. Under the regularity conditions that $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) < \infty$ and $E(y_0^2) < \infty$, the consistency and limiting distributions of the least squares estimators (LSE) of fixed $\beta_1$ and $\beta_2$ and the change-point estimator of $\tau_0(= k_0/T)$ were developed in Chong (2001) for the following three cases: (1) $|\beta_1| < 1$ and $|\beta_2| < 1$, (2) $|\beta_1| < 1$ and $\beta_2 = 1$, and (3) $\beta_1 = 1$ and $|\beta_2| < 1$.

Since heavy-tailed distributions such as the Student's $t$-distribution with degrees of freedom 2, Pareto distribution with index 2 and stable random variables are often used to model asset returns in empirical studies (Mandelbrot (1963)), Pang and Zhang (2015) employed the truncation technique to weaken the moment conditions of $y_0$ and $\varepsilon_t$'s in case (1) of Chong (2001). For any constant $c$, Pang et al. (2014) examined the asymptotics for the case where $|\beta_1| < 1$ and $\beta_2 = \beta_{2T} = 1 - c/T$ as well as the case where $\beta_1 = \beta_{1T} = 1 - c/T$ and $|\beta_2| < 1$. In these two cases, one of $\beta_1$ and $\beta_2$ is fixed and smaller than one in absolute value, while the other is local to unity. The limiting distributions obtained in Pang et al. (2014) are complicated. In the special case of $c = 0$, the results in Pang et al. (2014) are reduced to the two main theorems in Chong (2001). Both Pang et al. (2014) and Pang and Zhang (2015) only require $\varepsilon_t$'s to be in the domain of asymptotic normality (DAN) with zero mean and possibly infinite variance, and $y_0$ to be any random variable of an order smaller than $\sqrt{T}$ in probability.

The asymptotic theory for the near unit root model was first studied by Phillips (1987) and Chan and Wei (1987) independently. Their studies bridge the gap between the stationary AR(1) model and the unit root model. This paper is related to that of Phillips and Magdalinos (2007a), who attempted to bridge the gap between the asymptotic theories of the stationary, the explosive and the local-to-unity AR(1) models. In particular, they investigated the limiting distribution of the LSE of the AR parameter for the following AR(1) model: $y_t = \beta y_{t-1} + \varepsilon_t,\ t = 1, 2, \cdots, T$, with $\beta = \beta_T = 1 - c/k_T$ and $\beta = \beta_T = 1 + c/k_T$, where $c$ is a positive constant, and $k_T$ is a sequence of positive constants increasing to $\infty$ such that $k_T = o(T)$. They proved the asymptotic normality for the LSE of $\beta$ with convergence rate $\sqrt{Tk_T}$ for the mildly integrated AR(1) model when $\beta = \beta_T = 1 - c/k_T$ and showed a Cauchy limiting distribution for the LSE of $\beta$ with convergence rate $k_T\beta_T^T$ for the mildly explosive AR(1) model when $\beta = \beta_T = 1 + c/k_T$. The mildly explosive and the integrated AR(1) models have been widely used to model financial bubbles and collapses, respectively. The reader is referred to Phillips and Yu (2011), Phillips et al. (2011), Phillips et al. (2015a, 2015b, 2015c) and Phillips and Shi (2017) for more details.

Motivated by the works of Chong (2001) and Phillips and Magdalinos (2007a) and the importance of the mildly integrated and explosive AR(1) models in applications, we aim to study in this article the structural change in mildly integrated and mildly explosive AR(1) models from the theoretical perspective. In particular, we are interested in the following four cases: (I) $\beta_1 = \beta_{1T} = 1 - c/k_T$, $\beta_2 = 1$; (II) $\beta_1 = 1$, $\beta_2 = \beta_{2T} = 1 - c/k_T$; (III) $\beta_1 = 1$, $\beta_2 = \beta_{2T} = 1 + c/k_T$; and (IV) $\beta_1 = \beta_{1T} = 1 + c/k_T$, $\beta_2 = 1$, where $c > 0$, and $k_T$ shares the same assumption in Phillips and Magdalinos (2007a). Since the models before and after the time $k_0$ are either non-stationary or nearly non-stationary, and the difference between $\beta_1$ and $\beta_2$ converges to zero as the sample size tends to infinity, obtaining the closed-form limiting distributions of the LSEs of $\beta_1$ and $\beta_2$ and the estimator of the change point will be a challenging task. The main contribution of this paper is to derive the closed-form solution of the limiting distributions of these estimators for the cases above when the distribution of the error term belongs to the DAN.*

A sequence of i.i.d. random variables $\{X_i, i \geq 1\}$ belongs to the DAN if there exist two constant sequences $\{A_n, n \geq 1\}$ and $\{B_n, n \geq 1\}$ such that $Z_n := B_n^{-1}(X_1 + \cdots + X_n) - A_n$ converges to a standard normal random variable in distribution as $n$ tends to infinity (Feller (1971)), where $B_n$ takes the form $\sqrt{n}h(n)$, and $h(n)$ is a slowly varying function at infinity.

For the models studied in this article, we make the following assumptions:

- C1: $\{\varepsilon_t, t \geq 1\}$ is a sequence of i.i.d. random variables which are in the DAN with zero mean and possibly infinite variance.
- C2: $\{k_T, t \geq 1\}$ is a sequence of positive constants increasing to infinity slowly such that $k_T = o(T)$.
- C3: $y_0$ is an arbitrary random variable satisfying $y_0 = o_p(\sqrt{T})$ when $\beta_1 = 1$ and $y_0 = o_p(\sqrt{k_T})$ when $\beta_1 = 1 - c/k_T$ or $\beta_1 = 1 + c/k_T$.
- C4: $\tau_0 \in [\underline{\tau}, \overline{\tau}] \subset (0, 1)$.

**Remark 1.1** *The assumption of i.i.d. in C1 is only for convenience of exposition in the proofs. If the DAN condition is replaced by some appropriate moment conditions, one can extend our results to allow for dependence. One may refer to Phillips and Magdalinos (2007b) and Magdalinos (2012) for details. Assumption C2 is the same as that in Phillips and Magdalinos (2007a). Assumption C3 states that $y_0$ will not affect the asymptotic properties of the estimators of $\beta_1$, $\beta_2$ and the change point. Assumption C4 is standard in the*

*We present the limiting distributions of the $t$-ratios instead of the LSEs for the AR parameters. The latter can be derived in a similar fashion.

*change-point literature (Bai (1997), Chong (2001) and Pitarakis (2004)), which ensures the identifiability of the AR parameters and the change point. In empirical studies, one can set* $[\underline{\tau}, \overline{\tau}] = [0.05, 0.95]$.

Let $[a]$ denote the integer part of $a$. For any given $0 < \tau < 1$, the LSEs of the AR parameters $\beta_1$ and $\beta_2$ are given by

$$\hat{\beta}_1(\tau) = \frac{\sum_{t=1}^{[\tau T]} y_t y_{t-1}}{\sum_{t=1}^{[\tau T]} y_{t-1}^2} \quad \text{and} \quad \hat{\beta}_2(\tau) = \frac{\sum_{t=[\tau T]+1}^{T} y_t y_{t-1}}{\sum_{t=[\tau T]+1}^{T} y_{t-1}^2}.$$

The change-point estimator is defined as

$$\hat{\tau}_T = \arg\min_{\tau \in (0,1)} RSS_T(\tau),$$

where

$$RSS_T(\tau) = \sum_{t=1}^{[\tau T]} \Big(y_t - \hat{\beta}_1(\tau) y_{t-1}\Big)^2 + \sum_{t=[\tau T]+1}^{T} \Big(y_t - \hat{\beta}_2(\tau) y_{t-1}\Big)^2.$$

Once we obtain the change-point estimator $\hat{\tau}_T$, the final LSEs of $\beta_1$ and $\beta_2$ are defined by $\hat{\beta}_1(\hat{\tau}_T)$ and $\hat{\beta}_2(\hat{\tau}_T)$ respectively.

We define some notations before proceeding to our main results. Let $W(\cdot)$, $\overline{W}(\cdot)$ and $\widetilde{W}(\cdot)$ be three independent standard Brownian motions defined on $[0,1]$, $[0,1]$ and $R_+$ respectively; and $W_1(\cdot)$ and $W_2(\cdot)$ be two independent Brownian motions defined on $R_+$. "$\Rightarrow$" denotes the weak convergence of the associated probability measures. "$\xrightarrow{p}$" denotes convergence in probability and "$\stackrel{d}{=}$" means being identical in distribution. The limits in this paper are all taken as $T \to \infty$ unless specified otherwise. We denote $\hat{k} = [T\hat{\tau}_T]$, and the notation $a_T \asymp b_T$ means there exist two positive constants $c_1$ and $c_2$ such that $c_1 \leq a_T/b_T \leq c_2$ for all large $T$, where $a_T$ and $b_T$ are two positive functions of $T$. In addition, in order to deal with possibly heavy-tailed distributions, we employ the following truncation technique letting

$$\begin{cases} l(u) = E(\varepsilon_1^2 I\{|\varepsilon_1| \leq u\}), \quad b = \inf\{u \geq 1 : l(u) > 0\}, \\ \eta_j = \inf\{s : s \geq b+1, \dfrac{l(s)}{s^2} \leq \dfrac{1}{j}\}, \quad \text{for } j = 1, 2, 3, \cdots. \end{cases} \tag{1.2}$$

When $\varepsilon_1$ belongs to the DAN, $l(u)$ is a slowly varying function approaching a constant (when $\varepsilon_1$ has finite variance) or infinity (when $\varepsilon_1$ has infinite variance) as $u$ tends to infinity. Finally, we denote

$$t_1 = \sqrt{\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{l(\eta_T)}}(\hat{\beta}_1(\hat{\tau}_T) - \beta_1) \quad \text{and} \quad t_2 = \sqrt{\frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{l(\eta_T)}}(\hat{\beta}_2(\hat{\tau}_T) - \beta_2)$$

as the $t$-ratios of $\beta_1$ and $\beta_2$ respectively.

Under assumptions C1-C4, we have the following results.

**Theorem 1.1** *In Model (1.1), if $\beta_1 = \beta_{1T} = 1 - c/k_T$ and $\beta_2 = 1$, where $c$ is a fixed positive constant, under assumptions C1-C4, we have*

*(a) $\hat{\tau}_T$ is consistent, and its limiting distribution is given by*

$$\frac{cT}{k_T}(\hat{\tau}_T - \tau_0) \Rightarrow \arg\max_{\nu \in R} \left\{ \frac{C^*(\nu)}{B_c(\frac{1}{2})} - \frac{|\nu|}{2} \right\}, \tag{1.3}$$

*where $B_c(\frac{1}{2}) = \sqrt{c}\int_0^\infty \exp(-cs)dW_1(s)$, and $C^*(\nu)$ is defined to be $C^*(\nu) = W_1(-\nu)$ for $\nu \leq 0$ and*

$$C^*(\nu) = -W_2(\nu) - \int_0^\nu \frac{W_2(s)}{B_c(\frac{1}{2})} dW_2(s) - \int_0^\nu \Big(\frac{W_2(s)}{2B_c(\frac{1}{2})} + 1\Big) W_2(s) ds$$

*for $\nu > 0$.*

*(b) $\hat{\beta}_1(\hat{\tau}_T)$ is consistent, and its limiting distribution is given by*

$$t_1 \Rightarrow N(0, 1). \tag{1.4}$$

*(c) $\hat{\beta}_2(\hat{\tau}_T)$ is also consistent, and its limiting distribution is given by*

$$t_2 \Rightarrow \frac{\overline{W}^2(1) - 1}{2\sqrt{\int_0^1 \overline{W}^2(s)ds}}. \tag{1.5}$$

**Theorem 1.2** *In Model (1.1), if $\beta_1 = 1$ and $\beta_2 = \beta_{2T} = 1 - c/k_T$, where $c$ is a fixed positive constant, under assumptions C1-C4, we have*

*(a) (i) When $k_T = o(\sqrt{T})$, $\hat{\tau}_T$ is exactly $T$-consistent, i.e.,*

$$P(\hat{k} \neq k_0) \to 0.$$

*(ii) When $k_T \asymp \sqrt{T}$, $\hat{\tau}_T$ is $T$-consistent, i.e.,*

$$|\hat{k} - k_0| = O_p(1).$$

*(iii) When $\sqrt{T} = o(k_T)$, $\hat{\tau}_T$ is not $T$-consistent, but $\hat{\tau}_T$ is consistent, and its limiting distribution is given by*

$$\frac{c^2 T^2}{k_T^2}(\hat{\tau}_T - \tau_0) \Rightarrow \arg\max_{\nu \in R} \left\{ \frac{W^*(\nu)}{W_1(\tau_0)} - \frac{|\nu|}{2} \right\}, \tag{1.6}$$

*where $W^*(\nu)$ is a two-sided Brownian motion on $R$ defined to be $W^*(\nu) = W_1(-\nu)$ for $\nu \leq 0$ and $W^*(\nu) = W_2(\nu)$ for $\nu > 0$.*

*(b) $\hat{\beta}_1(\hat{\tau}_T)$ is consistent, and its limiting distribution is given by*

$$t_1 \Rightarrow \frac{W^2(1) - 1}{2\sqrt{\int_0^1 W^2(s)ds}}. \tag{1.7}$$

*(c) $\hat{\beta}_2(\hat{\tau}_T)$ is also consistent, and its limiting distribution is given by*

$$t_2 \Rightarrow \frac{\widetilde{W}(1)}{\sqrt{W^2(\tau_0) + (1 - \tau_0)}}. \tag{1.8}$$

**Theorem 1.3** *In Model (1.1), if $\beta_1 = 1$ and $\beta_2 = \beta_{2T} = 1 + c/k_T$, where $c$ is a fixed positive constant, under assumptions C1-C4, we have*

*(a) (i) When $k_T = o(\sqrt{T})$, $\hat{\tau}_T$ is exactly $T$-consistent, i.e.,*

$$P(\hat{k} \neq k_0) \to 0.$$

*(ii) When $k_T \asymp \sqrt{T}$, $\hat{\tau}_T$ is $T$-consistent, i.e.,*

$$|\hat{k} - k_0| = O_p(1).$$

*(iii) When $\sqrt{T} = o(k_T)$, $\hat{\tau}_T$ is not $T$-consistent, but $\hat{\tau}_T$ is consistent, and its limiting distribution is given by (1.6).*

*(b) $\hat{\beta}_1(\hat{\tau}_T)$ is consistent, and its limiting distribution is given by*

$$t_1 \Rightarrow \frac{W^2(1) - 1}{2\sqrt{\int_0^1 W^2(s)ds}}. \tag{1.9}$$

*(c) $\hat{\beta}_2(\hat{\tau}_T)$ is also consistent, and its limiting distribution is given by*

$$t_2 \Rightarrow N(0, 1). \tag{1.10}$$

**Theorem 1.4** *In Model (1.1), if $\beta_1 = \beta_{1T} = 1 + c/k_T$ and $\beta_2 = 1$, where $c$ is a fixed positive constant, under assumptions C1-C4, the estimators $\hat{k}, \hat{\beta}_1(\hat{\tau}_T)$ and $\hat{\beta}_2(\hat{\tau}_T)$ are all consistent and the following results hold:*

$$P(\hat{k} \neq k_0) \to 0, \tag{1.11}$$

$$t_1 \Rightarrow N(0, 1), \tag{1.12}$$

$$t_2 \Rightarrow N(0, 1). \tag{1.13}$$

**Remark 1.2** *The statistics $t_1$ and $t_2$ in Theorems 1.1-1.4 are not pivotal. We denote*

$$t_1' = \sqrt{\frac{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2}{\hat{\sigma}^2}}(\hat{\beta}_1(\hat{\tau}_T) - \beta_1) \text{ and } t_2' = \sqrt{\frac{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2}{\hat{\sigma}^2}}(\hat{\beta}_2(\hat{\tau}_T) - \beta_2),$$

*with*

$$\hat{\sigma}^2 = \frac{1}{T}\left\{\sum_{t=1}^{[\hat{\tau}_T T]}(y_t - \hat{\beta}_1(\hat{\tau}_T)y_{t-1})^2 + \sum_{t=[\hat{\tau}_T T]+1}^{T}(y_t - \hat{\beta}_2(\hat{\tau}_T)y_{t-1})^2\right\},$$

*and it can be proved that*

$$\frac{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} \xrightarrow{p} 1, \quad \frac{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} \xrightarrow{p} 1 \text{ and } \frac{\hat{\sigma}^2}{l(\eta_T)} \xrightarrow{p} 1$$

*hold in Theorems 1.1-1.4. Therefore, Theorems 1.1-1.4 will still hold when $t_1$ and $t_2$ are replaced by $t_1'$ and $t_2'$, respectively.*

**Remark 1.3** *As pointed out by the referees, an AR(1) model with a drift is more realistic in most applications. However, the closed-form expression of $RSS_T(\tau) - RSS_T(\tau_0)$ is difficult to obtain in an AR(1) model with a drift. In Phillips et al. (2015a, 2015b) and Phillips and Shi (2017), they specified the drift as $c_0/T^\gamma$, with $c_0$ being a constant and $\gamma > 1/2$; we can absorb this drift into the error term since it is asymptotically negligible, and denote $\varepsilon_t' = \varepsilon_t + c_0/T^\gamma, t = 1, \cdots, T$ as the new error term and then use the previous change-point estimation procedure to conduct statistical inference. Note that $\varepsilon_t = O_p(\sqrt{l(\eta_T)})$, and hence $\varepsilon_t' = \varepsilon_t \cdot (1 + O_p(\frac{1}{T^\gamma\sqrt{l(\eta_T)}}))$; one can prove that the theoretical results in this paper still hold.*

**Remark 1.4** *Our study is related to the works of Phillips and Yu (2011), Phillips et al. (2011), Phillips et al. (2015a, 2015b, 2015c) and Phillips and Shi (2017). In the aforementioned papers, the authors proposed a structural change AR(1) model with a bubble process and dated the origination of the explosive episode based on recursive right-tailed unit root tests. The explosive AR parameter is estimated by using a demeaning procedure and the least squares method. As pointed out by a referee, our study is also related to a recent work of Harvey et al. (2017), which assumes that $y_t = \mu + u_t$, where $\mu$ is a constant, and $\{u_t, t = 1, \cdots, T\}$ is generated according to a unit root model with a bubble process and a collapse process. They applied the least squares method to the first-differenced data and obtained consistent estimators for the regime change points. Both Harvey et al. (2017) and our work mainly focus on the estimation of the change point by the least squares method, and their results seem better. However, in Harvey et al. (2017), the explosive and the stationary AR (1) models, instead of the mildly explosive and the mildly integrated AR (1)* *models, are used to model the bubble process and the collapse process, respectively. Hence, the differences between the AR parameters are all fixed in Harvey et al. (2017), which leads to the consistency of change-point estimators. In fact, a similar result had been obtained in Theorem 4 of Chong (2001). In our paper, the differences between the AR parameters tend to zero when the sample size tends to infinity, thus our asymptotics are more complicated, and one cannot obtain consistent estimators for all cases. In addition, in this paper, apart from the change point, we also examine the asymptotics of the AR parameters.*

In this paper, we study the $t$-ratios of $\beta_1$ and $\beta_2$ rather than the LSEs of $\beta_1$ and $\beta_2$. For the unit root model, the limiting distribution of the LSE of the AR parameter is $\frac{W^2(1)-1}{2\int_0^1 W^2(s)ds}$ with the convergence rate $T$, while the limiting distribution of the $t$-ratio for the AR parameter is $\frac{W^2(1)-1}{2\sqrt{\int_0^1 W^2(s)ds}}$ with the convergence rate $T/\sqrt{l(\eta_T)}$. Although there is a possible reduction in the convergence rate, the reduction is negligible since $l(\cdot)$ is a slowly varying function. The advantage of using the $t$-ratio is that $\frac{W^2(1)-1}{2\sqrt{\int_0^1 W^2(s)ds}}$ is less skewed compared to $\frac{W^2(1)-1}{2\int_0^1 W^2(s)ds}$. For the mildly integrated AR(1) model, the limiting distribution of the $t$-ratio for the AR parameter, $\beta_T = 1 - c/k_T$, is $N(0,1)$ with the convergence rate $\sqrt{Tk_T/l(\eta_T)}$ by Theorem 3.2 in Phillips and Magdalinos (2007a). In this case, the benefit of $t$-ratio is not obvious as the limiting distribution of the LSE of the AR parameter is also normal with the convergence rate $\sqrt{Tk_T}$. However, the benefit of using the $t$-ratio in the mildly explosive AR(1) model becomes significant, since the limiting distribution of the $t$-ratio for the AR coefficient, $\beta_T = 1 + c/k_T$, is $N(0,1)$ with the convergence rate $\beta_T^T k_T/\sqrt{l(\eta_T)}$ by Theorem 4.3 in Phillips and Magdalinos (2007a). Compared with the result (b) of Theorem 4.3 in Phillips and Magdalinos (2007a), although there is a possible reduction in the convergence rate for the $t$-ratio, the reduction is negligible, and the limiting distribution is no longer a Cauchy distribution, which has infinite mean and variance. The $t$-ratio significantly improves the estimation accuracy in this case.

It follows from Theorems 1.1-1.4 that, for the $t$-ratio for $\beta_2$, the first sub-samples $\{y_1, \cdots, y_{[\tau_0 T]}\}$ will not affect its limiting distribution when $\beta_1 < \beta_2$. That is, the limiting distribution of the $t$-ratio for $\beta_2$ in this case is the same as that in an AR(1) model without a structural change. Meanwhile, it is not the case when $\beta_1 > \beta_2$. For example, the limiting distribution of $t_2$ in Theorem 1.4 is $N(0,1)$ instead of $\frac{\overline{W}^2(1)-1}{2\sqrt{\int_0^1 \overline{W}^2(s)ds}}$, which is caused by the influence of the first sub-sample.

Note that in Theorem 1.1, since $\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2 = O_p(Tk_T l(\eta_T))$ and $\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2 = O_p(T^2 l(\eta_T))$ by Lemma A.1 and Lemma A.2 below respectively, it can be easily seen that the convergence rates of the LSEs of $\beta_1$ and $\beta_2$ are $\sqrt{Tk_T}$ and $T$ respectively. Similarly, it can be shown that the convergence rates of the LSEs of $\beta_1$ and $\beta_2$ are $T$ and $\sqrt{Tk_T}$ respectively in Theorem 1.2, are $T$ and $\sqrt{Tk_T}(1 + c/k_T)^{T-[\tau_0 T]}$ in Theorem 1.3, and are $k_T(1 + c/k_T)^{[\tau_0 T]}$ and $\sqrt{Tk_T}(1 + c/k_T)^{[\tau_0 T]}$ in Theorem 1.4. Hence, the convergence rate of the LSE when the AR parameter equals one is faster than that in the case when the AR parameter equals $1 - c/k_T$, since the signal from the regressor $y_{t-1}$ in the former case is stronger than that in the latter case. It should also be noted that the convergence rate of $\hat{\tau}_T$ in Theorem 1.2 is faster than that in Theorem 1.1. This is because the signal from the regressor $y_{t-1}$ when $(\beta_1, \beta_2) = (1, 1 - c/k_T)$ is stronger than that from the regressor $y_{t-1}$ when $(\beta_1, \beta_2) = (1 - c/k_T, 1)$. However, these findings are not totally applicable to Theorem 1.3 and Theorem 1.4. Note that in Theorem 1.3, the convergence rate of the LSE of $\beta_2$ is faster than that of the LSE of $\beta_1$. Moreover, the convergence rate of $\hat{\tau}_T$ in Theorem 1.4 is faster than that in Theorem 1.3, since the signal from the regressor $y_{t-1}$ when $(\beta_1, \beta_2) = (1+c/k_T, 1)$ is stronger compared to the case when $(\beta_1, \beta_2) = (1, 1+c/k_T)$. These are consistent with Theorem 1.1 and Theorem 1.2. However, it is surprising that the convergence rate of the LSE of $\beta_2$ is faster than that of the LSE of $\beta_1$ in Theorem 1.4. This is completely different from the findings of Theorem 1.1 and Theorem 1.2. The reason is that the second sub-sample is more affected by the first sub-sample in Theorem 1.4 than that in Theorem 1.3.

The precision of $\hat{k}$ depends on both the strength of the signal from the model and the difference between $\beta_1$ and $\beta_2$ (i.e., $c/k_T$). The strength of the signal from the model increases from Theorem 1.1 to Theorem 1.4. In general, the signal from the model in Theorem 1.1 is too weak for $k_0$ to be located for any $k_T = o(T)$, while the signal from the model in Theorem 1.4 is so strong that $k_0$ can be located for any $k_T = o(T)$. For the models in Theorems 1.2 and 1.3, although the signal from the model in Theorem 1.3 is stronger than that in Theorem 1.2, it is surprising that $k_0$ can be located consistently only when $k_T = o(\sqrt{T})$ in both models. This is because the increment in the signal from Theorem 1.2 to Theorem 1.3 is so small that the difference between $RSS_T(\tau_0)$ and $RSS_T(\tau_0 \pm \frac{m}{T})$ for a fixed integer $m$ in Theorem 1.3 is asymptotically the same as that in Theorem 1.2, which can be seen from the proofs of Theorems 1.2 and 1.3. Although the increment in the signal from the model in Theorem 1.2 to the model in Theorem 1.3 does not help to locate $k_0$, it improves the convergence rate of the LSE of $\beta_2$.

It is interesting to find that the conventional $T$-consistent estimate for $\hat{\tau}_T$ occurs only when $k_T \asymp \sqrt{T}$ in both Theorems 1.2 and 1.3. When $k_T$ is of an order higher than $\sqrt{T}$, $\beta_1$ and $\beta_2$ will be very close to one another, and the estimation error for $\hat{k}$ will be huge. When $k_T$ has a smaller magnitude than $\sqrt{T}$, it implies that $\beta_1$ and $\beta_2$ have enough difference, and the estimation error for $\hat{k}$ reduces tremendously.

The rest of the paper is organized as follows. Section 2 demonstrates some finite-sample Monte Carlo results for our theoretical findings in this paper. Section 3 concludes the paper. The proofs of Theorems 1.1-1.4 are relegated to Appendices A-D respectively, with some technical proofs being moved to online supplementary material to this article which is available at Cambridge Journals Online (journals.cambridge.org/ect).

# 2 Simulations

For empirical applications, we perform the following two experiments to see how well the finite-sample properties of the estimators follow the asymptotics. In both experiments, the sample size is set at $T = 600$, the interval $[\underline{\tau}, \overline{\tau}]$ is taken as $[0.05, 0.95]$ (hence the search for the break fraction is conducted within this interval in our experiments), the true break fraction is set at $\tau_0 = 0.5$ (hence $k_0 = 300$), and the number of replications is set at $N = 50,000$; $\{y_t\}_{t=1}^T$ is generated from Model (1.1), $y_0$ is set at zero for simplicity, and $\{\varepsilon_t\}_{t=1}^T$ are generated independently from $N(0,1)$, we hence take $l(\eta_T) = 1$ since $T = 600$ is large, and $k_T = T^\alpha$ with $\alpha = 0.3, 0.5$ or $0.7$. The case where $\alpha = 0.3$ implies $k_T = o(\sqrt{T})$, the case where $\alpha = 0.5$ implies $k_T \asymp \sqrt{T}$, and the case where $\alpha = 0.7$ implies $\sqrt{T} = o(k_T)$. The graph of the distribution of $\frac{W^2(1)-1}{2\sqrt{\int_0^1 W^2(s)ds}}$ is plotted by dividing the interval $[0,1]$ into $5,000$ equally spaced sub-intervals first and then using the corresponding Riemann sums to approximate the integral. The number of replications is also set at $N = 50,000$.

## 2.1 Experiment 1

First, we conduct experiments to verify Theorem 1.1 and Theorem 1.2. We take $c = 3$ in this experiment. Note that the two AR parameters have a large difference ($= 3/600^{0.3} = 0.440$) when $\alpha = 0.3$, have a moderate difference ($= 3/600^{0.5} = 0.122$) when $\alpha = 0.5$ and have a very small difference ($= 3/600^{0.7} = 0.034$) when $\alpha = 0.7$. Figure 1 and Figure 2 show the histograms of $\hat{k}$ and the distributions of $t_1$ and $t_2$ for Theorems 1.1 and 1.2 respectively. Theorem 1.1 states that $\hat{k}$ is not a consistent estimator of $k_0$ and the estimation error is of $O_p(k_T)$. This is supported by Figure 1. Part (b) of Theorem 1.1 predicts that $t_1$ should have a normal distribution, and part (c) of Theorem 1.1 predicts that $t_2$ should have a

Dickey-Fuller $t$-distribution. Figure 1 agrees with these results. Part (a) of Theorem 1.2 predicts that $\hat{k}$ is a consistent estimator of $k_0$ when $k_T = o(\sqrt{T})$, has a finite estimation error in probability when $k_T \asymp o(\sqrt{T})$ and has a larger estimation error in probability when $\sqrt{T} = o(k_T)$. These theoretical findings are all supported by Figure 2. Part (b) of Theorem 1.2 predicts that $t_1$ should have a Dickey-Fuller $t$-distribution, and part (c) of Theorem 1.2 predicts that $t_2$ should have a symmetric distribution around zero that looks like a normal distribution. These theoretical results are also supported by Figure 2.

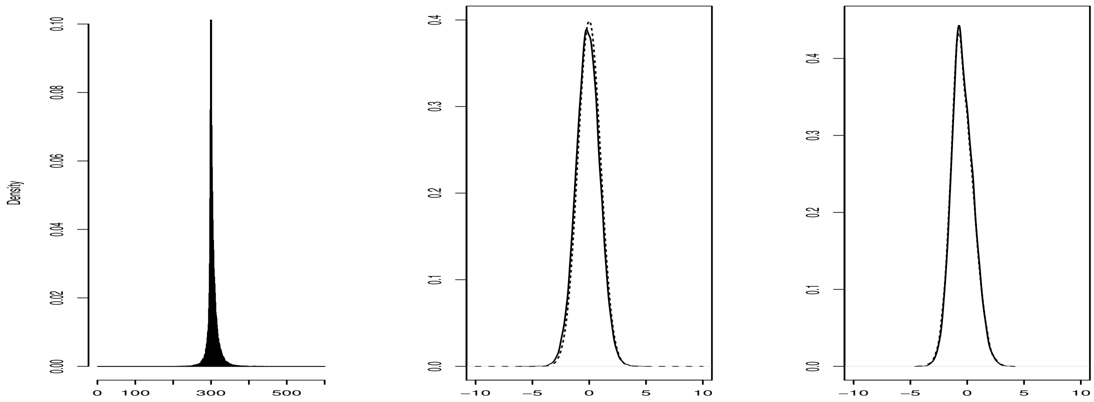
Figure 1 and Figure 2 also indicate that the smaller the magnitude of change, the larger the estimation error for $\hat{k}$ and the poorer the finite-sample performance of $t_1$ and $t_2$, which agrees with our intuition.

### 2.2 Experiment 2

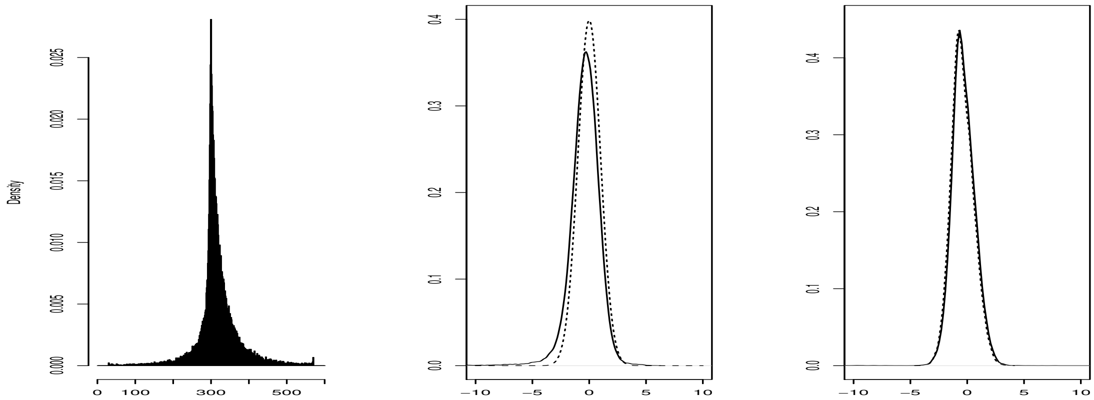
Second, we conduct experiments for Theorems 1.3 and 1.4. Here, we take $c = 0.7$. We have also conducted experiments with larger $c$. However, it is found that the finite-sample distributions of $t_2$ in Theorem 1.3 and $t_1$ and $t_2$ in Theorem 1.4 suffer from shape distortion. This phenomenon can be partially explained by the findings in Anderson (1959), which showed that, in general, the limiting distributions of the LSE and the $t$-ratio for the AR parameter in an explosive AR(1) model may not exist. Hence, we use $c = 0.7$ in this experiment, which guarantees that the mildly explosive AR parameter is not too far away from unity. Figure 3 and Figure 4 show the distributions of $\hat{k}$, $t_1$ and $t_2$ for Theorems 1.3 and 1.4 respectively. It can be shown that (1) Theorem 1.3 is supported by Figure 3; (2) Theorem 1.4 is supported by Figure 4, except that the histograms of $\hat{k}$ when $\alpha = 0.7$ is not very satisfactory due to the close distance between $\beta_1$ and $\beta_2$; and (3) the smaller the magnitude of change, the larger the estimation error for $\hat{k}$ and the poorer the finite-sample performance of $t_1$ and $t_2$.

## 3 Conclusions

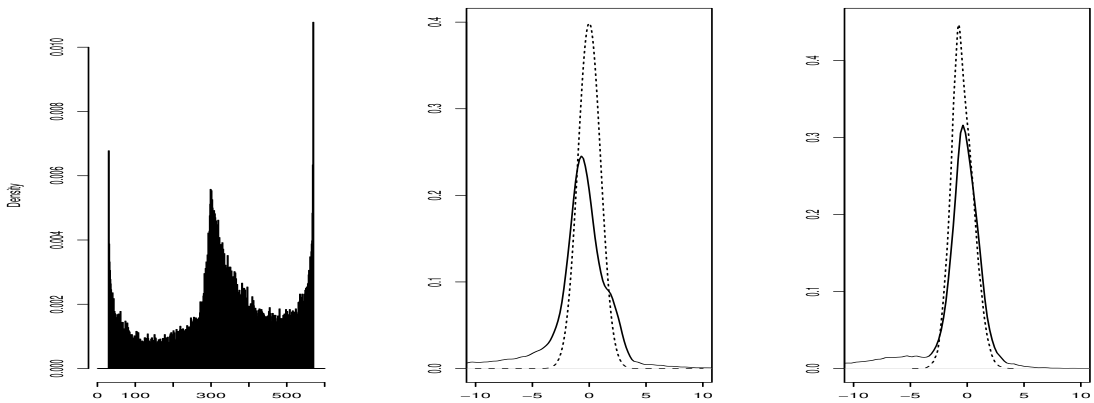
In this article, we examined the asymptotic properties of the LSE of the change point and the $t$-ratios for the AR parameters in a mildly integrated AR(1) model and a mildly explosive AR(1) model with a structural change. Some interesting findings are obtained: (1) the stronger the signals from the model are, the easier it is for the change point to be located. This suggests that, in general, the estimation of the change point in a mildly explosive AR(1) model is easier than that in a mildly integrated AR(1) model. However,




(a) $\beta_1 = 1 - c/T^{\alpha}, \beta_2 = 1, \alpha = 0.3$

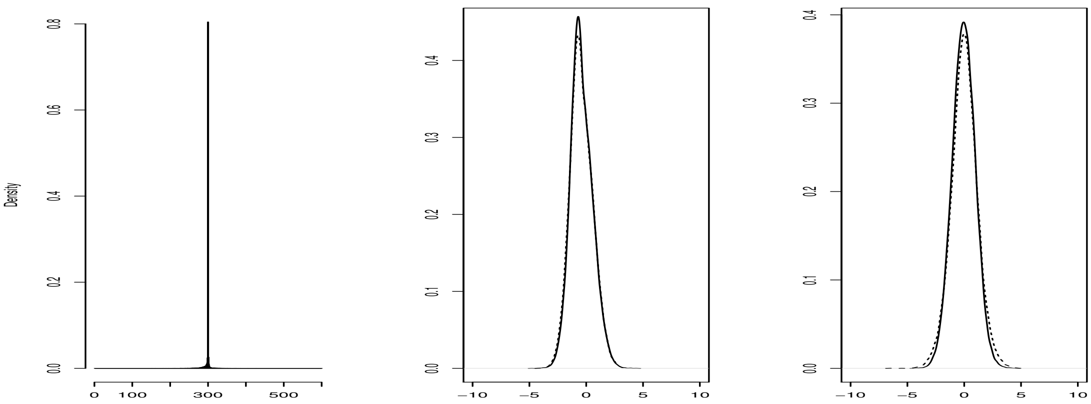



(b) $\beta_1 = 1 - c/T^{\alpha}, \beta_2 = 1, \alpha = 0.5$

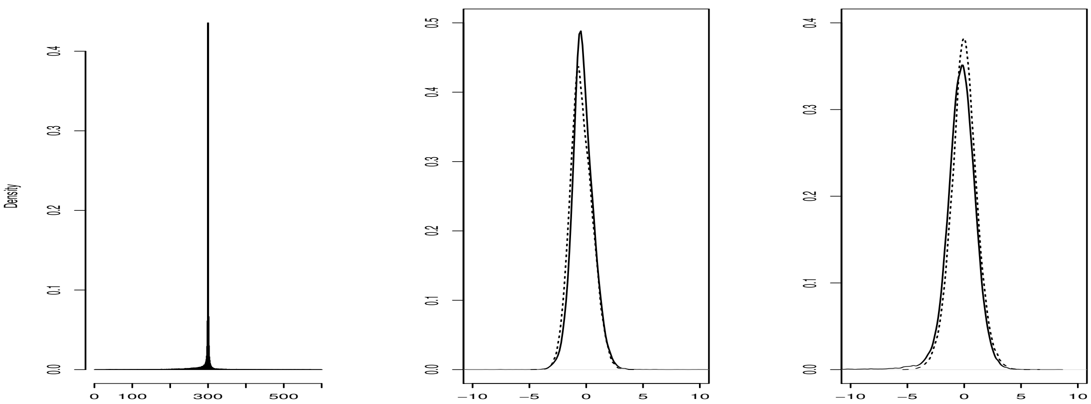



(c) $\beta_1 = 1 - c/T^{\alpha}, \beta_2 = 1, \alpha = 0.7$

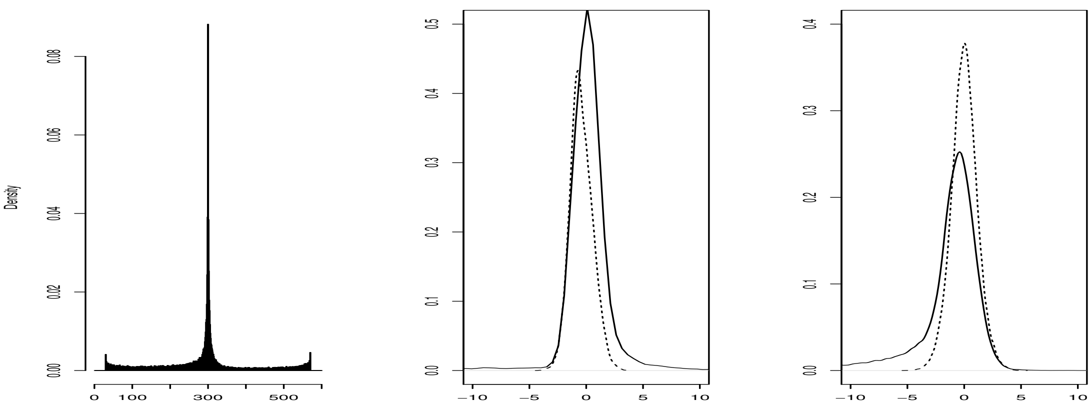
Figure 1: Histograms of $\hat{k}$ as well as the finite-sample distributions and the corresponding limiting distributions of the statistics $t_1$ and $t_2$ under the situation where $c = 3$ and $T = 600$. The solid lines represent the graphs when $T = 600$ and the dashed lines represent the graph when $T = \infty$.




(a) $\beta_1 = 1, \beta_2 = 1 - c/T^{\alpha}, \alpha = 0.3$

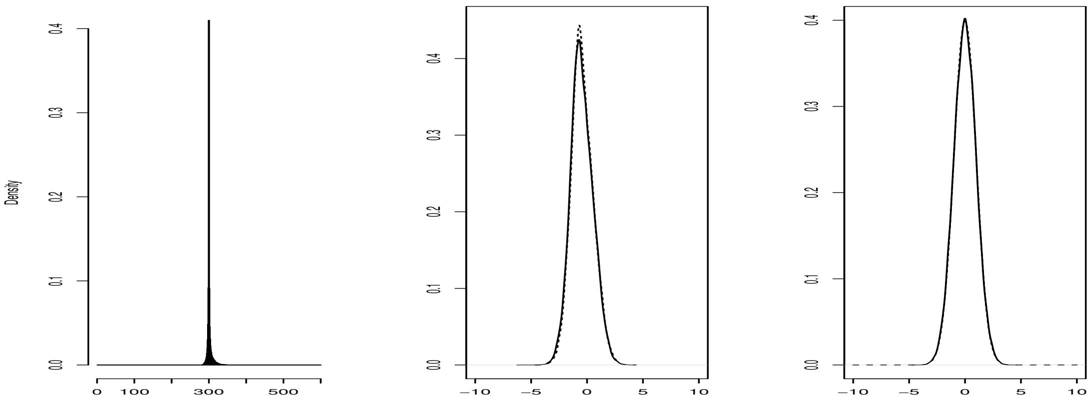



(b) $\beta_1 = 1, \beta_2 = 1 - c/T^{\alpha}, \alpha = 0.5$

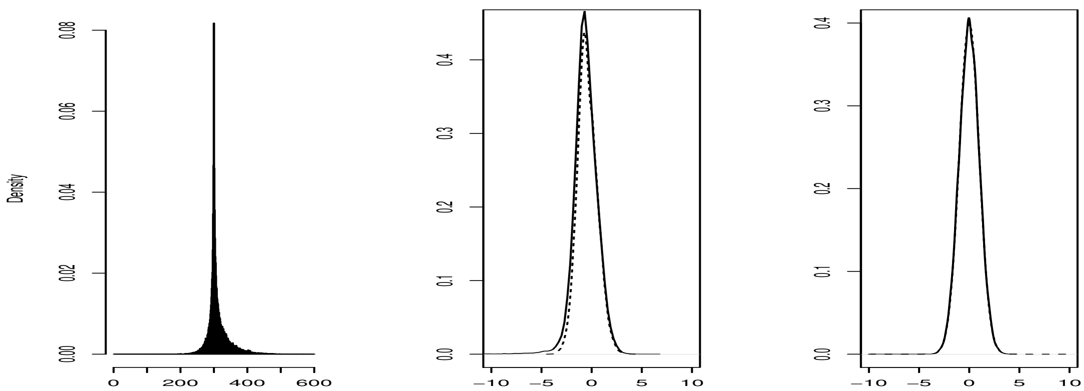



(c) $\beta_1 = 1, \beta_2 = 1 - c/T^{\alpha}, \alpha = 0.7$

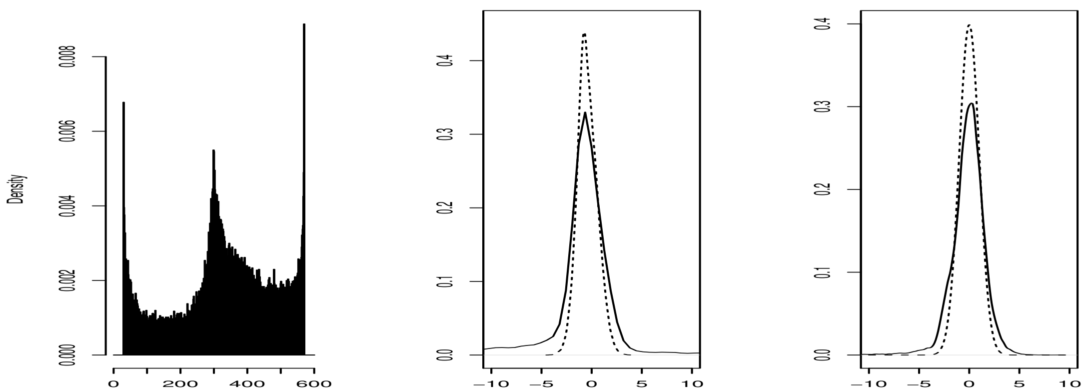
Figure 2: Histograms of $\hat{k}$ as well as the finite-sample distributions and the corresponding limiting distributions of the statistics $t_1$ and $t_2$ under the situation where $c = 3$ and $T = 600$. The solid lines represent the graphs when $T = 600$ and the dashed lines represent the graph when $T = \infty$.




(a) $\beta_1 = 1, \beta_2 = 1 + c/T^{\alpha}, \alpha = 0.3$

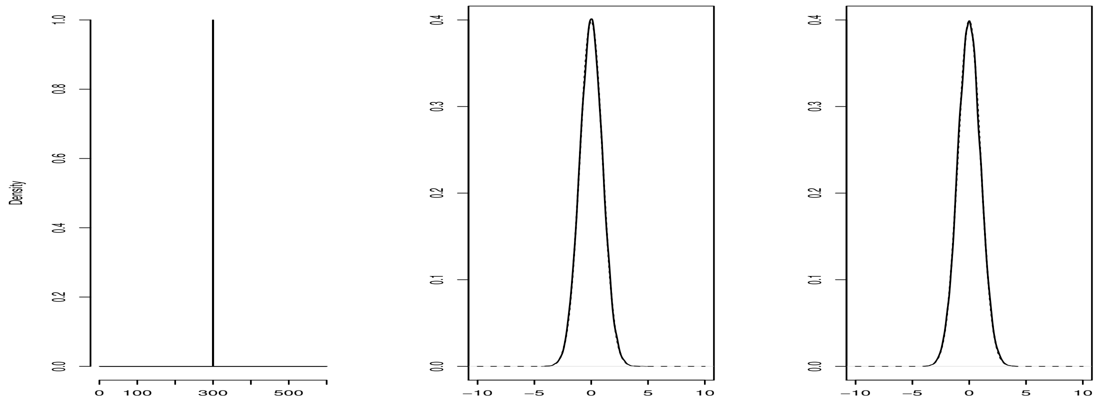



(b) $\beta_1 = 1, \beta_2 = 1 + c/T^{\alpha}, \alpha = 0.5$

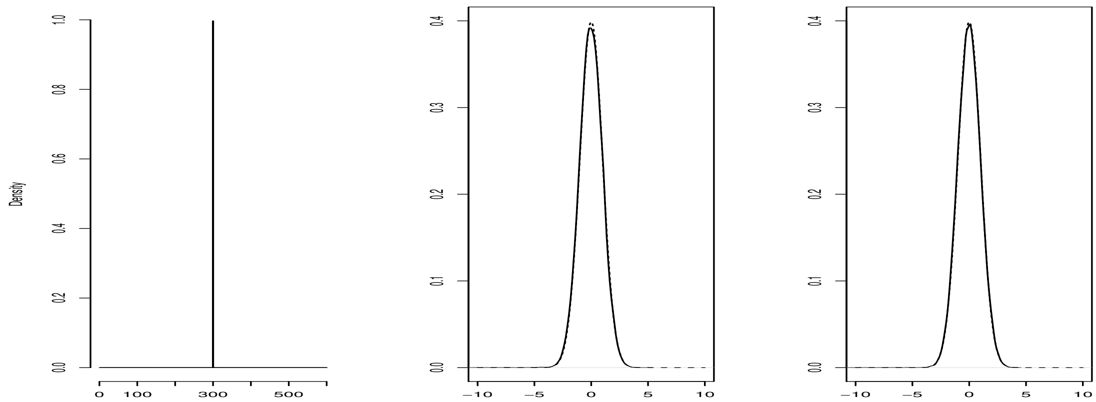



(c) $\beta_1 = 1, \beta_2 = 1 + c/T^{\alpha}, \alpha = 0.7$

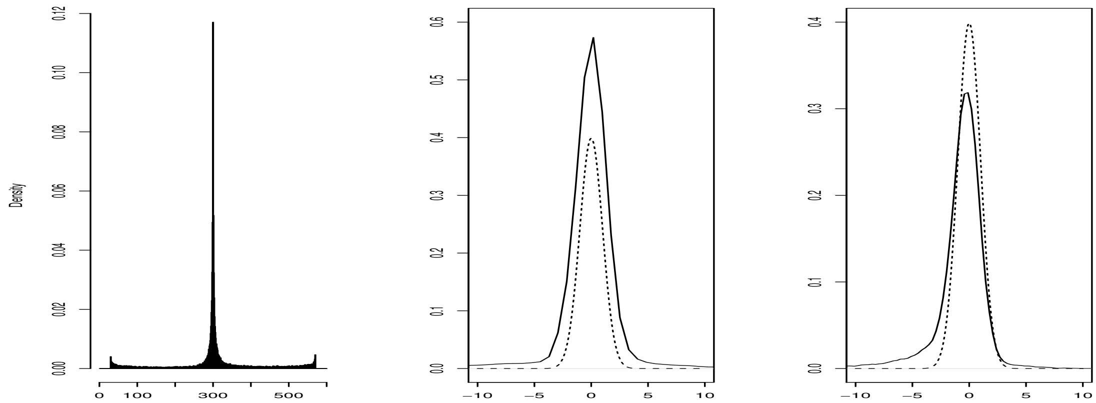
Figure 3: Histograms of $\hat{k}$ together with the finite-sample distributions and the corresponding limiting distributions of the statistics $t_1$ and $t_2$ under the situation where $c = 0.7$ and $T = 600$. The solid lines represent the graphs when $T = 600$ and the dashed lines represent the graph when $T = \infty$.




(a) $\beta_1 = 1 + c/T^{\alpha}, \beta_2 = 1, \alpha = 0.3$




(b) $\beta_1 = 1 + c/T^{\alpha}, \beta_2 = 1, \alpha = 0.5$




(c) $\beta_1 = 1 + c/T^{\alpha}, \beta_2 = 1, \alpha = 0.7$

Figure 4: Histograms of $\hat{k}$ together with the finite-sample distributions and the corresponding limiting distributions of the statistics $t_1$ and $t_2$ under the situation where $c = 0.7$ and $T = 600$. The solid lines represent the graphs when $T = 600$ and the dashed lines represent the graph when $T = \infty$.

if the first sub-sample comes from a unit root model, then it is more difficult to locate the change point regardless of the order of the second sub-sample; (2) in the presence of a change point, the first sub-sample will not affect the limiting distribution of the $t$-ratio for the second AR parameter when $\beta_1 < \beta_2$, while this is not the case when $\beta_1 > \beta_2$; (3) when a unit root model switches to a mildly integrated or mildly explosive AR(1) model, the asymptotic properties of the LSE of $k_0$ are the same. In particular, in both situations, our results reveal that $P(\hat{k} \neq k_0) \to 0$ when $k_T = o(\sqrt{T})$, $|\hat{k} - k_0| = O_p(1)$ when $k_T \asymp \sqrt{T}$ and $|\hat{k} - k_0| = O_p(k_T^2/T)$ when $\sqrt{T} = o(k_T)$. The phase transition for the estimation error of $\hat{k}$ occurs when $k_T \asymp \sqrt{T}$; (4) compared with the LSE of the AR parameters, the $t$-ratios for the AR parameters have better estimation accuracy without any reduction in convergence rate when the variance of the model errors is finite and with a reduction in convergence rate when the variance of the model errors is infinite, but this reduction is asymptotically negligible.

**Acknowledgements**. The authors would like to thank the Editor, Peter C. B. Phillips, a co-editor and three anonymous referees for their constructive and illuminating comments that have significantly improved the present article.

## References

[1] Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *The Annals of Mathematical Statistics* 30 (3): 676-687.

[2] Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory* **13** (3): 315-352.

[3] Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** (1): 47-78.

[4] Bai, J., Chen H., Chong, T. T. L. and Wang, S. X. (2008). Generic consistency of the break-point estimators under specification errors in a multiple-break model. *The Econometrics Journal* **11** (2): 287-307.

[5] Barsky, R. B. (1987). The Fisher hypothesis and the forecastability and persistence of inflation. *Journal of Monetary Economics* **19** (1): 3-24.

[6] Berkes, I., Horváth, L., Ling, S. and Schauer, J. (2011). Testing for structural change of AR model to threshold AR model. *Journal of Time Series Analysis* **32** (5): 547-565.

[7] Burdekin R. C. K. and Siklos, P. L. (1999). Exchange rate regimes and shifts in inflation persistence: Does nothing else matter? *Journal of Money, Credit and Banking* **31** (2): 235-247.

[8] Chan, N. H. and Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *The Annals of Statistics* **15** (3): 1050-1063.

[9] Chong, T. T. L. (2001). Structural change in AR(1) models. *Econometric Theory* **17** (1): 87-155.

[10] Chong, T. T. L. (2003). Generic consistency of the break-point estimator under specification errors. *The Econometrics Journal* **6** (1): 167-192.

[11] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (2nd Edition). Volume 2. John Wiley and Sons, New York.

[12] Halunga, A. G. and Osborn, D. R. (2012). Ratio-based estimators for a change point in persistence. *Journal of Econometrics* **171** (1): 24-31.

[13] Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in U.S. labor productivity. *The Journal of Economic Perspectives* **15** (4): 117-128.

[14] Hakkio, C. S. and Rush, M. (1991). Is the budget deficit "too large"? *Economic Inquiry* **29** (3): 429-445.

[15] Harvey, D. I., Leybourne, S. J. and Sollis, R. (2017). Improving the accuracy of asset price bubble start and end date estimators. *Journal of Empirical Finance* **40**: 121-138.

[16] Harvey, D. I., Leybourne, S. J. and Taylor, A. M. R. (2006). Modified tests for a change in persistence. *Journal of Econometrics* **134** (2): 441-469, and Corrigendum, *Journal of Econometrics* **168** (2): 407.

[17] Huang S. H., Pang, T. X. and Weng, C. G. (2014). Limit theory for moderate deviations from a unit root under innovations with a possibly infinite variance. *Methodology and Computing in Applied Probability* **16** (1): 187-206.

[18] Kejriwal, M., Perron, P. and Zhou, J. (2013). Wald tests for detecting multiple structural changes in persistence. *Econometric Theory* **29** (2): 289-323.

[19] Kim, J. and Pollard, D. (1990). Cube root asymptotics. *The Annals of Statistics* **18** (1): 191-219.

[20] Lee, S., Seo, M. H. and Shin, Y. (2016). The lasso for high dimensional regression with a possible change point. *Journal of the Royal Statistical Society Statistical Methodology Series B* **78** (1): 193-210.

[21] Leybourne, S., Kim, T., Smith, V. and Newbold, P. (2003). Tests for a change in persistence against the null of difference-stationarity. *The Econometrics Journal* **6** (2): 291-311.

[22] Magdalinos, T. (2012). Mildly explosive autoregression under weak and strong dependence. *Journal of Econometrics* **169** (2): 179-187.

[23] Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business* **36** (4): 394-419.

[24] Mankiw, N. G. and Miron, J. A. (1986). The changing behavior of the term structure of interest rates. *The Quarterly Journal of Economics* **101** (2): 211-228.

[25] Mankiw, N. G., Miron, J. A. and Weil, D. N. (1987). The adjustment of expectations to a change in regime: A study of the founding of the Federal Reserve. *The American Economic Review* **77** (3): 358-374.

[26] Pang, T. and Zhang, D. (2015). Asymptotic inferences for an AR(1) model with a change point and possibly infinite variance. *Communications in Statistics-Theory and Methods* **44** (22): 4848-4865.

[27] Pang, T., Zhang, D. and Chong, T. T. L. (2014). Asymptotic inferences for an AR(1) model with a change point: stationary and nearly non-stationary cases. *Journal of Time Series Analysis* **35** (2): 133-150.

[28] Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika* **74** (3): 535-547.

[29] Phillips, P. C. B. and Magdalinos, T. (2007a). Limit theory for moderate deviations from a unit root. *Journal of Econometrics* **136** (1): 115–130.

[30] Phillips, P. C. B. and Magdalinos, T. (2007b). Limit theory for moderate deviations from unity under weak dependence. In: Phillips, G.D.A., Tzavalis, E. (Eds.), The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis. Cambridge: Cambridge University Press, pp.123-162.

[31] Phillips, P. C. B. and Shi, S. P. (2017). Financial bubble implosion and reverse regression. Econometric Theory, forthcoming.

[32] Phillips, P. C. B., Shi, S. and Yu, J. (2015a). Testing for Multiple Bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review* **56** (4): 1043-1078.

[33] Phillips, P. C. B., Shi, S. and Yu, J. (2015b). Testing for Multiple Bubbles: Limit Theory of Real-Time Detectors. *International Economic Review* **56** (4): 1079-1134.

[34] Phillips, P. C. B., Shi, S. and Yu, J. (2015c). Supplement to Two Papers on Multiple Bubbles. *International Economic Review*, Online Supplementary Material.

[35] Phillips, P. C. B., Wu, Y. and Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: when did exuberance escalate asset values? *International Economic Review* **52** (1): 201-226.

[36] Phillips, P. C. B. and Yu, J. (2011). Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics* **2** (3): 455-491.

[37] Pitarakis, J.-Y. (2004). Least squares estimation and tests of breaks in mean and variance under misspecification. *The Econometrics Journal* **7** (1): 32-54.

[38] Qu, Z. (2008). Testing for structural change in regression quantiles. *Journal of Econometrics* **146** (1): 170-184.

[39] Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics* **14** (1): 11-30.

[40] Stock, J. H. and Watson, M. W. (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In R. F. Engle and H. White (Eds.), *Cointegration, Causality and Forcasting: A Festschrift in Honour of Clive. W. J. Granger*, pp. 1-44. Oxford: Oxford University Press.

## Appendix A: Proof of Theorem 1.1

Recalling the definitions in (1.2), it can be shown that $Tl(\eta_T) \leq \eta_T^2$ for all $T \geq 1$ and $\eta_T^2 \approx Tl(\eta_T)$ for large $T$. In addition, for each given $T$, we let

$$\begin{cases} \varepsilon_t^{(1)} = \varepsilon_t I\{|\varepsilon_t| \leq \eta_T\} - E(\varepsilon_t I\{|\varepsilon_t| \leq \eta_T\}) \\ \varepsilon_t^{(2)} = \varepsilon_t I\{|\varepsilon_t| > \eta_T\} - E(\varepsilon_t I\{|\varepsilon_t| > \eta_T\}) \end{cases} \tag{A.1}$$

for $t = 1, \cdots, T$. This is a well-known truncation technique for dealing with the weak convergence of the random variables from the DAN with zero mean and possibly infinite variance. Huang et al. (2014) successfully extended the results in Phillips and Magdalinos (2007a) to the DAN case by applying this truncation technique.

The following three lemmas are needed in the proof of Theorem 1.1, and their proofs can be found in the online supplementary material.

**Lemma A.1** *Let $\{y_t, t \geq 1\}$ be generated according to Model (1.1), where $\beta_1 = \beta_{1T} = 1 - c/k_T$ for a positive constant $c$. Then under assumptions C1-C4, the following results hold jointly:*

*(a)* $\frac{1}{\sqrt{Tk_T l(\eta_T)}} \sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t \Rightarrow N(0, \frac{\tau_0}{2c})$,

*(b)* $\frac{1}{Tk_T l(\eta_T)} \sum_{t=1}^{[\tau_0 T]} y_{t-1}^2 \xrightarrow{p} \frac{\tau_0}{2c}$,

*(c)* $\frac{y_{[rT]}}{\sqrt{k_T l(\eta_T)}} \Rightarrow \int_0^\infty \exp(-cs)dW(s)$ *for any* $0 < r \leq \tau_0$.

**Lemma A.2** *Let $\{y_t, t \geq 1\}$ be generated according to Model (1.1), where $\beta_1 = \beta_{1T} = 1-c/k_T$ for a positive constant $c$ and $\beta_2 = 1$. Then under assumptions C1-C4, the following results hold jointly:*

*(a)* $\frac{1}{Tl(\eta_T)} \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t \Rightarrow \frac{1}{2}(\overline{W}(1) - \overline{W}(\tau_0))^2 - \frac{1}{2}(1-\tau_0)$,

*(b)* $\frac{1}{T^2 l(\eta_T)} \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2 \Rightarrow \int_{\tau_0}^1 (\overline{W}(s) - \overline{W}(\tau_0))^2 ds$.

**Lemma A.3** *Let $\{y_t, t \geq 1\}$ be generated according to Model (1.1), where $\beta_1 = \beta_{1T} = 1-c/k_T$ for a positive constant $c$ and $\beta_2 = 1$. Then under assumptions C1-C4, the following results hold:*

*(a)* $\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}^2 = O_p(k_T^2 l(\eta_T))$ *and* $\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t = O_p(k_T l(\eta_T))$ *when* $\hat{\tau}_T \leq \tau_0$,

*(b)* $\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2 = O_p(k_T^2 l(\eta_T))$ *and* $\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}\varepsilon_t = O_p(k_T l(\eta_T))$ *when* $\hat{\tau}_T > \tau_0$.

**Proof of Theorem 1.1**. To derive the limiting distribution of $\hat{\tau}_T$, one can follow Appendix G in Chong (2001) with the following two main modifications: (1) let $g(T) = k_T/c$ in Appendix G in Chong (2001); (2) replace $\int_0^\infty \exp(-s)dW_1(s)$ by $\sqrt{c}\int_0^\infty \exp(-cs)dW_1(s)$, which is the limiting distribution of $y_{[\tau_0 T]-t-1}/\sqrt{g(T)l(\eta_T)}$ for $0 \leq t \leq [|\nu|g(T)] - 1$ (where $\nu$ is a constant) by noting that

$$\begin{aligned}
\frac{y_{[\tau_0 T]-t-1}}{\sqrt{g(T)l(\eta_T)}} &= \frac{1}{\sqrt{g(T)l(\eta_T)}}\Big((1-\frac{c}{k_T})^{[\tau_0 T]-t-1}y_0 + \sum_{i=0}^{[\tau_0 T]-t-2}(1-\frac{c}{k_T})^i \varepsilon_{[\tau_0 T]-t-1-i}\Big) \\
&= \sqrt{c}\cdot \sum_{i=0}^{[\tau_0 T]-t-2}\big((1-\frac{c}{k_T})^{k_T}\big)^{i/k_T}\frac{\varepsilon_{[\tau_0 T]-t-1-i}}{\sqrt{k_T l(\eta_T)}} + o_p(1) \\
&\Rightarrow \sqrt{c}\int_0^\infty \exp(-cs)dW_1(s) = B_c(\frac{1}{2}).
\end{aligned}$$

Moreover, it is worth mentioning that (1.3) is a special case of the limiting distribution of $\hat{\tau}_T$ in Theorem 3 in Chong (2001) when the moment conditions $E(y_0^2) < \infty$ and $E(\varepsilon_t^4) < \infty$

in his paper are satisfied and $c = 1$. To relax these moment conditions, we only need to apply the truncation technique (A.1). The details are omitted for brevity.

To find the limiting distribution of $t_1$ under $k_T = o(T)$, one can follow Appendix G in Chong (2001) and apply Lemma A.1 and Lemma A.3 to have

$$\begin{aligned}
&\sqrt{Tk_T}(\hat\beta_1(\hat\tau_T) - \hat\beta_1(\tau_0)) \\
= \;& \sqrt{Tk_T}\Big(\frac{\sum_{t=1}^{[\hat\tau_T T]} y_t y_{t-1}}{\sum_{t=1}^{[\hat\tau_T T]} y_t^2} - \frac{\sum_{t=1}^{[\tau_0 T]} y_t y_{t-1}}{\sum_{t=1}^{[\tau_0 T]} y_t^2}\Big) \\
= \;& I\{\hat\tau_T \le \tau_0\}\sqrt{Tk_T}\Big(\frac{\sum_{t=[\hat\tau_T T]+1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=1}^{[\hat\tau_T T]} y_{t-1}^2}\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} - \frac{\sum_{t=[\hat\tau_T T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\hat\tau_T T]} y_{t-1}^2}\Big) \\
& + I\{\hat\tau_T > \tau_0\}\sqrt{Tk_T}\Big( - \frac{\sum_{t=[\tau_0 T]+1}^{[\hat\tau_T T]} y_{t-1}^2}{\sum_{t=1}^{[\hat\tau_T T]} y_{t-1}^2}\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} + \frac{\sum_{t=[\tau_0 T]+1}^{[\hat\tau_T T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\hat\tau_T T]} y_{t-1}^2} \\
& \qquad + (\beta_2 - \beta_{1T})\frac{\sum_{t=[\tau_0 T]+1}^{[\hat\tau_T T]} y_{t-1}^2}{\sum_{t=1}^{[\hat\tau_T T]} y_{t-1}^2}\Big) \\
= \;& I\{\hat\tau_T \le \tau_0\}\sqrt{Tk_T}\Big(O_p(\frac{k_T^2 l(\eta_T)}{Tk_T l(\eta_T)})O_p(\frac{1}{\sqrt{Tk_T}}) + O_p(\frac{k_T l(\eta_T)}{Tk_T l(\eta_T)})\Big) \\
& + I\{\hat\tau_T > \tau_0\}\sqrt{Tk_T}\Big(O_p(\frac{k_T^2 l(\eta_T)}{Tk_T l(\eta_T)})O_p(\frac{1}{\sqrt{Tk_T}}) + O_p(\frac{k_T l(\eta_T)}{Tk_T l(\eta_T)}) + O_p(\frac{k_T^2 l(\eta_T)}{Tk_T^2 l(\eta_T)})\Big) \\
= \;& o_p(1),
\end{aligned}$$

which implies

$$\sqrt{\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{l(\eta_T)}}(\hat\beta_1(\hat\tau_T) - \hat\beta_1(\tau_0)) = o_p(1)$$

by Lemma A.1. Thus, $\hat\beta_1(\hat\tau_T)$ and $\hat\beta_1(\tau_0)$ have the same asymptotic distribution. Applying Lemma A.1 again, we have

$$\sqrt{\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{l(\eta_T)}}(\hat\beta_1(\tau_0) - \beta_{1T}) = \frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sqrt{l(\eta_T)\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}} \Rightarrow N(0,1),$$

which immediately implies that

$$t_1 = \sqrt{\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{l(\eta_T)}}(\hat\beta_1(\hat\tau_T) - \beta_{1T}) \Rightarrow N(0,1).$$

To find the limiting distribution of $t_2$ under $k_T = o(T)$, one can also follow Appendix G in Chong (2001) and apply Lemma A.2 and Lemma A.3 to have

$$\begin{aligned}
&T(\hat\beta_2(\hat\tau_T) - \hat\beta_2(\tau_0)) \\
= \;& T\Big(\frac{\sum_{t=[\hat\tau_T T]+1}^{T} y_t y_{t-1}}{\sum_{t=[\hat\tau_T T]+1}^{T} y_t^2} - \frac{\sum_{t=[\tau_0 T]+1}^{T} y_t y_{t-1}}{\sum_{t=[\tau_0 T]+1}^{T} y_t^2}\Big)
\end{aligned}$$

$$
\begin{aligned}
&= I\{\hat\tau_T \le \tau_0\} T\Big( -\frac{\sum_{t=[\hat\tau_T T]+1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=[\hat\tau_T T]+1}^{T} y_{t-1}^2} \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} + \frac{\sum_{t=[\hat\tau_T T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=[\hat\tau_T T]+1}^{T} y_{t-1}^2} \\
&\qquad + (\beta_{1T} - \beta_2) \frac{\sum_{t=[\hat\tau_T T]+1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=[\hat\tau_T T]+1}^{T} y_{t-1}^2} \Big) \\
&\quad + I\{\hat\tau_T > \tau_0\} T\Big( \frac{\sum_{t=[\tau_0 T]+1}^{[\hat\tau_T T]} y_{t-1}^2}{\sum_{t=[\hat\tau_T T]+1}^{T} y_{t-1}^2} \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} - \frac{\sum_{t=[\tau_0 T]+1}^{[\hat\tau_T T]} y_{t-1}\varepsilon_t}{\sum_{t=[\hat\tau_T T]+1}^{T} y_{t-1}^2} \Big) \\
&= I\{\hat\tau_T \le \tau_0\} T\Big( O_p(\frac{k_T^2 l(\eta_T)}{T^2 l(\eta_T)}) O_p(\frac{1}{T}) + O_p(\frac{k_T l(\eta_T)}{T^2 l(\eta_T)}) + O_p(\frac{k_T^2 l(\eta_T)}{k_T T^2 l(\eta_T)}) \Big) \\
&\quad + I\{\hat\tau_T > \tau_0\} T\Big( O_p(\frac{k_T^2 l(\eta_T)}{T^2 l(\eta_T)}) O_p(\frac{1}{T}) + O_p(\frac{k_T l(\eta_T)}{T^2 l(\eta_T)}) \Big) \\
&= o_p(1),
\end{aligned}
$$

which implies

$$
\sqrt{\frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{l(\eta_T)}} (\hat\beta_2(\hat\tau_T) - \hat\beta_2(\tau_0)) = o_p(1)
$$

by Lemma A.2. Thus, $\hat\beta_2(\hat\tau_T)$ and $\hat\beta_2(\tau_0)$ also have the same asymptotic distribution. Applying Lemma A.2 again, we have

$$
\sqrt{\frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{l(\eta_T)}} (\hat\beta_2(\tau_0) - \beta_2) = \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sqrt{l(\eta_T) \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}} \Rightarrow \frac{(\overline{W}(1) - \overline{W}(\tau_0))^2 - (1-\tau_0)}{2\sqrt{\int_{\tau_0}^1 (\overline{W}(s) - \overline{W}(\tau_0))^2 ds}},
$$

which immediately implies that

$$
t_2 = \sqrt{\frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{l(\eta_T)}} (\hat\beta_2(\hat\tau_T) - \beta_2) \Rightarrow \frac{(\overline{W}(1) - \overline{W}(\tau_0))^2 - (1-\tau_0)}{2\sqrt{\int_{\tau_0}^1 (\overline{W}(s) - \overline{W}(\tau_0))^2 ds}}.
$$

From the properties of Brownian motion and applying the change of variables, it is trivial that

$$
\frac{(\overline{W}(1) - \overline{W}(\tau_0))^2 - (1-\tau_0)}{2\sqrt{\int_{\tau_0}^1 (\overline{W}(s) - \overline{W}(\tau_0))^2 ds}} \overset{d}{=} \frac{\overline{W}^2(1) - 1}{2\sqrt{\int_0^1 \overline{W}^2(s) ds}}.
$$

Hence, (1.5) holds. □

## Appendix B: Proof of Theorem 1.2

The following six lemmas are used in the proof of Theorem 1.2, and their proofs can be found in the online supplementary materials.

**Lemma B.1** *Let $\{y_t, t \geq 1\}$ be generated by Model (1.1) with $\beta_1 = 1$. Then under assumptions C1-C4, the following results hold jointly:*
*(a)* $\frac{1}{Tl(\eta_T)} \sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t \Rightarrow \frac{1}{2}(W^2(\tau_0) - \tau_0)$,
*(b)* $\frac{1}{T^2 l(\eta_T)} \sum_{t=1}^{[\tau_0 T]} y_{t-1}^2 \Rightarrow \int_0^{\tau_0} W^2(s)ds$.

**Lemma B.2** *Let $\beta_2 = \beta_{2T} = 1 - c/k_T$ with $c > 0$. Then under assumptions C1-C4, we have, for any $\tau_0 < s \leq 1$,*

$$\frac{1}{\sqrt{k_T l(\eta_T)}} \sum_{t=[\tau_0 T]+1}^{[sT]} \beta_{2T}^{[sT]-t} \varepsilon_t \Rightarrow N(0, \frac{1}{2c}).$$

**Lemma B.3** *Let $\{y_t, t \geq 1\}$ be generated by Model (1.1), where $\beta_1 = 1$ and $\beta_2 = \beta_{2T} = 1 - c/k_T$ with $c > 0$. Then under assumptions C1-C4, the following results hold jointly:*
*(a)* $\frac{1}{\sqrt{Tk_T} l(\eta_T)} \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t \Rightarrow \widetilde{W}(\frac{1}{2c})$,
*(b)* $\frac{1}{Tk_T l(\eta_T)} \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2 \Rightarrow \frac{1}{2c}(W^2(\tau_0) + 1 - \tau_0)$.

**Lemma B.4** *Let $\{y_t, t \geq 1\}$ be generated according to Model (1.1), where $\beta_1 = 1$ and $\beta_2 = \beta_{2T} = 1 - c/k_T$ with $c > 0$. Then under assumptions C1-C4 with $k_T = O(\sqrt{T})$, we have*

$$\begin{cases} A_1 = \dfrac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} = o_p(1/k_T), \\ A_2 = \sup\limits_{m \in D_{1T}} \dfrac{\sum_{t=m+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=m+1}^{[\tau_0 T]} y_{t-1}^2} = o_p(1/k_T), \\ A_3 = \sup\limits_{m \in D_{1T}} \left| \dfrac{\sum_{t=m+1}^{T} y_{t-1}^2}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2 \sum_{t=m+1}^{[\tau_0 T]} y_{t-1}^2} \Lambda_T(\dfrac{m}{T}) \right| = o_p(1/k_T^2), \\ A_4 = \dfrac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} = o_p(1/k_T), \\ A_5 = \sup\limits_{m \in D_{2T}} \dfrac{\sum_{t=[\tau_0 T]+1}^{m} y_{t-1}\varepsilon_t}{\sum_{t=[\tau_0 T]+1}^{m} y_{t-1}^2} = o_p(1/k_T), \\ A_6 = \sup\limits_{m \in D_{2T}} \left| \dfrac{\sum_{t=1}^{m} y_{t-1}^2}{\sum_{t=[\tau_0 T]+1}^{m} y_{t-1}^2 \sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} \Lambda_T(\dfrac{m}{T}) \right| = o_p(1/k_T^2), \end{cases} \tag{B.1}$$

*where*

$$\Lambda_T(\frac{m}{T})$$
$$= \frac{\left(\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t\right)^2}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} - \frac{\left(\sum_{t=1}^{m} y_{t-1}\varepsilon_t\right)^2}{\sum_{t=1}^{m} y_{t-1}^2} + \frac{\left(\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t\right)^2}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} - \frac{\left(\sum_{t=m+1}^{T} y_{t-1}\varepsilon_t\right)^2}{\sum_{t=m+1}^{T} y_{t-1}^2},$$

*and*

$$\begin{cases} D_{1T} = \{m : m \in Z_T, m < [\tau_0 T] - M_T\}, \\ D_{2T} = \{m : m \in Z_T, m > [\tau_0 T] + M_T\} \end{cases}$$

*with $M_T > 0$ such that $M_T \to \infty$ arbitrary slowly, and $Z_T$ denotes the set $\{0, 1, 2, \cdots, T\}$.*

**Lemma B.5** *Let $\{y_t, t \geq 1\}$ be generated according to Model (1.1), where $\beta_1 = 1$ and $\beta_2 = \beta_{2T} = 1 - c/k_T$ with $c > 0$. Then under assumptions C1-C4 with $k_T = o(\sqrt{T})$, we have, for any fixed integer $m \geq 0$,*

*(a)* $\frac{k_T^2}{Tl(\eta_T)}\Big(RSS_T(\tau_0 - \frac{m}{T}) - RSS_T(\tau_0)\Big) \Rightarrow c^2 m W^2(\tau_0)$,

*(b)* $\frac{k_T^2}{Tl(\eta_T)}\Big(RSS_T(\tau_0 + \frac{m}{T}) - RSS_T(\tau_0)\Big) \Rightarrow c^2 m W^2(\tau_0)$.

**Lemma B.6** *Let $\{y_t, t \geq 1\}$ be generated according to Model (1.1), where $\beta_1 = 1$ and $\beta_2 = \beta_{2T} = 1 - c/k_T$ with $c > 0$. Then under assumptions C1-C4 with $\sqrt{T} = o(k_T)$, we have*

*(a)* $\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}^2 = O_p(k_T^2 l(\eta_T))$ *and* $\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t = O_p(k_T l(\eta_T))$ *when* $\hat{\tau}_T \leq \tau_0$,

*(b)* $\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2 = O_p(k_T^2 l(\eta_T))$ *and* $\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}\varepsilon_t = O_p(k_T l(\eta_T))$ *when* $\hat{\tau}_T > \tau_0$.

**Proof of Theorem 1.2**. To derive the first and second parts of Theorem 1.2(a), i.e., to prove $P(\hat{k} \neq k_0) \to 0$ when $k_T = o(\sqrt{T})$ and $|\hat{k} - k_0| = O_p(1)$ when $k_T \asymp \sqrt{T}$, we prove

$$|\hat{\tau}_T - \tau_0| = O_p(1/T) \quad \text{when} \quad k_T = O(\sqrt{T}) \tag{B.2}$$

first. According to the proof of Theorem 3 in Chong (2001), it is sufficient to show that

$$P(\lambda_T^2 + 2\lambda_T A_1 - 2\lambda_T A_2 - A_3 < 0) + P(\lambda_T^2 - 2\lambda_T A_4 + 2\lambda_T A_5 - A_6 < 0) \to 0,$$

where $\lambda_T = \beta_2 - \beta_1 = -c/k_T$, and the definitions of $A_1, \cdots, A_6$ can be found in (B.1). Since $\lambda_T^2 > 0$, it suffices to prove that

$$A_i = o_p(1/k_T), \quad i = 1, 2, 4, 5 \quad \text{and} \quad A_j = o_p(1/k_T^2), \quad j = 3, 6, \quad \text{when } k_T = O(\sqrt{T}).$$

This has been proved in Lemma B.4. Therefore, (B.2) is verified. This implies the second part of Theorem 1.2(a).

To prove the first part of Theorem 1.2(a), i.e., to prove $P(\hat{k} \neq k_0) \to 0$ when $k_T = o(\sqrt{T})$, it is noted that $|\hat{\tau}_T - \tau_0| = O_p(1/T)$, and for any $\eta > 0$, there exists a positive integer $M$ such that

$$P(\hat{k} \neq k_0) \;=\; P(|\hat{k} - k_0| > M) + P(|\hat{k} - k_0| \leq M, \hat{k} \neq k_0)$$

$$
\begin{aligned}
\leq \quad & \eta + \sum_{m=1}^{M} P\Big(\frac{k_T^2}{Tl(\eta_T)}\big(RSS_T(\tau_0 - \frac{m}{T}) - RSS_T(\tau_0)\big) < 0\Big) \\
& + \sum_{m=1}^{M} P\Big(\frac{k_T^2}{Tl(\eta_T)}\big(RSS_T(\tau_0 + \frac{m}{T}) - RSS_T(\tau_0)\big) < 0\Big).
\end{aligned}
$$

Applying Lemma B.5 to the above inequality, one immediately has $P(\hat{k} \neq k_0) \to 0$ due to the finiteness of $M$ and arbitrariness of $\eta$.

To prove the third part of Theorem 1.2(a), i.e., the limiting distribution of $\hat{\tau}_T$ when $\sqrt{T} = o(k_T)$, we follow Appendix K in Chong (2001) and let $g(T) = k_T^2/(c^2T)$. Note that (1.6) is a special case of the limiting distribution of $\hat{\tau}_T$ in Theorem 4 in Chong (2001) when $\sqrt{T} = o(k_T)^{\dagger}$, the moment conditions $E(y_0^2) < \infty$ and $E(\varepsilon_t^4) < \infty$ in his paper are satisfied and $c = 1$. One can apply the truncation technique (A.1) to weaken the conditions $E(y_0^2) < \infty$ and $E(\varepsilon_t^4) < \infty$ and accommodate assumptions C1 and C3. The details are omitted for brevity.

To find the limiting distribution of $t_1$, note that we have proven that $|\hat{\tau}_T - \tau_0| = o_p(1/T)$ when $k_T = o(\sqrt{T})$, $|\hat{\tau}_T - \tau_0| = O_p(1/T)$ when $k_T \asymp \sqrt{T}$ and $|\hat{\tau}_T - \tau_0| = O_p(k_T^2/T^2)$ when $\sqrt{T} = o(k_T)$, hence we study the limiting distribution of $t_1$ under the above three cases separately.

When $\sqrt{T} = o(k_T)$, applying Lemma B.1 and Lemma B.6 and following Appendix G in Chong (2001), we have

$$
\begin{aligned}
& T(\hat{\beta}_1(\hat{\tau}_T) - \hat{\beta}_1(\tau_0)) \\
= \quad & T\Bigg(\frac{\sum_{t=1}^{[\hat{\tau}_T T]} y_t y_{t-1}}{\sum_{t=1}^{[\hat{\tau}_T T]} y_t^2} - \frac{\sum_{t=1}^{[\tau_0 T]} y_t y_{t-1}}{\sum_{t=1}^{[\tau_0 T]} y_t^2}\Bigg) \\
= \quad & I\{\hat{\tau}_T \leq \tau_0\} T\Bigg(\frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2} \frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} - \frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2}\Bigg) \\
& + I\{\hat{\tau}_T > \tau_0\} T\Bigg( - \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2} \frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} + \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2}
\end{aligned}
$$

$^{\dagger}$In Theorem 4 in Chong (2001), the condition $T^{3/4}(1 - \beta_{2T}) \to \infty$, that is $k_T = o(T^{3/4})$ in our case, is imposed so as to derive the limiting distribution of $\hat{\tau}_T$. By checking his proof carefully, this condition is imposed only for showing that $\frac{g(T)}{\sqrt{T}} \sum_{t=0}^{[|v|g(T)]-1} \left(\frac{1}{\sqrt{g(T)}} \sum_{i=0}^{t} \varepsilon_{[\tau_0 T]-i}\varepsilon_{[\tau_0 T]-t}\right) \frac{1}{g(T)} = o_p(1)$ when $\varepsilon_t$'s have finite 4th moments (see page 153 in Chong (2001)). However, since $\sum_{t=0}^{[|v|g(T)]-1} \left(\frac{1}{\sqrt{g(T)}} \sum_{i=0}^{t} \varepsilon_{[\tau_0 T]-i}\varepsilon_{[\tau_0 T]-t}\right) \frac{1}{\sqrt{g(T)}} = \sum_{t=0}^{[|v|g(T)]-1} \left(\frac{1}{\sqrt{g(T)}} \sum_{i=0}^{t-1} \varepsilon_{[\tau_0 T]-i}\right) \frac{\varepsilon_{[\tau_0 T]-t}}{\sqrt{g(T)}} + O_p(1) = O_p(1)$, the condition $\sqrt{\frac{g(T)}{T}} = o(1)$ is sufficient to achieve the goal. Note that $g(T) \asymp k_T^2/T$, hence, the condition $k_T = o(T)$ is sufficient. Therefore, the condition $T^{3/4}(1 - \beta_{2T}) \to \infty$ is not needed in Theorem 4 in Chong (2001).

$$
\begin{aligned}
& \quad + (\beta_{2T} - \beta_1) \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2} \Big) \\
&= I\{\hat{\tau}_T \le \tau_0\} T \Big( O_p(\frac{k_T^2 l(\eta_T)}{T^2 l(\eta_T)}) O_p(\frac{1}{T}) + O_p(\frac{k_T l(\eta_T)}{T^2 l(\eta_T)}) \Big) \\
& \quad + I\{\hat{\tau}_T > \tau_0\} T \Big( O_p(\frac{k_T^2 l(\eta_T)}{T^2 l(\eta_T)}) O_p(\frac{1}{T}) + O_p(\frac{k_T l(\eta_T)}{T^2 l(\eta_T)}) + O_p(\frac{k_T^2 l(\eta_T)}{k_T T^2 l(\eta_T)}) \Big) \\
&= o_p(1), \qquad\qquad (B.3)
\end{aligned}
$$

which implies

$$
\sqrt{\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{l(\eta_T)}} (\hat{\beta}_1(\hat{\tau}_T) - \hat{\beta}_1(\tau_0)) = o_p(1)
$$

by Lemma B.1. Thus, $\hat{\beta}_1(\hat{\tau}_T)$ and $\hat{\beta}_1(\tau_0)$ have the same asymptotic distribution. We then invoke Lemma B.1 again to obtain

$$
\sqrt{\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{l(\eta_T)}} (\hat{\beta}_1(\tau_0) - \beta_1) = \frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sqrt{l(\eta_T) \sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}} \Rightarrow \frac{W^2(\tau_0) - \tau_0}{2\sqrt{\int_0^{\tau_0} W^2(s) ds}},
$$

which immediately leads to

$$
t_1 = \sqrt{\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{l(\eta_T)}} (\hat{\beta}_1(\hat{\tau}_T) - \beta_1) \Rightarrow \frac{W^2(\tau_0) - \tau_0}{2\sqrt{\int_0^{\tau_0} W^2(s) ds}}. \qquad (B.4)
$$

From the properties of Brownian motions and applying a change of variables, it follows that

$$
\frac{W^2(\tau_0) - \tau_0}{2\sqrt{\int_0^{\tau_0} W^2(s) ds}} \stackrel{d}{=} \frac{W^2(1) - 1}{2\sqrt{\int_0^{1} W^2(s) ds}}.
$$

Hence, (1.7) is proved.

For the case where $k_T \asymp \sqrt{T}$, note that since $|\hat{\tau}_T - \tau_0| = O_p(1/T)$, it is trivial that

$$
\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}^2 = O_p(T l(\eta_T)), \qquad \sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t = O_p(\sqrt{T} l(\eta_T)) \qquad (B.5)
$$

and

$$
\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2 = O_p(T l(\eta_T)), \qquad \sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}\varepsilon_t = O_p(\sqrt{T} l(\eta_T)). \qquad (B.6)
$$

Consequently, similar to (B.3), we have

$$
\begin{aligned}
& T(\hat{\beta}_1(\hat{\tau}_T) - \hat{\beta}_1(\tau_0)) \\
&= I\{\hat{\tau}_T \le \tau_0\} T \Big( \frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2} \frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} - \frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2} \Big) \\
& \quad + I\{\hat{\tau}_T > \tau_0\} T \Big( - \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2} \frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} + \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2}
\end{aligned}
$$

$$
\begin{aligned}
& \qquad + (\beta_{2T} - \beta_1) \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2} \Big) \\
= \;& I\{\hat{\tau}_T \le \tau_0\} T \Big( O_p(\frac{T l(\eta_T)}{T^2 l(\eta_T)}) O_p(\frac{1}{T}) + O_p(\frac{\sqrt{T} l(\eta_T)}{T^2 l(\eta_T)}) \Big) \\
& + I\{\hat{\tau}_T > \tau_0\} T \Big( O_p(\frac{T l(\eta_T)}{T^2 l(\eta_T)}) O_p(\frac{1}{T}) + O_p(\frac{\sqrt{T} l(\eta_T)}{T^2 l(\eta_T)}) + O_p(\frac{T l(\eta_T)}{k_T T^2 l(\eta_T)}) \Big) \\
= \;& o_p(1),
\end{aligned}
$$

which means (B.4) still holds when $k_T \asymp \sqrt{T}$.

For the case where $k_T = o(\sqrt{T})$, note that since $P(\hat{k} \neq k_0) \to 0$, following the proof of Theorem 4 in Chong (2001), one can show that (B.4) still holds. The details are omitted.

To find the limiting distribution of $t_2$, we also consider the following three cases where $\sqrt{T} = o(k_T)$, $k_T \asymp \sqrt{T}$ and $k_T = o(\sqrt{T})$ separately.

When $\sqrt{T} = o(k_T)$, applying Lemma B.3 and Lemma B.6 and following Appendix G in Chong (2001), we have

$$
\begin{aligned}
& \sqrt{T k_T} (\hat{\beta}_2(\hat{\tau}_T) - \hat{\beta}_2(\tau_0)) \\
= \;& \sqrt{T k_T} \Big( \frac{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_t y_{t-1}}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_t^2} - \frac{\sum_{t=[\tau_0 T]+1}^{T} y_t y_{t-1}}{\sum_{t=[\tau_0 T]+1}^{T} y_t^2} \Big) \\
= \;& I\{\hat{\tau}_T \le \tau_0\} \sqrt{T k_T} \Big( - \frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2} \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} + \frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2} \\
& \qquad + (\beta_1 - \beta_{2T}) \frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2} \Big) \\
& + I\{\hat{\tau}_T > \tau_0\} \sqrt{T k_T} \Big( \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2} \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} - \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}\varepsilon_t}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2} \Big) \\
= \;& I\{\hat{\tau}_T \le \tau_0\} \sqrt{T k_T} \Big( O_p(\frac{k_T^2 l(\eta_T)}{T k_T l(\eta_T)}) O_p(\frac{1}{\sqrt{T k_T}}) + O_p(\frac{k_T l(\eta_T)}{T k_T l(\eta_T)}) + O_p(\frac{k_T^2 l(\eta_T)}{T k_T^2 l(\eta_T)}) \Big) \\
& + I\{\hat{\tau}_T > \tau_0\} \sqrt{T k_T} \Big( O_p(\frac{k_T^2 l(\eta_T)}{T k_T l(\eta_T)}) O_p(\frac{1}{\sqrt{T k_T}}) + O_p(\frac{k_T l(\eta_T)}{T k_T l(\eta_T)}) \Big) \\
= \;& o_p(1),
\end{aligned}
$$

which implies

$$
\sqrt{\frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{l(\eta_T)}} (\hat{\beta}_2(\hat{\tau}_T) - \hat{\beta}_2(\tau_0)) = o_p(1)
$$

by Lemma B.3. Thus, $\hat{\beta}_2(\hat{\tau}_T)$ and $\hat{\beta}_2(\tau_0)$ have the same asymptotic distribution. Applying Lemma B.3 again, we have

$$
\sqrt{\frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{l(\eta_T)}} (\hat{\beta}_2(\tau_0) - \beta_{2T}) = \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sqrt{l(\eta_T) \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}} \Rightarrow \frac{\widetilde{W}(1)}{\sqrt{W^2(\tau_0) + 1 - \tau_0}},
$$

which immediately leads to

$$t_2 = \sqrt{\frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{l(\eta_T)}}(\hat{\beta}_2(\hat{\tau}_T) - \beta_{2T}) \Rightarrow \frac{\widetilde{W}(1)}{\sqrt{W^2(\tau_0) + 1 - \tau_0}}. \tag{B.7}$$

When $k_T \asymp \sqrt{T}$, applying (B.5) and (B.6), we have

$$\begin{aligned}
& \sqrt{Tk_T}(\hat{\beta}_2(\hat{\tau}_T) - \hat{\beta}_2(\tau_0)) \\
= \;& I\{\hat{\tau}_T \le \tau_0\}\sqrt{Tk_T}\Big( - \frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2} \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} + \frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2} \\
& \qquad + (\beta_1 - \beta_{2T}) \frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2}\Big) \\
& + I\{\hat{\tau}_T > \tau_0\}\sqrt{Tk_T}\Big( \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2} \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} - \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}\varepsilon_t}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2}\Big) \\
= \;& I\{\hat{\tau}_T \le \tau_0\}\sqrt{Tk_T}\Big(O_p(\frac{Tl(\eta_T)}{Tk_T l(\eta_T)})O_p(\frac{1}{\sqrt{Tk_T}}) + O_p(\frac{\sqrt{T}l(\eta_T)}{Tk_T l(\eta_T)}) + O_p(\frac{Tl(\eta_T)}{Tk_T^2 l(\eta_T)})\Big) \\
& + I\{\hat{\tau}_T > \tau_0\}\sqrt{Tk_T}\Big(O_p(\frac{Tl(\eta_T)}{Tk_T l(\eta_T)})O_p(\frac{1}{\sqrt{Tk_T}}) + O_p(\frac{\sqrt{T}l(\eta_T)}{Tk_T l(\eta_T)})\Big) \\
= \;& o_p(1),
\end{aligned}$$

which means that (B.7) still holds.

When $k_T = o(\sqrt{T})$, applying Lemma B.5 and following the proof of Theorem 4 in Chong (2001), one can show that (B.7) holds. The details are omitted. □

## Appendix C: Proof of Theorem 1.3

Lemma B.1 and the following five lemmas are key ingredients in the proof of Theorem 1.3, and their proofs can be found in the online supplementary materials.

**Lemma C.1** *Suppose assumptions C1 and C2 are fulfilled and $c > 0$, then*
*(a) for any $0 < \tau \le 1$, $\frac{1}{\sqrt{k_T l(\eta_T)}} \sum_{t=1}^{[\tau T]} (1 + \frac{c}{k_T})^{t-1-[\tau T]} \varepsilon_t \Rightarrow X$,*
*(b) for any $0 < \tau \le 1$, $\frac{1}{\sqrt{k_T l(\eta_T)}} \sum_{t=1}^{[\tau T]} (1 + \frac{c}{k_T})^{-t} \varepsilon_t \Rightarrow Y$,*
*where $X$ and $Y$ are independent $N(0, \frac{1}{2c})$ random variables.*

**Lemma C.2** *Let $\{y_t, t \ge 1\}$ be generated according to Model (1.1), where $\beta_1 = 1$ and $\beta_2 = \beta_{2T} = 1 + c/k_T$ with $c > 0$. Then under assumptions C1-C4, the following results hold jointly:*

*(a)* $\frac{\beta_{2T}^{-(T-[\tau_0 T])}}{\sqrt{T}k_T l(\eta_T)} \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t \Rightarrow XW(\tau_0)$,
*(b)* $\frac{\beta_{2T}^{-2(T-[\tau_0 T])}}{Tk_T l(\eta_T)} \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2 \Rightarrow \frac{1}{2c}W^2(\tau_0)$,
*where* $X$ *is a random variable obeying* $N(0, \frac{1}{2c})$ *and is independent of* $W(\tau_0)$.

**Lemma C.3** *Let* $\{y_t, t \geq 1\}$ *be generated according to Model (1.1), where* $\beta_1 = 1$ *and* $\beta_2 = \beta_{2T} = 1 + c/k_T$ *with* $c > 0$. *Then under assumptions C1-C4 with* $k_T = O(\sqrt{T})$, *the results in (B.1) hold.*

**Lemma C.4** *Let* $\{y_t, t \geq 1\}$ *be generated according to Model (1.1), where* $\beta_1 = 1$ *and* $\beta_2 = \beta_{2T} = 1 + c/k_T$ *with* $c > 0$. *Then under assumptions C1-C4 with* $\sqrt{T} = o(k_T)$, *we have*
*(a)* $\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}^2 = O_p(k_T^2 l(\eta_T))$ *and* $\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t = O_p(k_T l(\eta_T))$ *when* $\hat{\tau}_T \leq \tau_0$,
*(b)* $\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2 = O_p(k_T^2 l(\eta_T))$ *and* $\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}\varepsilon_t = O_p(k_T l(\eta_T))$ *when* $\hat{\tau}_T > \tau_0$.

**Lemma C.5** *Let* $\{y_t, t \geq 1\}$ *be generated according to Model (1.1), where* $\beta_1 = 1$ *and* $\beta_2 = \beta_{2T} = 1 + c/k_T$ *with* $c > 0$. *Then under assumptions C1-C4 with* $k_T = o(\sqrt{T})$, *we have for any fixed integer* $m \geq 0$,
*(a)* $\frac{k_T^2}{Tl(\eta_T)}\Big(RSS_T(\tau_0 - \frac{m}{T}) - RSS_T(\tau_0)\Big) \Rightarrow c^2 m W^2(\tau_0)$,
*(b)* $\frac{k_T^2}{Tl(\eta_T)}\Big(RSS_T(\tau_0 + \frac{m}{T}) - RSS_T(\tau_0)\Big) \Rightarrow c^2 m W^2(\tau_0)$.

**Proof of Theorem 1.3**. We first prove the first and second parts of Theorem 1.3(a). Similar to the proof of Theorem 1.2(a), we can prove (B.2) by using the arguments in the proof of Theorem 1.2 and invoking Lemma C.3. Hence, the second part of Theorem 1.3(a) is verified.

To prove the first part of Theorem 1.3(a), i.e., to prove $P(\hat{k} \neq k_0) \to 0$ when $k_T = o(\sqrt{T})$, one can refer to the proof of Theorem 1.2(a) and apply Lemma C.5. The details are not provided here for brevity.

To derive the third part of Theorem 1.3(a), we follow Appendix K in Chong (2001). Since asymptotics in the mildly explosive model are more complex than that in the mildly integrated model, we offer more details concerning the proof. We write $\beta_{2T} = 1 + c/k_T = 1 + 1/(\sqrt{Tg(T)})$, with $g(T) = k_T^2/(c^2 T)$. Then, $g(T) \to \infty$ and $g(T) = o(k_T)$ since $\sqrt{T} = o(k_T)$ and $k_T = o(T)$. We denote

$$\Lambda_T(\tau) = \frac{\Big(\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t\Big)^2}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} - \frac{\Big(\sum_{t=1}^{[\tau T]} y_{t-1}\varepsilon_t\Big)^2}{\sum_{t=1}^{[\tau T]} y_{t-1}^2} + \frac{\Big(\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t\Big)^2}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} - \frac{\Big(\sum_{t=[\tau T]+1}^{T} y_{t-1}\varepsilon_t\Big)^2}{\sum_{t=[\tau T]+1}^{T} y_{t-1}^2}.$$

The following observation is a simple generalization of Proposition A.1 in Phillips and Magdalinos (2007a), hence the proof is omitted:

$$\left(\frac{T}{k_T}\right)^a = o((1+c/k_T)^{bT}), \quad \text{for any } a>0 \text{ and } b>0. \tag{C.1}$$

For $\tau = \tau_0 + \nu g(T)/T$ and $\nu \le 0$, by applying (C.1) and Lemmas B.1 and C.2, we have the following results:

$$\frac{\Lambda_T(\tau)}{l(\eta_T)} = O_p(1), \quad \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{\sum_{t=[\tau T]+1}^{T} y_{t-1}^2} \xrightarrow{p} 1;$$

$$\begin{aligned}
&\left| \frac{(\beta_{2T}-1)}{l(\eta_T)} \frac{\sum_{t=[\tau T]+1}^{[\tau_0 T]} y_{t-1}^2 \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=[\tau T]+1}^{T} y_{t-1}^2} \right| \\
=\ & O_p\left( \frac{Tg(T)\sqrt{Tk_T}\beta_{2T}^{T-[\tau_0 T]}}{k_T(Tg(T) + Tk_T\beta_{2T}^{2(T-[\tau_0 T])})} \right) \\
=\ & O_p\left( \frac{Tg(T)\sqrt{Tk_T}}{Tk_T^2\beta_{2T}^{(T-[\tau_0 T])}} \right) = o_p\left( \left(\frac{T}{k_T}\right)^{1/2} \cdot \frac{1}{\beta_{2T}^{(T-[\tau_0 T])}} \right) \\
=\ & o_p(1);
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{\sqrt{Tg(T)}l(\eta_T)} \sum_{t=0}^{[|\nu|g(T)]-1} y_{[\tau_0 T]-t-1}\varepsilon_{[\tau_0 T]-t} \\
=\ & -\frac{y_{[\tau_0 T]}}{\sqrt{Tl(\eta_T)}} \cdot \frac{1}{\sqrt{g(T)l(\eta_T)}} \sum_{t=0}^{[|\nu|g(T)]-1} (-\varepsilon_{[\tau_0 T]-t}) \cdot (1+o_p(1)) \\
\Rightarrow\ & -W_1(\tau_0)W_1(|\nu|)
\end{aligned}$$

and

$$\frac{1}{Tg(T)l(\eta_T)} \sum_{t=0}^{[|\nu|g(T)]-1} y_{[\tau_0 T]-t-1}^2 \Rightarrow |\nu| W_1^2(\tau_0).$$

One is referred to Appendix K in Chong (2001) for more details.

Then, using equation (B.2) in Chong (2001), we have

$$\begin{aligned}
&\frac{RSS_T(\tau) - RSS_T(\tau_0)}{l(\eta_T)} \\
=\ & -\frac{2(\beta_{2T}-1)}{l(\eta_T)} \sum_{t=[\tau T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t \cdot (1+o_p(1)) + \frac{(\beta_{2T}-1)^2}{l(\eta_T)} \sum_{t=[\tau T]+1}^{[\tau_0 T]} y_{t-1}^2 \cdot (1+o_p(1)) + o_p(1) \\
=\ & -\frac{2(1+o_p(1))}{\sqrt{Tg(T)}l(\eta_T)} \sum_{t=0}^{[|\nu|g(T)]-1} y_{[\tau_0 T]-t-1}\varepsilon_{[\tau_0 T]-t} + \frac{1+o_p(1)}{Tg(T)l(\eta_T)} \sum_{t=0}^{[|\nu|g(T)]-1} y_{[\tau_0 T]-t-1}^2 + o_p(1) \\
\Rightarrow\ & -2W_1(\tau_0)W_1(|\nu|) + |\nu| W_1^2(\tau_0).
\end{aligned} \tag{C.2}$$

For $\tau = \tau_0 + \nu g(T)/T$ and $\nu > 0$, we have

$$\begin{aligned}\sum_{t=[\tau_0 T]+1}^{[\tau_0 T+\nu g(T)]} y_{t-1}^2 &= \sum_{t=[\tau_0 T]+1}^{[\tau_0 T+\nu g(T)]} \Big(\beta_{2T}^{t-1-[\tau_0 T]} y_{[\tau_0 T]} + \sum_{i=[\tau_0 T]+1}^{t-1} \beta_{2T}^{t-1-i}\varepsilon_i\Big)^2 \\ &= y_{[\tau_0 T]}^2 \sum_{t=[\tau_0 T]+1}^{[\tau_0 T+\nu g(T)]} \beta_{2T}^{2(t-1-[\tau_0 T])} + \sum_{t=[\tau_0 T]+1}^{[\tau_0 T+\nu g(T)]} \Big(\sum_{i=[\tau_0 T]+1}^{t-1} \beta_{2T}^{t-1-i}\varepsilon_i\Big)^2 \\ &\quad +2y_{[\tau_0 T]} \sum_{t=[\tau_0 T]+1}^{[\tau_0 T+\nu g(T)]} \beta_{2T}^{t-1-[\tau_0 T]} \sum_{i=[\tau_0 T]+1}^{t-1} \beta_{2T}^{t-1-i}\varepsilon_i. \qquad (C.3)\end{aligned}$$

Note that

$$\sum_{t=[\tau_0 T]+1}^{[\tau_0 T+\nu g(T)]} \beta_{2T}^{2(t-1-[\tau_0 T])} = O\Big(\Big|\frac{k_T(1-\beta_{2T}^{2[\nu g(T)]})}{k_T(1-\beta_{2T}^2)}\Big|\Big) = O\Big(\Big|\frac{k_T\cdot g(T)/k_T}{k_T(1-\beta_{2T}^2)}\Big|\Big) = O(g(T)),$$

which implies

$$y_{[\tau_0 T]}^2 \sum_{t=[\tau_0 T]+1}^{[\tau_0 T+\nu g(T)]} \beta_{2T}^{2(t-1-[\tau_0 T])} = O_p(Tg(T)l(\eta_T)). \qquad (C.4)$$

In addition, it can be shown that

$$\begin{aligned}E\left(\sum_{t=[\tau_0 T]+1}^{[\tau_0 T+\nu g(T)]} \Big(\sum_{i=[\tau_0 T]+1}^{t-1} \beta_{2T}^{t-1-i}\varepsilon_i^{(1)}\Big)^2\right) &= \sum_{t=[\tau_0 T]+1}^{[\tau_0 T+\nu g(T)]} \sum_{i=[\tau_0 T]+1}^{t-1} \beta_{2T}^{2(t-1-i)} l(\eta_T)\cdot(1+o(1)) \\ &= O\left(\frac{g(T)}{\beta_{2T}^2-1} l(\eta_T)\right) + O\left(\frac{(1-\beta_{2T}^{\nu g(T)})l(\eta_T)}{(1-\beta_{2T}^2)^2}\right) \\ &= O(k_T g(T) l(\eta_T)) + O(k_T^2 l(\eta_T)\cdot g(T)/k_T) \\ &= O(k_T g(T) l(\eta_T)),\end{aligned}$$

implying

$$\sum_{t=[\tau_0 T]+1}^{[\tau_0 T+\nu g(T)]} \Big(\sum_{i=[\tau_0 T]+1}^{t-1} \beta_{2T}^{t-1-i}\varepsilon_i\Big)^2 = O_p(k_T g(T) l(\eta_T)) = o_p(Tg(T)l(\eta_T)). \qquad (C.5)$$

Combining (C.3), (C.4) and (C.5) and applying the Cauchy-Schwarz inequality, we have

$$\sum_{t=[\tau_0 T]+1}^{[\tau_0 T+\nu g(T)]} y_{t-1}^2 = O_p(Tg(T)l(\eta_T)). \qquad (C.6)$$

The above arguments also imply that

$$\max_{1\le t\le[\nu g(T)]-1} |y_{[\tau_0 T]+t} - y_{[\tau_0 T]}| = o_p(\sqrt{Tl(\eta_T)}).$$

Then, applying (C.6), we have the following results:

$$\frac{\Lambda_T(\tau)}{l(\eta_T)} = O_p(1), \quad \frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=1}^{[\tau T]} y_{t-1}^2} \xrightarrow{p} 1;$$

$$\left|\frac{(\beta_{2T}-1)}{l(\eta_T)}\cdot\frac{-\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t \sum_{t=[\tau_0 T]+1}^{[\tau T]} y_{t-1}^2}{\sum_{t=1}^{[\tau T]} y_{t-1}^2}\right| = O_p\Big(\frac{T^2 g(T)}{k_T(T^2+Tg(T))}\Big) = O_p\Big(\frac{g(T)}{k_T}\Big) = o_p(1)$$

by Lemma B.1 and $g(T)=o(k_T)$.

Now, making use of equation (B.4) in Chong (2001), we have

$$\begin{aligned}
&\frac{RSS_T(\tau)-RSS_T(\tau_0)}{l(\eta_T)} \\
= \ & \frac{2(\beta_{2T}-1)}{l(\eta_T)}\sum_{t=[\tau_0 T]+1}^{[\tau T]} y_{t-1}\varepsilon_t\cdot(1+o_p(1)) + \frac{(\beta_{2T}-1)^2}{l(\eta_T)}\sum_{t=[\tau_0 T]+1}^{[\tau T]} y_{t-1}^2\cdot(1+o_p(1)) + o_p(1) \\
= \ & \frac{2(1+o_p(1))}{\sqrt{Tg(T)}l(\eta_T)}\sum_{t=0}^{[|\nu|g(T)]-1} y_{[\tau_0 T]+t}\varepsilon_{[\tau_0 T]+t+1} + \frac{1+o_p(1)}{Tg(T)l(\eta_T)}\sum_{t=0}^{[|\nu|g(T)]-1} y_{[\tau_0 T]+t+1}^2 + o_p(1) \\
= \ & \frac{-2y_{[\tau_0 T]}}{\sqrt{Tl(\eta_T)}}\cdot\frac{1+o_p(1)}{\sqrt{g(T)l(\eta_T)}}\sum_{t=0}^{[|\nu|g(T)]-1}(-\varepsilon_{[\tau_0 T]+t+1}) + \nu\frac{y_{[\tau_0 T]}^2}{Tl(\eta_T)}\cdot(1+o_p(1)) + o_p(1) \\
\Rightarrow \ & -2W_1(\tau_0)W_2(\nu) + \nu W_1^2(\tau_0). \qquad \text{(C.7)}
\end{aligned}$$

Finally, applying the continuous mapping theorem for argmax/argmin functionals (cf. Kim and Pollard (1990)), it follows from (C.2) and (C.7) that

$$\begin{aligned}
\frac{c^2T^2}{k_T^2}(\hat\tau_T-\tau_0) &= \frac{T}{g(T)}(\hat\tau_T-\tau_0) = \hat\nu \\
&= \arg\min_{\nu\in R}\{RSS_T(\tau)-RSS_T(\tau_0)\} \\
&= \arg\min_{\nu\in R}\left\{\frac{RSS_T(\tau)-RSS_T(\tau_0)}{l(\eta_T)}\right\} \\
&\Rightarrow \arg\min_{\nu\in R}\left\{-2W_1^2(\tau_0)\cdot\Big(\frac{W^*(\nu)}{W_1(\tau_0)}-\frac{|\nu|}{2}\Big)\right\} = \arg\max_{\nu\in R}\left\{\frac{W^*(\nu)}{W_1(\tau_0)}-\frac{|\nu|}{2}\right\},
\end{aligned}$$

where $W^*(\nu)$ is a two-sided Brownian motion on $R$ defined to be $W^*(\nu)=W_1(-\nu)$ for $\nu\le 0$ and $W^*(\nu)=W_2(\nu)$ for $\nu>0$.

To prove Theorem 1.3(b), note that we have shown that $|\hat\tau_T-\tau_0| = o_p(1/T)$ when $k_T = o(\sqrt{T})$, $|\hat\tau_T-\tau_0| = O_p(1/T)$ when $k_T \asymp \sqrt{T}$ and $|\hat\tau_T-\tau_0| = O_p(k_T^2/T^2)$ when $\sqrt{T} = o(k_T)$, hence we study the limiting distribution of $t_1$ under the above three cases separately.

Consider the case where $\sqrt{T}=o(k_T)$ first. In this case, applying Lemmas B.1 and C.4 and following Appendix G in Chong (2001), we have

$$\begin{aligned}
&T(\hat\beta_1(\hat\tau_T)-\hat\beta_1(\tau_0)) \\
= \ & T\Big(\frac{\sum_{t=1}^{[\hat\tau_T T]} y_t y_{t-1}}{\sum_{t=1}^{[\hat\tau_T T]} y_t^2} - \frac{\sum_{t=1}^{[\tau_0 T]} y_t y_{t-1}}{\sum_{t=1}^{[\tau_0 T]} y_t^2}\Big) \\
= \ & I\{\hat\tau_T\le\tau_0\}T\Big(\frac{\sum_{t=[\hat\tau_T T]+1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=1}^{[\hat\tau_T T]} y_{t-1}^2}\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} - \frac{\sum_{t=[\hat\tau_T T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\hat\tau_T T]} y_{t-1}^2}\Big)
\end{aligned}$$

$$
\begin{aligned}
&+I\{\hat{\tau}_T > \tau_0\}T\Big( - \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2} \frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} + \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2} \\
&\qquad + (\beta_{2T} - \beta_1)\frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2}\Big) \\
= \; & I\{\hat{\tau}_T \le \tau_0\}T\Big(O_p(\frac{k_T^2 l(\eta_T)}{T^2 l(\eta_T)})O_p(\frac{1}{T}) + O_p(\frac{k_T l(\eta_T)}{T^2 l(\eta_T)})\Big) \\
&+I\{\hat{\tau}_T > \tau_0\}T\Big(O_p(\frac{k_T^2 l(\eta_T)}{T^2 l(\eta_T)})O_p(\frac{1}{T}) + O_p(\frac{k_T l(\eta_T)}{T^2 l(\eta_T)}) + O_p(\frac{k_T^2 l(\eta_T)}{k_T T^2 l(\eta_T)})\Big) \\
= \; & o_p(1), \qquad\qquad (C.8)
\end{aligned}
$$

implying

$$
\sqrt{\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{l(\eta_T)}}(\hat{\beta}_1(\hat{\tau}_T) - \hat{\beta}_1(\tau_0)) = o_p(1)
$$

by Lemma B.1. Thus, $\hat{\beta}_1(\hat{\tau}_T)$ and $\hat{\beta}_1(\tau_0)$ have the same asymptotic distribution. Then, applying Lemma B.1 again, we have

$$
\sqrt{\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{l(\eta_T)}}(\hat{\beta}_1(\tau_0) - \beta_1) = \frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sqrt{l(\eta_T)\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}} \Rightarrow \frac{W^2(\tau_0) - \tau_0}{2\sqrt{\int_0^{\tau_0} W^2(t)dt}} \overset{d}{=} \frac{W^2(1) - 1}{2\sqrt{\int_0^1 W^2(s)ds}},
$$

implying

$$
t_1 = \sqrt{\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{l(\eta_T)}}(\hat{\beta}_1(\hat{\tau}_T) - \beta_1) \Rightarrow \frac{W^2(1) - 1}{2\sqrt{\int_0^1 W^2(s)ds}}. \qquad (C.9)
$$

When $k_T \asymp \sqrt{T}$, since $|\hat{k} - k_0| = O_p(1)$, one can follow the proofs in Theorem 1.2 to show that $T(\hat{\beta}_1(\hat{\tau}_T) - \hat{\beta}_1(\tau_0)) = o_p(1)$ still holds by using (B.5) and (B.6). The details are omitted here. Hence, (C.9) still holds when $k_T \asymp \sqrt{T}$.

When $k_T = o(\sqrt{T})$, since $P(\hat{k} \neq k_0) \to 0$, one can follow the lines in the proof of Theorem 4 in Chong (2001) to show that (C.9) still holds. The details are also omitted.

Next, we shall prove Theorem 1.3(c). Similarly, when $\sqrt{T} = o(k_T)$, making use of Lemmas C.2 and C.4 and following Appendix G in Chong (2001), we have

$$
\begin{aligned}
& \beta_{2T}^{T-[\tau_0 T]}\sqrt{Tk_T}(\hat{\beta}_2(\hat{\tau}_T) - \hat{\beta}_2(\tau_0)) \\
= \; & \beta_{2T}^{T-[\tau_0 T]}\sqrt{Tk_T}\Big(\frac{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_t y_{t-1}}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_t^2} - \frac{\sum_{t=[\tau_0 T]+1}^{T} y_t y_{t-1}}{\sum_{t=[\tau_0 T]+1}^{T} y_t^2}\Big) \\
= \; & I\{\hat{\tau}_T \le \tau_0\}\beta_{2T}^{T-[\tau_0 T]}\sqrt{Tk_T}\Big( - \frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2} \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} + \frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2} \\
&\qquad + (\beta_1 - \beta_{2T})\frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2}\Big)
\end{aligned}
$$

$$
\begin{aligned}
& +I\{\hat{\tau}_T > \tau_0\}\beta_{2T}^{T-[\tau_0 T]}\sqrt{Tk_T}\Big(\frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2}\frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} - \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}\varepsilon_t}{\sum_{t=[\hat{\tau}_T T]+1}^{T} y_{t-1}^2}\Big) \\
= \ & I\{\hat{\tau}_T \le \tau_0\}\beta_{2T}^{T-[\tau_0 T]}\sqrt{Tk_T}\Big(O_p(\frac{k_T^2 l(\eta_T)}{\beta_{2T}^{2(T-[\tau_0 T])}Tk_T l(\eta_T)})O_p(\frac{1}{\beta_{2T}^{T-[\tau_0 T]}\sqrt{Tk_T}l(\eta_T)}) \\
& + O_p(\frac{k_T l(\eta_T)}{\beta_{2T}^{2(T-[\tau_0 T])}Tk_T l(\eta_T)}) + O_p(\frac{k_T^2 l(\eta_T)}{\beta_{2T}^{2(T-[\tau_0 T])}Tk_T^2 l(\eta_T)})\Big) \\
& +I\{\hat{\tau}_T > \tau_0\}\beta_{2T}^{T-[\tau_0 T]}\sqrt{Tk_T}\Big(O_p(\frac{k_T^2 l(\eta_T)}{\beta_{2T}^{2(T-[\tau_0 T])}Tk_T l(\eta_T)})O_p(\frac{1}{\beta_{2T}^{T-[\tau_0 T]}\sqrt{Tk_T}}) \\
& + O_p(\frac{k_T l(\eta_T)}{\beta_{2T}^{2(T-[\tau_0 T])}Tk_T l(\eta_T)})\Big) \\
= \ & o_p(1)
\end{aligned}
$$

by using (C.1), which implies

$$\sqrt{\frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{l(\eta_T)}}(\hat{\beta}_2(\hat{\tau}_T) - \hat{\beta}_2(\tau_0)) = o_p(1)$$

by Lemma C.2. Thus, $\hat{\beta}_2(\hat{\tau}_T)$ and $\hat{\beta}_2(\tau_0)$ have the same asymptotic distribution. Applying Lemma C.2 again, we have

$$\sqrt{\frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{l(\eta_T)}}(\hat{\beta}_2(\tau_0) - \beta_{2T}) = \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sqrt{l(\eta_T)\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}} \Rightarrow N(0,1),$$

which implies

$$t_2 = \sqrt{\frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{l(\eta_T)}}(\hat{\beta}_2(\hat{\tau}_T) - \beta_{2T}) \Rightarrow N(0,1). \tag{C.10}$$

The result (C.10) when $k_T \asymp \sqrt{T}$ or $k_T = o(\sqrt{T})$ can also be proved by similar arguments in the proofs of Theorem 1.2, and the details are omitted. □

## Appendix D: Proof of Theorem 1.4

The following four lemmas are needed in the proof of Theorem 1.4, and their proofs can be found in the online supplementary materials.

**Lemma D.1** *Let $\{y_t, t \ge 1\}$ be generated according to Model (1.1), where $\beta_1 = \beta_{1T} = 1 + c/k_T$ with $c > 0$. Then under assumptions C1-C4, the following results hold jointly:*
*(a)* $\frac{\beta_{1T}^{-[\tau_0 T]}}{k_T l(\eta_T)}\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t \Rightarrow XY$,
*(b)* $\frac{\beta_{1T}^{-2[\tau_0 T]}}{k_T^2 l(\eta_T)}\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2 \Rightarrow \frac{1}{2c}Y^2$,
*where $X$ and $Y$ are independent $N(0, \frac{1}{2c})$ random variables.*

**Lemma D.2** *Let $\{y_t, t \geq 1\}$ be generated according to Model (1.1), where $\beta_1 = \beta_{1T} = 1 + c/k_T$ with $c > 0$ and $\beta_2 = 1$. Then under assumptions C1-C4, the following results hold jointly:*

*(a)* $\frac{\beta_{1T}^{-[\tau_0 T]}}{\sqrt{T k_T} l(\eta_T)} \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t \Rightarrow Y(\overline{W}(1) - \overline{W}(\tau_0))$,

*(b)* $\frac{\beta_{1T}^{-2[\tau_0 T]}}{T k_T l(\eta_T)} \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2 \Rightarrow (1-\tau_0)Y^2$,

*where $Y$ is as defined in Lemma D.1 and is independent of $\overline{W}(1) - \overline{W}(\tau_0)$.*

**Lemma D.3** *Let $\{y_t, t \geq 1\}$ be generated according to Model (1.1), where $\beta_1 = \beta_{1T} = 1 + c/k_T$ and $\beta_2 = 1$ with $c > 0$. Then under assumptions C1-C4, the results in (B.1) hold.*

**Lemma D.4** *Let $\{y_t, t \geq 1\}$ be generated according to Model (1.1), where $\beta_1 = \beta_{1T} = 1 + c/k_T$ and $\beta_2 = 1$ with $c > 0$. Then under assumptions C1-C4, we have, for any fixed integer $m \geq 0$,*

*(a)* $\frac{k_T}{\beta_{1T}^{2[\tau_0 T]} l(\eta_T)} \Big(RSS_T(\tau_0 - \frac{m}{T}) - RSS_T(\tau_0)\Big) \Rightarrow c^2 m Y^2$,

*(b)* $\frac{k_T}{\beta_{1T}^{2[\tau_0 T]} l(\eta_T)} \Big(RSS_T(\tau_0 + \frac{m}{T}) - RSS_T(\tau_0)\Big) \Rightarrow c^2 m Y^2$,

*where $Y$ is as defined in Lemma D.1.*

**Proof of Theorem 1.4**. To prove (1.11), we first show that $|\hat{\tau}_T - \tau_0| = O_p(1/T)$. It can be proved by following the similar arguments in the proof of Theorem 1.2 and using Lemma D.3. The proof of $P(\hat{k} \neq k_0) \to 0$ can then be completed by using Lemma D.4. The details are omitted for brevity.

To obtain the limiting distribution of $t_1$, we use the fact that the limiting distributions of $\hat{\beta}_1(\hat{\tau}_T)$ and $\hat{\beta}_1(\tau_0)$ are identical when $P(\hat{k} \neq k_0) \to 0$ by applying the similar arguments in the proof of Theorem 4 in Chong (2001). The details are omitted. Then, invoking Lemma D.1, we have

$$\sqrt{\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{l(\eta_T)}}(\hat{\beta}_1(\tau_0) - \beta_{1T}) = \frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sqrt{l(\eta_T)\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}} \Rightarrow N(0,1),$$

which implies

$$t_1 = \sqrt{\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{l(\eta_T)}}(\hat{\beta}_1(\hat{\tau}_T) - \beta_{1T}) \Rightarrow N(0,1).$$

Analogously, $\hat{\beta}_2(\hat{\tau}_T)$ and $\hat{\beta}_2(\tau_0)$ have the same asymptotic distribution. Then, applying Lemma D.2, we have

$$\sqrt{\frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{l(\eta_T)}}(\hat{\beta}_2(\tau_0) - \beta_2) = \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sqrt{l(\eta_T)\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}} \Rightarrow N(0,1),$$

which implies

$$t_2 = \sqrt{\frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{l(\eta_T)}}(\hat{\beta}_2(\hat{\tau}_T) - \beta_2) \Rightarrow N(0,1).$$

□