Asymptotic Inferences for an AR(1) Model with a Change Point: Stationary and Nearly Non-stationary Cases

Pang, Tianxiao and Zhang, Danna and Chong, Terence Tai-Leung

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

30 December 2013

Asymptotic inferences for an AR(1) model with a change point: stationary and nearly non-stationary cases*

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

[1]Department of Mathematics, Yuquan Campus, Zhejiang University, Hangzhou 310027, P.R.China

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

[3]Department of Economics and Institute of Global Economics and Finance,

The Chinese University of Hong Kong, Hong Kong

Abstract. This paper examines the asymptotic inference for AR(1) models with a possible structural break in the AR parameter β near the unity at an unknown time k_0. Consider the model $y_t = \beta_1 y_{t-1} I\{t \le k_0\} + \beta_2 y_{t-1} I\{t > k_0\} + \varepsilon_t,\ t = 1, 2, \cdots, T$, where $I\{\cdot\}$ denotes the indicator function. We examine two cases: Case (I) $|\beta_1| < 1, \beta_2 = \beta_{2T} = 1 - c/T$; and case (II) $\beta_1 = \beta_{1T} = 1 - c/T, |\beta_2| < 1$, where c is a fixed constant, and $\{\varepsilon_t, t \ge 1\}$ is a sequence of i.i.d. random variables which are in the domain of attraction of the normal law with zero means and possibly infinite variances. We derive the limiting distributions of the least squares estimators of β_1 and β_2, and that of the break-point estimator for shrinking break for the aforementioned cases. Monte Carlo simulations are conducted to demonstrate the finite sample properties of the estimators. Our theoretical results are supported by Monte Carlo simulations.

Keywords: AR(1) model, Change point, Domain of attraction of the normal law, Limiting distribution, Least squares estimator.

AMS 2010 subject classification: 60F05, 62F12.

*Tianxiao Pang and Danna Zhang's research was supported by the National Natural Science Foundation of China (No. 11001236, No. 11071213 and No. J1210038) and the Fundamental Research Funds for the Central Universities.

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

†Corresponding author.

1 Introduction

The change-point problem has received considerable attention in the literature over the past three decades (Mankiw and Miron, 1986; Mankiw, Miron and Weil, 1987; Hansen, 1992; Chong, 2001). This paper extends the work of Chong (2001), who studies an AR(1) model with a structural break in the AR parameter β at an unknown time k_0. We consider the following model:

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

where $I\{\cdot\}$ denotes the indicator function and $\{\varepsilon_t, t \geq 1\}$ is a sequence of i.i.d. random variables. Under some regularity conditions that $E\varepsilon_t^4 < \infty$ and $Ey_0^2 < \infty$, Chong (2001) proves the consistency and derives the limiting distributions of the least squares estimators of β_1, β_2 and τ_0 for three cases: (1) $|\beta_1| < 1$ and $|\beta_2| < 1$; (2) $|\beta_1| < 1$ and $\beta_2 = 1$; (3) $\beta_1 = 1$ and $|\beta_2| < 1$.

In the present paper, we focus on Model (1.1) where one of the pre-shift and post-shift AR parameters is less than one in absolute value while the other is local to unity. This case is omitted in Chong (2001). Specifically, we focus on the following two cases: (I) $|\beta_1| < 1$, $\beta_2 = \beta_{2T} = 1 - c/T$; (II) $\beta_1 = \beta_{1T} = 1 - c/T, |\beta_2| < 1$, where c is a fixed constant. The case of local to unity in AR(1) model was first independently studied by Chan and Wei (1987) and Phillips (1987). Their studies bridge the gap between stationary AR(1) model and unit root model. Moreover, since heavy-tailed distributions, such as Student's t distribution with degrees of freedom 2 and Pareto distribution with index 2, are commonly found in insurance, econometrics and other literature, it is more appropriate to impose weaker moment conditions on the ε_t's and y_0 than those in Chong (2001). The primary contribution of this paper is to derive the consistency and limiting distributions of the least squares estimators of β_1, β_2 and the estimator of τ_0 under a more general setting.

Throughout the rest of the present paper, we shall focus on the random variables which are in the domain of attraction of the normal law (DAN), which is an important subclass of heavy-tailed random variables. 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 (Feller, 1971), where B_n takes the form $\sqrt{n}h(n)$ and $h(n)$ is a slowly varying function at infinity. We make the following assumptions:

- C1: $\{\varepsilon_t, t \geq 1\}$ is a sequence of i.i.d. random variables which are in the domain of attraction of the normal law with zero means and possibly infinite variances.

- C2: y_0 is an arbitrary random variable such that $y_0 = o_p(\sqrt{T})$, where T is the sample size.

- C3: $\tau_0 \in [\underline{\tau}, \overline{\tau}] \subset (0, 1)$.

Remark 1. Assumption C2 is a weak initial condition. It not only allows y_0 to be a finite random variable, but also allows it to be a random variable of order smaller than $\sqrt{T}$ in probability.

For any given τ, the ordinary least squares estimators of parameters β_1 and β_2 are given by

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

The symbol $[a]$ denotes the integer part of a and the change-point estimator satisfies

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

where

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

We introduce some notations before presenting our main results. Let $W_1(\cdot)$ and $W_2(\cdot)$ be two independent Brownian motions defined on the non-negative half real R_+; $W(\cdot)$ and $\overline{W}(\cdot)$ be two independent Brownian motions defined on $[0, 1]$ and R_+ respectively; "$\Rightarrow$" signifies the weak convergence of the associated probability measures; "$\xrightarrow{p}$" represents convergence in probability; "$\overset{d}{=}$" denotes identical in distribution. Let C be a finite constant. The limits in this paper are all taken as $T \to \infty$ unless specified otherwise.

Under assumptions C1-C3, we have

Theorem 1.1 *In Model (1.1), if $|\beta_1| < 1$, $\beta_2 = \beta_{2T} = 1 - c/T$, where c is a fixed constant, and the assumptions C1-C3 are satisfied, then the estimators $\hat{\tau}_T, \hat{\beta}_1(\hat{\tau}_T)$ and $\hat{\beta}_2(\hat{\tau}_T)$ are all consistent, and*

$$\begin{cases} |\hat{\tau}_T - \tau_0| = O_p(1/T), \\ \sqrt{T}(\hat{\beta}_1(\hat{\tau}_T) - \beta_1) \Rightarrow N(0, (1 - \beta_1^2)/\tau_0), \\ T(\hat{\beta}_2(\hat{\tau}_T) - \beta_2) \Rightarrow \dfrac{\frac{1}{2} F^2(W, c, \tau_0, 1) + c \int_{\tau_0}^1 e^{2c(1-t)} F^2(W, c, \tau_0, t) dt - \frac{1}{2}(1 - \tau_0)}{\int_{\tau_0}^1 e^{2c(1-t)} F^2(W, c, \tau_0, t) dt}, \end{cases} \tag{1.2}$$

where

$$F(W,c,\tau_0,t) = e^{-c(1-t)}(W(t)-W(\tau_0)) - c\int_{\tau_0}^{t} e^{-c(1-s)}(W(s)-W(\tau_0))ds.$$

If we also let β_{1T} be a sequence of β_1 such that $|\beta_{2T}-\beta_{1T}| \to 0$ and $T(\beta_{2T}-\beta_{1T}) \to \infty$, then the limiting distribution of $\hat{\tau}_T$ is given by

$$(\beta_{2T}-\beta_{1T})T(\hat{\tau}_T-\tau_0) \Rightarrow \arg\max_{\nu\in R}\left\{\frac{C^*(\nu)}{B_a(\frac{1}{2})} - \frac{|\nu|}{2}\right\},$$

where $B_a(\frac{1}{2})$ is generated by $\int_0^\infty \exp(-s)dW_1(s)$ and $C^*(\nu)$ is defined to be $C^*(\nu) = W_1(-\nu)$ for $\nu \leq 0$ and

$$\begin{aligned} C^*(\nu) &= -I(W_2,c,\tau_0,\nu) - \int_0^\nu \frac{I(W_2,c,\tau_0,t)}{B_a(\frac{1}{2})}dI(W_2,c,\tau_0,t) \\ &\quad - \int_0^\nu \Big(\frac{I(W_2,c,\tau_0,t)}{2B_a(\frac{1}{2})}+1\Big)I(W_2,c,\tau_0,t)dt \end{aligned}$$

for $\nu > 0$ with

$$I(W_2,c,\tau_0,t) = W_2(\tau_0+t) - W_2(\tau_0) - c\int_{\tau_0}^{\tau_0+t} e^{-c(\tau_0+t-s)}(W_2(s)-W_2(\tau_0))ds.$$

Theorem 1.2 *In Model (1.1), if $\beta_1 = \beta_{1T} = 1-c/T$, where c is a fixed constant and $|\beta_2| < 1$, and the assumptions C1-C3 are satisfied, then the estimators $\hat{\tau}_T, \hat{\beta}_1(\hat{\tau}_T)$ and $\hat{\beta}_2(\hat{\tau}_T)$ are all consistent and*

$$\begin{cases} \mathbb{P}(\hat{k} \neq k_0) \to 0, \\ T(\hat{\beta}_1(\hat{\tau}_T)-\beta_1) \Rightarrow \dfrac{\frac{1}{2}e^{2c(1-\tau_0)}G^2(W,c,\tau_0) + c\int_0^{\tau_0} e^{2c(1-t)}G^2(W,c,t)dt - \frac{\tau_0}{2}}{\int_0^{\tau_0} e^{2c(1-t)}G^2(W,c,t)dt}, \\ \sqrt{T}(\hat{\beta}_2(\hat{\tau}_T)-\beta_2) \Rightarrow \dfrac{\sqrt{1-\beta_2^2}\cdot W(B(c,\tau_0))}{1-\tau_0+e^{2c(1-\tau_0)}G^2(W,c,\tau_0)}, \end{cases} \tag{1.3}$$

where

$$G(W,c,t) = e^{-c(1-t)}W(t) - c\int_0^t e^{-c(1-s)}W(s)ds$$

and

$$B(c,\tau_0) = (1-e^{-2c\tau_0})/(2c) + 1 - \tau_0.$$

Suppose we also let β_{2T} be a sequence of β_2 such that $\sqrt{T}(\beta_{2T}-\beta_{1T}) \to 0$ and $T^{3/4}(\beta_{1T} - \beta_{2T}) \to \infty$, then the limiting distribution of $\hat{\tau}_T$ is given by

$$(\beta_{2T}-\beta_{1T})^2T^2(\hat{\tau}_T-\tau_0) \Rightarrow \arg\max_{\nu\in R}\left\{\frac{B^*(\nu)}{e^{c(1-\tau_0)}G(W_1,c,\tau_0)} - \frac{|\nu|}{2}\right\},$$

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

Remark 2. In Theorem 1.1, letting $c = 0$, it is clear that

$$\begin{aligned}
& \frac{\frac{1}{2}F^2(W,c,\tau_0,1) + c\int_{\tau_0}^1 e^{2c(1-t)}F^2(W,c,\tau_0,t)dt - \frac{1}{2}(1-\tau_0)}{\int_{\tau_0}^1 e^{2c(1-t)}F^2(W,c,\tau_0,t)dt}\Big|_{c=0} \\
= & \frac{\frac{1}{2}(W(1)-W(\tau_0))^2 - \frac{1}{2}(1-\tau_0)}{\int_{\tau_0}^1 (W(t)-W(\tau_0))^2 dt} \\
\stackrel{d}{=} & \frac{W^2(1)-1}{2(1-\tau_0)\int_0^1 W^2(t)dt}
\end{aligned}$$

and for $\nu > 0$

$$\begin{aligned}
& \Big\{ -I(W_2,c,\tau_0,\nu) - \int_0^\nu \frac{I(W_2,c,\tau_0,t)}{B_a(\frac{1}{2})} dI(W_2,c,\tau_0,t) \\
& \quad - \int_0^\nu \Big(\frac{I(W_2,c,\tau_0,t)}{2B_a(\frac{1}{2})} + 1\Big) I(W_2,c,\tau_0,t)dt \Big\}\Big|_{c=0} \\
= & -(W_2(\tau_0+\nu) - W_2(\tau_0)) - \int_0^\nu \frac{W_2(\tau_0+t)-W_2(\tau_0)}{B_a(\frac{1}{2})} d(W_2(\tau_0+t)-W_2(\tau_0)) \\
& \quad - \int_0^\nu \Big(\frac{W_2(\tau_0+t)-W_2(\tau_0)}{2B_a(\frac{1}{2})} + 1\Big)(W_2(\tau_0+t)-W_2(\tau_0))dt \\
\stackrel{d}{=} & -W_2(\nu) - \int_0^\nu \frac{W_2(t)}{B_a(\frac{1}{2})} dW_2(t) - \int_0^\nu \Big(\frac{W_2(t)}{2B_a(\frac{1}{2})} + 1\Big) W_2(t)dt.
\end{aligned}$$

The above two expressions coincide with the third term of (15) and $C^*(\nu)$ with $\nu > 0$ in Chong (2001), respectively. Hence, our Theorem 1.1 is reduced to Theorem 3 in Chong (2001) by taking $c = 0$.

Similarly, letting $c = 0$ in Theorem 1.2, we have

$$\frac{\frac{1}{2}e^{2c(1-\tau_0)}G^2(W,c,\tau_0) + c\int_0^{\tau_0} e^{2c(1-t)}G^2(W,c,t)dt - \frac{\tau_0}{2}}{\int_0^{\tau_0} e^{2c(1-t)}G^2(W,c,t)dt}\Big|_{c=0} = \frac{W^2(\tau_0)-\tau_0}{2\int_0^{\tau_0} W^2(t)dt},$$

$$\frac{\sqrt{1-\beta_2^2}\cdot \overline{W}(B(c,\tau_0))}{1-\tau_0+e^{2c(1-\tau_0)}G^2(W,c,\tau_0)}\Big|_{c=0} = \frac{\sqrt{1-\beta_2^2}\,\overline{W}(1)}{1-\tau_0+W^2(\tau_0)}$$

and

$$e^{c(1-\tau_0)}G(W_1,c,\tau_0)\big|_{c=0} = W_1(\tau_0),$$

indicating that Theorem 1.2 is reduced to Theorem 4 in Chong (2001) when $c = 0$.

Note that the assumptions on the ε_t's and y_0 are weaker than those in Chong (2001).

Remark 3. The limiting distributions of $\hat{\beta}_1(\hat{\tau}_T)$ and $\hat{\beta}_2(\hat{\tau}_T)$ in Theorem 1.2 could be simplified if assumption C2 is more specific. For example, if the initial value y_0 is defined as

$y_0 = y_{T,0} = \sum_{j=0}^{\infty} \rho_T^j \varepsilon_{-j}$ with ρ_T satisfying $T(1-\rho_T) = h_T \to 0$ and $\{\varepsilon_{-j}, j \geq 0\}$ being a sequence of i.i.d. random variables sharing the same distribution with ε_1, then similar arguments of Lemma 3 in Andrews and Guggenberger (2008) will lead to $\sqrt{2h_T} y_0 / \sqrt{Tl(\eta_T)} \Rightarrow N(0,1)$, where the definitions of η_T and the function $l(\cdot)$ can be found at the beginning of Section 3. Since y_0 dominates the asymptotic distribution of $\hat{\beta}_1(\hat{\tau}_T)$, we have

$$\begin{aligned} \frac{2h_T}{Tl(\eta_T)} \sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t &\xrightarrow{p} 0, \\ \frac{2h_T}{T^2 l(\eta_T)} \sum_{t=1}^{[\tau_0 T]} y_{t-1}^2 &\Rightarrow \frac{1-e^{-2c\tau_0}}{2c} W^2(1). \end{aligned}$$

Consequently, we have

$$T(\hat{\beta}_1(\hat{\tau}_T) - \beta_1) \xrightarrow{p} 0.$$

Similarly, from the proof of Lemma 4.3 in Pang, Zhang and Chong (2013), it can be shown that

$$\sqrt{\frac{T}{2h_T}} (\hat{\beta}_1(\hat{\tau}_T) - \beta_1) \Rightarrow (1-\beta_2^2) e^{c\tau_0} \pi(\beta_2)/W(1)$$

if the stationary distribution (denoted by $\pi(\beta_2)$) of the AR(1) process $y_t = \beta_2 y_{t-1} + \varepsilon_t/\sqrt{l(\eta_T)}$ with $y_0 = 0$ for $t = 1, \cdots, T - [\tau_0 T]$, exists. Note that $\pi(\beta_2)$ and $W(1)$ are independent.

Remark 4. Chong (2001) proves that $|\hat{\tau}_T - \tau_0| = O_p(1/T)$ in the case of $|\beta_1| < 1$ and $\beta_2 = 1$, while $\mathbb{P}(\hat{k} \neq k_0) \to 0$ in the case of $\beta_1 = 1$ and $|\beta_2| < 1$. This result also holds in the present paper. Note that the result about the estimator of k_0 in Theorem 1.2 is stronger than that in Theorem 1.1. This is because the signal from the regressor y_{t-1} when the serial correlation coefficient is $1 - c/T$ is stronger than that from the regressor y_{t-1} when the serial correlation coefficient is a fixed constant smaller than one in absolute value (as implied by the faster convergence rate of $\hat{\beta}_2(\hat{\tau}_T)$ in Theorem 1.1 and the faster convergence rate of $\hat{\beta}_1(\hat{\tau}_T)$ in Theorem 1.2), meanwhile, the signal from the regressor y_{t-1} under the situation of $(\beta_1, \beta_2) = (1 - c/T, c_0)$ is stronger than that under the situation of $(\beta_1, \beta_2) = (c_0, 1 - c/T)$, where c_0 is fixed and $|c_0| < 1$.

Remark 5. The statistical inference on the least squares estimators of β_1, β_2 and $\hat{\tau}_T$ for the following cases: (I) $\beta_1 = \beta_{1T} = 1 - c/T$, $\beta_2 = 1$; (II) $\beta_1 = 1, \beta_2 = \beta_{2T} = 1 - c/T$ are much more complicated and would be left for future research.

The rest of the paper is organized as follows: Section 2 presents the simulation results for the finite sample properties of the estimators in Theorems 1.1 and 1.2. Section 3 states some useful lemmas and provides the proof for Theorem 1.1. Section 4 provides the proof for Theorem 1.2. Note that all proofs of the lemmas in this paper are omitted for reasons of space, and the readers are referred to Pang, Zhang and Chong (2013) for details.

2 Simulations

We perform the following experiments to see how well our asymptotic results match the finite-sample properties of the estimators. In all experiments, the sample size is set at $T = 200$ and the number of replications is set at $N = 20,000$; $\{y_t\}_{t=1}^{T}$ is generated from Model (1.1); y_0 has the following probability density function

$$f(x) = \begin{cases} 0 & \text{if } x \leq -2, \\ \frac{3}{2(x+3)^{5/2}} & \text{if } x > -2. \end{cases}$$

Note that $E|y_0| < \infty$ and $E|y_0|^{3/2+\delta} = \infty$ for any $\delta \geq 0$. Note also that assumption C2 holds. The true change point is set at $\tau_0 = 0.3$ and 0.5. For the constant c and the distribution of ε_t's, we consider the following numerical setup:

$$\begin{cases} c = 1, \\ \varepsilon_t \in \{t(3), \;\; t(2)\}, \end{cases} \tag{2.4}$$

where $t(3)$ and $t(2)$ denote the student-t random variables with degrees of freedom 3 and 2 respectively. It is easy to verify that $t(3)$ and $t(2)$ are both in the domain of attraction of the normal law, and that $t(3)$ has finite variance but infinite fourth moment, while $t(2)$ has infinite variance. When the AR parameter which depends on the sample size T and the constant c is determined, the values of the fixed AR parameter are set to 0.5, 0.75 and 0.8. Note that we only present the simulations for the case where $\tau_0 = 0.5$ and the fixed AR parameter is equal to 0.5 in this paper in order to conserve space. Readers are referred to Pang, Zhang and Chong (2013) for more detailed simulations.

First, we conduct experiments to verify the second and the third results in (1.2) under (2.4), which predict that the finite-sample distribution of $\hat{\beta}_1(\hat{\tau}_T)$ is approximately normal, whereas $\hat{\beta}_2(\hat{\tau}_T)$ appears to have a Dickey-Fuller distribution. Figure 1-Figure 2 agree with our results.

Second, we conduct experiments to verify the second and the third results in (1.3) under (2.4), which predict that the finite-sample distribution of $\hat{\beta}_1(\hat{\tau}_T)$ is approximately the

Dickey-Fuller type, whereas $\hat{\beta}_2(\hat{\tau}_T)$ will have a normal distribution. Figure 3-Figure 4 also agree with our results.

In all of the experiments in Pang, Zhang and Chong (2013), we have the following observations: (I) the performance for the simulations on $\hat{\beta}_1(\hat{\tau}_T)$ is better for the case where $\tau_0 = 0.5$ compared to the case where $\tau_0 = 0.3$, since more data are used to generate the first subsample. Analogously, the performance for the simulations on $\hat{\beta}_2(\hat{\tau}_T)$ is better for the case where $\tau_0 = 0.3$ compared to the case where $\tau_0 = 0.5$. (II) the performance is better for the case where $\{\varepsilon_t\}_{t=1}^T \sim t(3)$ compared to the case where $\{\varepsilon_t\}_{t=1}^T \sim t(2)$ under (2.4). This is not surprising since $t(2)$ is a heavy-tailed distribution. We have also conducted the experiments for larger sample size when $\{\varepsilon_t\}_{t=1}^T \sim t(2)$, the performance does not improve much. Given the results, one should be cautious when conducting statistical inference under heavy-tailed innovations such as $t(2)$. The experiments when $c = -1$ are also studied. However, since the results are very similar to the case where $c = 1$, we do not report those simulations here to conserve space.

In the following figures, we let $c = 1$. The solid line shows the finite sample distribution when $T = 200$ while the dashed line shows the asymptotic distribution.

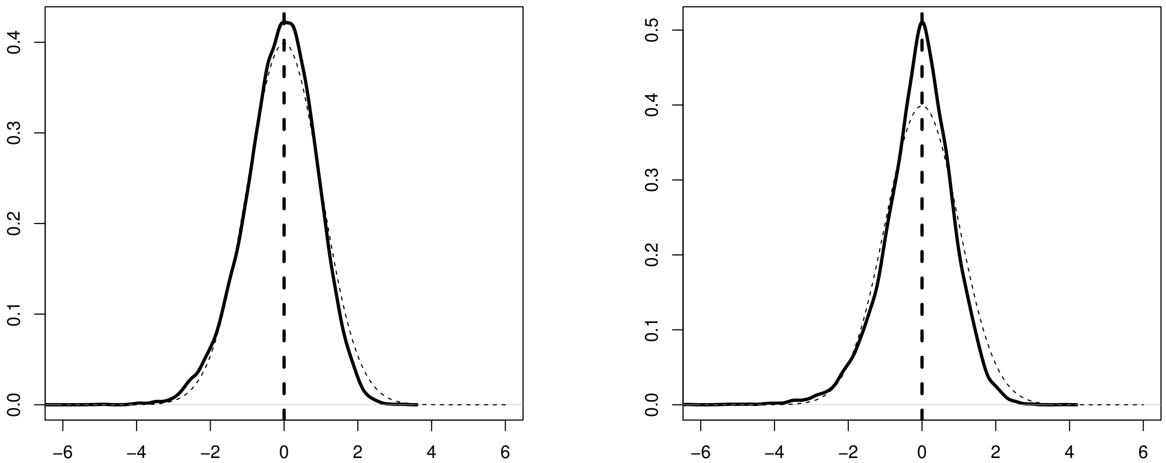

Figure 1: Distribution of $\sqrt{T}(\hat{\beta}_1(\hat{\tau}_T) - \beta_1)$ when $\tau_0 = 0.5$, $\beta_1 = 0.5$, $\beta_2 = 1 - 1/T$. Left: $\{\varepsilon_t\}_{t=1}^T \sim t(3)$; Right: $\{\varepsilon_t\}_{t=1}^T \sim t(2)$.

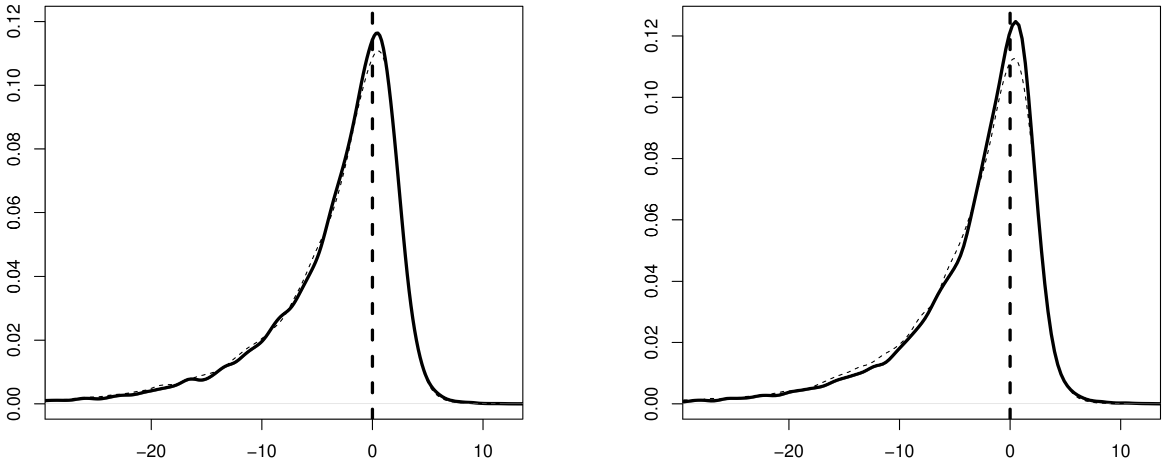

Figure 2: Distribution of $T(\hat{\beta}_2(\hat{\tau}_T) - \beta_2)$ when $\tau_0 = 0.5$, $\beta_1 = 0.5$, $\beta_2 = 1 - 1/T$. Left: $\{\varepsilon_t\}_{t=1}^T \sim t(3)$; Right: $\{\varepsilon_t\}_{t=1}^T \sim t(2)$.

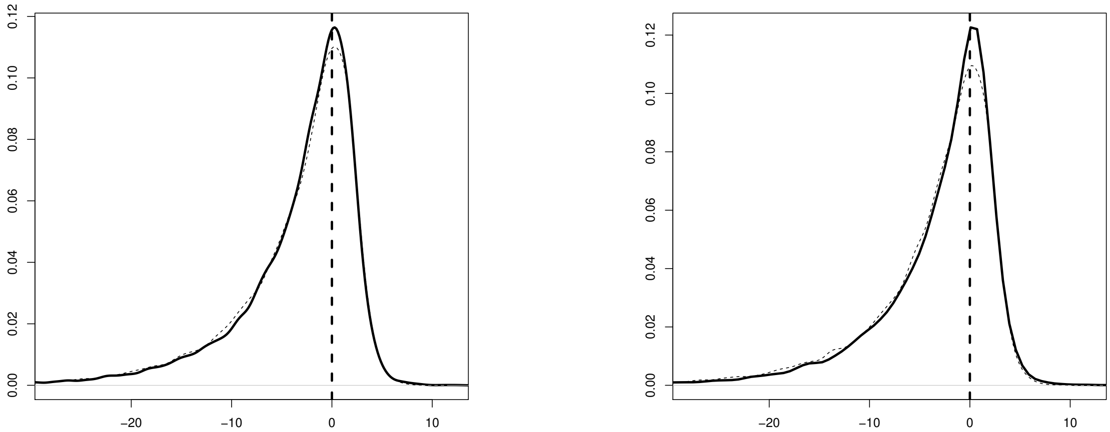

Figure 3: Distribution of $\sqrt{T}(\hat{\beta}_1(\hat{\tau}_T) - \beta_1)$ when $\tau_0 = 0.5$, $\beta_2 = 0.5$, $\beta_1 = 1 - 1/T$. Left: $\{\varepsilon_t\}_{t=1}^T \sim t(3)$; Right: $\{\varepsilon_t\}_{t=1}^T \sim t(2)$.

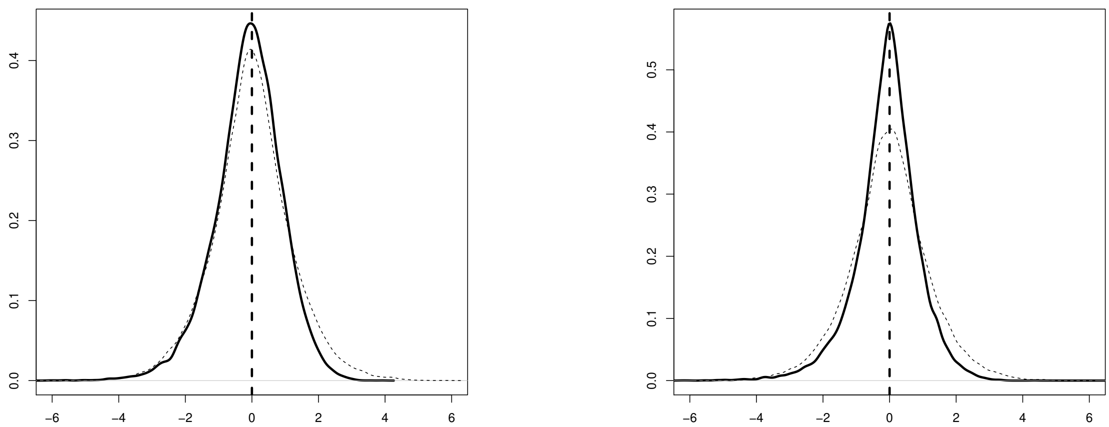

Figure 4: Distribution of $T(\hat{\beta}_2(\hat{\tau}_T) - \beta_2)$ when $\tau_0 = 0.5$, $\beta_2 = 0.5$, $\beta_1 = 1 - 1/T$ $\{\varepsilon_t\}_{t=1}^T \sim t(3)$; Right: $\{\varepsilon_t\}_{t=1}^T \sim t(2)$.

3 Proof of Theorem 1.1

To prove Theorem 1.1 and Theorem 1.2 when the ε_t's are heavy-tailed, we employ the truncation technique in this paper. We let

$$l(t) = E\varepsilon_1^2 I\{|\varepsilon_1| \le t\}, \quad b = \inf\{t \ge 1 : l(t) > 0\},$$

and

$$\eta_j = \inf\{s : s \ge b+1, \frac{l(s)}{s^2} \le \frac{1}{j}\}, \quad \text{for } j = 1, 2, 3, \cdots.$$

Note that $Tl(\eta_T) \le \eta_T^2$ for all $T \ge 1$ and $\eta_T^2 \approx Tl(\eta_T)$ for large T. In addition, for each T we let

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

for $t = 1, \cdots, T$.

The following two lemmas are taken from Csörgő et al. (2003) and Pang and Zhang (2011) respectively:

Lemma 3.1 *Let X be a random variable, and denote $l(x) = EX^2 I\{|X| \le x\}$. The following statements are equivalent:*

(1a) X *is in the domain of attraction of the normal law,*

(1b) $x^2\mathbb{P}(|X| > x) = o(l(x))$,

(1c) $xE(|X|I\{|X| > x\}) = o(l(x))$,

(1d) $E(|X|^n I\{|X| \le x\}) = o(x^{n-2}l(x))$ *for* $n > 2$.

Lemma 3.2 *Suppose assumptions C1-C3 are satisfied, then in Model (1.1) with $|\beta_1| < 1$, the following results hold jointly:*

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

(2b) $\frac{1}{Tl(\eta_T)} \sum_{t=1}^{[\tau_0 T]} y_{t-1}^2 \xrightarrow{p} \tau_0/(1-\beta_1^2)$.

The following two lemmas are useful in proving Theorem 1.1:

Lemma 3.3 *Suppose assumption C1 is satisfied and $\beta_2 = \beta_{2T} = 1 - c/T$, where c is a fixed constant, then for any $\tau_0 \le \tau \le 1$,*

$$\begin{aligned} X_T(\tau_0, \tau) &:= \frac{1}{\sqrt{Tl(\eta_T)}} \sum_{t=[\tau_0 T]+1}^{[\tau T]} \beta_{2T}^{T-t}\varepsilon_t \\ &\Rightarrow e^{-c(1-\tau)}(W(\tau) - W(\tau_0)) - c\int_{\tau_0}^{\tau} e^{-c(1-t)}(W(t) - W(\tau_0))dt = F(W, c, \tau_0, \tau) \end{aligned}$$

and for any $0 \le s \le 1-\tau_0$

$$
\begin{aligned}
Z_T(\tau_0, s) \quad &:= \quad \frac{1}{\sqrt{Tl(\eta_T)}} \sum_{t=[\tau_0 T]+1}^{[\tau_0 T]+[sT]} \beta_{2T}^{[\tau_0 T]+[sT]-t} \varepsilon_t \\
&\Rightarrow \quad W(\tau_0+s) - W(\tau_0) - c\int_{\tau_0}^{\tau_0+s} e^{-c(\tau_0+s-t)}(W(t)-W(\tau_0))dt = I(W, c, \tau_0, s).
\end{aligned}
$$

Lemma 3.4 *Let* $\{y_t, t \ge 1\}$ *be generated according to Model (1.1) with* $|\beta_1| < 1$ *and* $\beta_2 = \beta_{2T} = 1 - c/T$*, where* c *is a fixed constant. Under the assumptions C1-C3, the following results hold jointly:*

(3a) $\frac{1}{Tl(\eta_T)} \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t \Rightarrow \frac{1}{2}F^2(W, c, \tau_0, 1) + c\int_{\tau_0}^1 e^{2c(1-t)} F^2(W, c, \tau_0, t)dt - \frac{1}{2}(1-\tau_0)$,

(3b) $\frac{1}{T^2 l(\eta_T)} \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2 \Rightarrow \int_{\tau_0}^1 e^{2c(1-t)} F^2(W, c, \tau_0, t)dt.$

Proof of Theorem 1.1. Consider the first part of (1.2). Along the lines of the proof of Theorem 3 in Chong (2001), it is sufficient to prove that

$$
\begin{cases}
A_1 = \dfrac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} = o_p(1), \\
A_2 = \sup\limits_{m \in D_{1T}} \dfrac{\sum_{t=m+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=m+1}^{[\tau_0 T]} y_{t-1}^2} = o_p(1), \\
A_3 = \sup\limits_{m \in D_{1T}} \Big| \dfrac{\sum_{t=m+1}^{T} y_{t-1}^2}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2 \sum_{t=m+1}^{[\tau_0 T]} y_{t-1}^2} \Lambda_T(\dfrac{m}{T}) \Big| = o_p(1), \\
A_4 = \dfrac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} = o_p(1), \\
A_5 = \sup\limits_{m \in D_{2T}} \dfrac{\sum_{t=[\tau_0 T]+1}^{m} y_{t-1}\varepsilon_t}{\sum_{t=[\tau_0 T]+1}^{m} y_{t-1}^2} = o_p(1), \\
A_6 = \sup\limits_{m \in D_{2T}} \Big| \dfrac{\sum_{t=1}^{m} y_{t-1}^2}{\sum_{t=[\tau_0 T]+1}^{m} y_{t-1}^2 \sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} \Lambda_T(\dfrac{m}{T}) \Big| = o_p(1)
\end{cases}
\tag{3.1}
$$

in the case of $|\beta_1| < 1$ and $\beta_2 = \beta_{2T} = 1 - c/T$, where c is a fixed constant,

$$
\begin{aligned}
\Lambda_T(\frac{m}{T}) \quad = \quad & \frac{(\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t)^2}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} \Bigg(1 - \Big(\frac{\sum_{t=1}^{m} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t} \Big)^2 \frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=1}^{m} y_{t-1}^2} \Bigg) \\
& + \frac{(\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t)^2}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} \Bigg(1 - \Big(\frac{\sum_{t=m+1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t} \Big)^2 \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{\sum_{t=m+1}^{T} y_{t-1}^2} \Bigg)
\end{aligned}
\tag{3.2}
$$

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 $M_T \to \infty$ and $M_T/T \to 0$, where Z_T denotes the set $\{0, 1, 2, \cdots, T\}$.

We apply Lemmas 3.2 and 3.4 to prove (3.1). The results for A_1 and A_4 are obvious. For A_2, applying the Uniform Law of Large Numbers in Andrews (1987, Theorem 1) yields

$$|A_2| = \Big|\sup_{m\in D_{1T}} \frac{\sum_{t=m+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=m+1}^{[\tau_0 T]} y_{t-1}^2}\Big| \le O_p\Big(\frac{1}{\sqrt{M_T}}\Big) = o_p(1).$$

For A_3, A_5 and A_6, note that

$$\begin{aligned}
A_3 &= \sup_{m\in D_{1T}} \Big|\frac{\sum_{t=m+1}^{T} y_{t-1}^2}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2 \sum_{t=m+1}^{[\tau_0 T]} y_{t-1}^2}\Lambda_T(\frac{m}{T})\Big| \\
&\le \Big(\frac{1}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} + \frac{1}{\sum_{t=[\tau_0 T]-M_T}^{[\tau_0 T]} y_{t-1}^2}\Big)\Big(\sup_{m\in D_{1T}}\Big|\frac{(\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t)^2}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} - \frac{(\sum_{t=1}^{m} y_{t-1}\varepsilon_t)^2}{\sum_{t=1}^{m} y_{t-1}^2}\Big| \\
&\quad + \sup_{m\in D_{1T}}\Big|\frac{(\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t)^2}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} - \frac{(\sum_{t=m+1}^{T} y_{t-1}\varepsilon_t)^2}{\sum_{t=m+1}^{T} y_{t-1}^2}\Big|\Big) \\
&= \Big(O_p(\frac{1}{T^2 l(\eta_T)}) + O_p(\frac{1}{M_T l(\eta_T)})\Big)\cdot O_p(l(\eta_T)) \\
&= O_p(\frac{1}{M_T}) = o_p(1),
\end{aligned}$$

$$\begin{aligned}
A_5 &= \sup_{m\in D_{2T}} \frac{\sum_{t=[\tau_0 T]+1}^{m} y_{t-1}\varepsilon_t}{\sum_{t=[\tau_0 T]+1}^{m} y_{t-1}^2} \\
&= \sup_{m\in D_{2T}} O_p\Big(\frac{1}{m-[\tau_0 T]}\Big) \\
&= O_p(\frac{1}{M_T}) = o_p(1)
\end{aligned}$$

and

$$\begin{aligned}
A_6 &= \sup_{m\in D_{2T}} \Big|\frac{\sum_{t=1}^{m} y_{t-1}^2}{\sum_{t=[\tau_0 T]+1}^{m} y_{t-1}^2 \sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}\Lambda_T(\frac{m}{T})\Big| \\
&\le \Big(\frac{1}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} + \frac{1}{\sum_{t=[\tau_0 T]+1}^{[\tau_0 T]+M_T} y_{t-1}^2}\Big)\Big(\sup_{m\in D_{2T}}\Big|\frac{(\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t)^2}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} - \frac{(\sum_{t=1}^{m} y_{t-1}\varepsilon_t)^2}{\sum_{t=1}^{m} y_{t-1}^2}\Big| \\
&\quad + \sup_{m\in D_{2T}}\Big|\frac{(\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t)^2}{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} - \frac{(\sum_{t=m+1}^{T} y_{t-1}\varepsilon_t)^2}{\sum_{t=m+1}^{T} y_{t-1}^2}\Big|\Big) \\
&= \Big(O_p(\frac{1}{T l(\eta_T)}) + O_p(\frac{1}{M_T^2 l(\eta_T)})\Big)\cdot O_p(l(\eta_T)) \\
&= o_p(\frac{1}{M_T}) = o_p(1).
\end{aligned}$$

Hence, the first part of (1.2) is proved.

To find the limiting distribution of $\hat\beta_1(\hat\tau_T)$, first note that $\hat\tau_T - \tau_0 = O_p(1/T)$. Following

Appendix G in Chong (2001), we have

$$
\begin{aligned}
& \sqrt{T}(\hat{\beta}_1(\hat{\tau}_T) - \hat{\beta}_1(\tau_0)) \\
= \; & \sqrt{T}\Big(\frac{\sum_{t=1}^{[\hat{\tau}_T T]} y_t y_{t-1}}{\sum_{t=1}^{[\hat{\tau}_T T]} y_t^2} - \frac{\sum_{t=1}^{[\tau_0 T]} y_t y_{t-1}}{\sum_{t=1}^{[\tau_0 T]} y_t^2} \Big) \\
= \; & I\{\hat{\tau}_T \le \tau_0\}\sqrt{T}\Big(\frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2} \frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} - \frac{\sum_{t=[\hat{\tau}_T T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2} \Big) \\
& + I\{\hat{\tau}_T > \tau_0\}\sqrt{T}\Big(- \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2} \frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} + \frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2} \\
& \qquad + (\beta_{2T} - \beta_1)\frac{\sum_{t=[\tau_0 T]+1}^{[\hat{\tau}_T T]} y_{t-1}^2}{\sum_{t=1}^{[\hat{\tau}_T T]} y_{t-1}^2} \Big) \\
= \; & I\{\hat{\tau}_T \le \tau_0\}\sqrt{T}\Big(O_p(\frac{l(\eta_T)}{Tl(\eta_T)}) O_p(\frac{1}{\sqrt{T}}) - O_p(\frac{l(\eta_T)}{Tl(\eta_T)}) \Big) \\
& + I\{\hat{\tau}_T > \tau_0\}\sqrt{T}\Big(- O_p(\frac{l(\eta_T)}{Tl(\eta_T)}) O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{l(\eta_T)}{Tl(\eta_T)}) + O_p(\frac{l(\eta_T)}{Tl(\eta_T)}) \Big) \\
= \; & o_p(1),
\end{aligned}
$$

one is referred to Chong (2001) for more details. Thus, $\hat{\beta}_1(\hat{\tau}_T)$ and $\hat{\beta}_1(\tau_0)$ have the same asymptotic distribution. Applying Lemma 3.2, we have

$$
\sqrt{T}(\hat{\beta}_1(\hat{\tau}_T) - \beta_1) \overset{d}{=} \sqrt{T}(\hat{\beta}_1(\tau_0) - \beta_1) = \frac{\frac{1}{\sqrt{T}l(\eta_T)} \sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\frac{1}{Tl(\eta_T)} \sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} \Rightarrow N\big(0, \frac{1-\beta_1^2}{\tau_0}\big).
$$

Similarly, we have

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

Thus, $\hat{\beta}_2(\hat{\tau}_T)$ and $\hat{\beta}_2(\tau_0)$ also have the same asymptotic distribution. Applying Lemma 3.4, we have

$$\begin{aligned} T(\hat{\beta}_2(\hat{\tau}_T)-\beta_2) &\stackrel{d}{=} T(\hat{\beta}_2(\tau_0)-\beta_2) \\ &= \frac{\frac{1}{Tl(\eta_T)}\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\frac{1}{T^2 l(\eta_T)}\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2} \\ &\Rightarrow \frac{\frac{1}{2}F^2(W,c,\tau_0,1)+c\int_{\tau_0}^1 e^{2c(1-t)}F^2(W,c,\tau_0,t)dt-\frac{1}{2}(1-\tau_0)}{\int_{\tau_0}^1 e^{2c(1-t)}F^2(W,c,\tau_0,t)dt}. \end{aligned}$$

To derive the limiting distribution of $\hat{\tau}_T$ for shrinking shift, we let $\beta_2=\beta_{2T}=1-c/T$ and $\beta_1=\beta_{1T}=\beta_{2T}-1/g(T)$ in the remaining proof of Theorem 1.1, where $g(T)>0$ with $g(T)\to\infty$ and $g(T)/T\to 0$. Note that $\beta_{1T}=1-1/g(T)+o(1/g(T))$. Hence, the sequence $\{y_t, 1\le t\le [\tau_0 T]\}$ is generated from a mildly integrated AR(1) model, as a result, the results or ideas from Phillips and Magdalinos (2007) and Huang et al. (2012) could be applied directly. Following Chong (2001), first, for $\tau=\tau_0+\nu g(T)/T$ and $\nu\le 0$, by recalling (3.2), we have

$$\begin{aligned} |\Lambda_T(\tau)| &= O_p(l(\eta_T))\Big(1-(1-o_p(1))^2(1+o_p(1)\Big)+O_p(l(\eta_T))\Big(1-(1+o_p(1))^2(1-o_p(1))\Big) \\ &= o_p(l(\eta_T)). \end{aligned} \tag{3.3}$$

Second, for any $t=0,\cdots,[|\nu|g(T)]-1$, we have

$$\begin{aligned} \frac{y_{[\tau_0 T]-t-1}}{\sqrt{g(T)l(\eta_T)}} &= \frac{1}{\sqrt{g(T)l(\eta_T)}}\sum_{i=1}^{[\tau_0 T]-t-1}\beta_{1T}^{[\tau_0 T]-t-1-i}\varepsilon_i^{(1)} \\ &\quad +\frac{1}{\sqrt{g(T)l(\eta_T)}}\sum_{i=1}^{[\tau_0 T]-t-1}\beta_{1T}^{[\tau_0 T]-t-1-i}\varepsilon_i^{(2)}+\frac{\beta_{1T}^{[\tau_0 T]-t-1}y_0}{\sqrt{g(T)l(\eta_T)}}. \end{aligned} \tag{3.4}$$

It is not difficult to show that

$$\frac{\beta_{1T}^{[\tau_0 T]-t-1}y_0}{\sqrt{g(T)l(\eta_T)}}=\sqrt{\frac{T}{g(T)}}\beta_{1T}^{[\tau_0 T]-t-1}\cdot\frac{y_0}{\sqrt{Tl(\eta_T)}}\xrightarrow{p}0 \tag{3.5}$$

by $y_0=o_p(\sqrt{T})$ and $g(T)=o(T)$; see the proof of Proposition A.1 in Phillips and Magdalinos (2007) for more details. In addition, it follows from Lemma 3.1 that

$$\begin{aligned} &\frac{1}{\sqrt{g(T)l(\eta_T)}}\cdot E\Big|\sum_{i=1}^{[\tau_0 T]-t-1}\beta_{1T}^{[\tau_0 T]-t-1-i}\varepsilon_i^{(2)}\Big| \\ = &\frac{1}{\sqrt{g(T)l(\eta_T)}}\cdot\sum_{i=1}^{[\tau_0 T]-t-1}|\beta_{1T}|^{[\tau_0 T]-t-1-i}\cdot o(\frac{l(\eta_T)}{\eta_T})=o(1) \end{aligned}$$

by recalling that $\eta_T^2 \approx Tl(\eta_T)$ for large T. Then, for any $t = 0, \cdots, [|\nu|g(T)] - 1$

$$\begin{aligned}
\frac{y_{[\tau_0 T]-t-1}}{\sqrt{g(T)l(\eta_T)}} &= \frac{1}{\sqrt{g(T)l(\eta_T)}} \sum_{i=1}^{[\tau_0 T]-t-1} \beta_{1T}^{[\tau_0 T]-t-1-i} \varepsilon_i^{(1)} + o_p(1) \\
&\Rightarrow \int_0^\infty e^{-s} dW_1(s) \stackrel{d}{=} B_a(\frac{1}{2}),
\end{aligned} \tag{3.6}$$

see page 138 in Chong (2001) for details. Note also that

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

by functional central limit theorem for i.i.d. random variables from DAN. As a result, we have

$$\frac{1}{g(T)l(\eta_T)} \sum_{t=0}^{[|\nu|g(T)]-1} y_{[\tau_0 T]-t-1} \varepsilon_{[\tau_0 T]-t} \Rightarrow B_a(\frac{1}{2}) W_1(|\nu|) \tag{3.7}$$

and

$$\frac{1}{g^2(T)l(\eta_T)} \sum_{t=0}^{[|\nu|g(T)]-1} y^2_{[\tau_0 T]-t-1} \Rightarrow |\nu| B_a^2(\frac{1}{2}). \tag{3.8}$$

Moreover, note that

$$\frac{(\beta_{2T} - \beta_{1T})}{l(\eta_T)} \frac{\sum_{t=[\tau T]+1}^{[\tau_0 T]} y_{t-1}^2 \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=[\tau T]+1}^{T} y_{t-1}^2} = \frac{O_p(g^2(T)l(\eta_T)) O_p(Tl(\eta_T))}{g(T)l(\eta_T) O_p(T^2 l(\eta_T))} = o_p(1) \tag{3.9}$$

and

$$\frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}{\sum_{t=[\tau T]+1}^{T} y_{t-1}^2} \xrightarrow{p} 1. \tag{3.10}$$

Then, by recalling equation B.2 on page 117 in Chong (2001) when $\underline{\tau} \le \tau \le \tau_0$

$$\begin{aligned}
& RSS_T(\tau) - RSS_T(\tau_0) \\
= \; & 2(\beta_{2T} - \beta_{1T}) \Bigg(\frac{\sum_{t=[\tau T]+1}^{[\tau_0 T]} y_{t-1}^2 \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=[\tau T]+1}^{T} y_{t-1}^2} - \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2 \sum_{t=[\tau T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=[\tau T]+1}^{T} y_{t-1}^2} \Bigg) \\
& + (\beta_{2T} - \beta_{1T})^2 \frac{\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2 \sum_{t=[\tau T]+1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=[\tau T]+1}^{T} y_{t-1}^2} + \Lambda_T(\tau),
\end{aligned}$$

we have

$$\begin{aligned}
& \frac{RSS_T(\tau) - RSS_T(\tau_0)}{l(\eta_T)} \\
= \; & -\frac{2}{g(T)l(\eta_T)} \sum_{t=[\tau T]+1}^{[\tau_0 T]} y_{t-1}\varepsilon_t \cdot (1 + o_p(1)) + \frac{1}{g^2(T)l(\eta_T)} \sum_{t=[\tau T]+1}^{[\tau_0 T]} y_{t-1}^2 \cdot (1 + o_p(1)) + o_p(1) \\
\Rightarrow \; & -2B_a(\frac{1}{2}) W_1(|\nu|) + |\nu| B_a^2(\frac{1}{2})
\end{aligned}$$

by (3.7)-(3.10).

Similarly, for $\tau = \tau_0 + \nu g(T)/T$ with $\nu > 0$, we also have $\Lambda_T(\tau) = o_p(l(\eta_T))$. Moreover, it can be shown that

$$\begin{aligned}
& \frac{(\beta_{2T} - \beta_{1T})}{l(\eta_T)} \frac{\sum_{t=[\tau_0 T]+1}^{[\tau T]} y_{t-1}^2 \sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau T]} y_{t-1}^2} \\
= \;& \frac{O_p(g^2(T)l(\eta_T))O_p(\sqrt{Tg(T)}l(\eta_T))}{g(T)l(\eta_T)\big(O_p(Tg(T)l(\eta_T)) + O_p(g^2(T)l(\eta_T))\big)} \\
= \;& O_p\Big(\sqrt{\frac{g(T)}{T}}\Big) = o_p(1);
\end{aligned} \tag{3.11}$$

$$\frac{\sum_{t=1}^{[\tau T]} y_{t-1}^2}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} = 1 + \frac{\sum_{t=[\tau_0 T]+1}^{[\tau T]} y_{t-1}^2}{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} = 1 + \frac{O_p(g^2(T)l(\eta_T))}{O_p(Tg(T)l(\eta_T))} = 1 + o_p(1);$$

$$\begin{aligned}
\frac{y_{[\tau_0 T]}}{\sqrt{g(T)l(\eta_T)}} &= \Big(1 - \frac{c}{T} - \frac{1}{g(T)}\Big)^{[\tau_0 T]} \frac{y_0}{\sqrt{g(T)l(\eta_T)}} + \sum_{t=0}^{[\tau_0 T]-1} \Big(1 - \frac{c}{T} - \frac{1}{g(T)}\Big)^t \frac{\varepsilon_{[\tau_0 T]-t}}{\sqrt{g(T)l(\eta_T)}} \\
&\Rightarrow \int_0^\infty \exp(-s) dW_1(s) \overset{d}{=} B_a(\frac{1}{2})
\end{aligned}$$

whose proof is similar to those of (3.5) and (3.6) ;

$$\begin{aligned}
& \frac{1}{g^2(T)l(\eta_T)} \sum_{t=0}^{[\nu g(T)]-1} y_{[\tau_0 T]+t}^2 \\
= \;& \frac{1}{g(T)} \sum_{t=0}^{[\nu g(T)]-1} \Big(\frac{1}{\sqrt{g(T)l(\eta_T)}} \sum_{i=0}^{t-1} \beta_{2T}^i \varepsilon_{[\tau_0 T]+t-i} + \frac{\beta_{2T}^t y_{[\tau_0 T]}}{\sqrt{g(T)l(\eta_T)}} \Big)^2 \\
\Rightarrow \;& \int_0^\nu \Big(I(W_2, c, \tau_0, t) + B_a(\frac{1}{2}) \Big)^2 dt
\end{aligned}$$

and

$$\frac{1}{\sqrt{g(T)}} \sum_{t=0}^{[\nu g(T)]-1} \frac{y_{[\tau_0 T]+t}}{\sqrt{g(T)l(\eta_T)}} \frac{\varepsilon_{[\tau_0 T]+t+1}}{\sqrt{l(\eta_T)}} \Rightarrow \int_0^\nu \Big(I(W_2, c, \tau_0, t) + B_a(\frac{1}{2}) \Big) dI(W_2, c, \tau_0, t)$$

by virtue of Lemma 3.3. Thus, by recalling the equation B.4 on page 120 in Chong (2001) when $\tau_0 \le \tau \le \overline{\tau}$

$$\begin{aligned}
& RSS_T(\tau) - RSS_T(\tau_0) \\
= \;& 2(\beta_{2T} - \beta_{1T}) \Big(\frac{\sum_{t=1}^{[\tau_0 T]} y_{t-1}^2 \sum_{t=[\tau_0 T]+1}^{[\tau T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau T]} y_{t-1}^2} - \frac{\sum_{t=[\tau_0 T]+1}^{[\tau T]} y_{t-1}^2 \sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\sum_{t=1}^{[\tau T]} y_{t-1}^2} \Big) \\
& + (\beta_{2T} - \beta_{1T})^2 \frac{\sum_{t=[\tau_0 T]+1}^{[\tau T]} y_{t-1}^2 \sum_{t=1}^{[\tau_0 T]} y_{t-1}^2}{\sum_{t=1}^{[\tau T]} y_{t-1}^2} + \Lambda_T(\tau),
\end{aligned}$$

we have

$$
\begin{aligned}
&\frac{RSS_T(\tau) - RSS_T(\tau_0)}{l(\eta_T)} \\
= \;& \frac{2(\beta_{2T} - \beta_{1T})}{l(\eta_T)} \cdot \sum_{t=[\tau_0 T]+1}^{[\tau T]} y_{t-1}\varepsilon_t \cdot (1 + o_p(1)) + \frac{(\beta_{2T} - \beta_{1T})^2}{l(\eta_T)} \sum_{t=[\tau_0 T]+1}^{[\tau T]} y_{t-1}^2 \cdot (1 + o_p(1)) + o_p(1) \\
= \;& \frac{2(1 + o_p(1))}{g(T) l(\eta_T)} \sum_{t=0}^{[\nu g(T)]-1} y_{[\tau_0 T]+t}\varepsilon_{[\tau_0 T]+t+1} + \frac{(1 + o_p(1))}{g^2(T) l(\eta_T)} \sum_{t=0}^{[\nu g(T)]-1} y_{[\tau_0 T]+t}^2 + o_p(1) \\
\Rightarrow \;& 2\int_0^\nu \Big(I(W_2, c, \tau_0, t) + B_a(\frac{1}{2})\Big) dI(W_2, c, \tau_0, t) + \int_0^\nu \Big(I(W_2, c, \tau_0, t) + B_a(\frac{1}{2})\Big)^2 dt \\
= \;& -2B_a^2(\frac{1}{2}) \cdot \Bigg\{ -\frac{I(W_2, c, \tau_0, \nu)}{B_a(\frac{1}{2})} - \int_0^\nu \frac{I(W_2, c, \tau_0, t)}{B_a^2(\frac{1}{2})} dI(W_2, c, \tau_0, t) - \\
& \qquad \int_0^\nu \Big(\frac{I(W_2, c, \tau_0, t)}{2B_a(\frac{1}{2})} + 1\Big) \frac{I(W_2, c, \tau_0, t)}{B_a(\frac{1}{2})} dt - \frac{\nu}{2}\Bigg\}.
\end{aligned}
$$

Applying the continuous mapping theorem for argmax functionals (cf. Kim and Pollard (1990)), we have

$$
\begin{aligned}
(\beta_{2T} - \beta_{1T}) T(\hat{\tau}_T - \tau_0) \;&=\; \hat{\nu} = \arg\min_{\nu \in R} \big\{RSS_T(\tau) - RSS_T(\tau_0)\big\} \\
&=\; \arg\min_{\nu \in R} \Big\{\frac{RSS_T(\tau) - RSS_T(\tau_0)}{l(\eta_T)}\Big\} \\
&\Rightarrow\; \arg\min_{\nu \in R} \Big\{ -2B_a^2(\frac{1}{2}) \Big(\frac{C^*(\nu)}{B_a(\frac{1}{2})} - \frac{|\nu|}{2}\Big)\Big\} \\
&=\; \arg\max_{\nu \in R} \Big\{\frac{C^*(\nu)}{B_a(\frac{1}{2})} - \frac{|\nu|}{2}\Big\},
\end{aligned}
$$

where $C^*(\nu)$ is defined as in Theorem 1.1. The proofs are complete. □

4 Proof of Theorem 1.2

The following lemmas will be used in the proof of Theorem 1.2:

Lemma 4.1 *Suppose assumption C1 is satisfied and $\beta_1 = \beta_{1T} = 1 - c/T$, where c is a fixed constant, then for any $0 \le \tau \le \tau_0$,*

$$
Q_T(\tau) := \frac{1}{\sqrt{T l(\eta_T)}} \sum_{t=1}^{[\tau T]} \beta_{1T}^{T-t} \varepsilon_t \Rightarrow e^{-c(1-\tau)} W(\tau) - c\int_0^\tau e^{-c(1-s)} W(s) ds = G(W, c, \tau).
$$

Lemma 4.2 *Let $\{y_t\}$ be generated according to Model (1.1), where $\beta_1 = \beta_{1T} = 1 - c/T$ for a constant c. Under assumptions C1-C3, the following results hold jointly:*
(4a) $\frac{1}{Tl(\eta_T)} \sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t \Rightarrow \frac{1}{2} e^{2c(1-\tau_0)} G^2(W,c,\tau_0) + c\int_0^{\tau_0} e^{2c(1-t)} G^2(W,c,t)dt - \frac{\tau_0}{2}$,
(4b) $\frac{1}{T^2 l(\eta_T)} \sum_{t=1}^{[\tau_0 T]} y_{t-1}^2 \Rightarrow \int_0^{\tau_0} e^{2c(1-t)} G^2(W,c,t)dt$.

Lemma 4.3 *Let $\{y_t\}$ be generated according to Model (1.1), where $\beta_1 = \beta_{1T} = 1 - c/T$ for a fixed constant c and $|\beta_2| < 1$, Under assumptions C1-C3, the following results hold jointly:*
(5a) $\frac{1}{\sqrt{T l(\eta_T)}} \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t \Rightarrow \overline{W}(B(c,\tau_0))/\sqrt{1-\beta_2^2}$,
(5b) $\frac{1}{Tl(\eta_T)} \sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2 \Rightarrow \frac{1-\tau_0+e^{2c(1-\tau_0)}G^2(W,c,\tau_0)}{1-\beta_2^2}$,
where $B(c,\tau_0) = (1-e^{-2c\tau_0})/(2c) + 1 - \tau_0$.

Proof of Theorem 1.2. It is not difficult to show that $\hat{\tau}_T$ is T-consistent by similar arguments in the proof of Theorem 1.1. To prove a stronger result in (1.3), we follow Appendix K in Chong (2001) with some modifications. For $m = 0, 1, \cdots$ and $m/T \to 0$,

$$\frac{1}{Tl(\eta_T)} RSS_T(\tau_0 - \frac{m}{T}) \Rightarrow 1 + h_1(m)$$

with

$$h_1(m) = \frac{(1-\tau_0+e^{2c(1-\tau_0)}G^2(W,c,\tau_0))(1-\beta_2^2)me^{2c(1-\tau_0)}G^2(W,c,\tau_0)}{(1-\beta_2^2)me^{2c(1-\tau_0)}G^2(W,c,\tau_0)+1-\tau_0+e^{2c(1-\tau_0)}G^2(W,c,\tau_0)}$$

and

$$\frac{1}{Tl(\eta_T)} RSS_T(\tau_0 + \frac{m}{T}) \Rightarrow 1 + h_2(m)$$

with

$$h_2(m) = \frac{(1-\beta_2)e^{2c(1-\tau_0)}G^2(W,c,\tau_0)(1-\beta_2^{2m})}{1+\beta_2}.$$

Since both $h_1(m)$ and $h_2(m)$ are increasing functions with respect to m, the first result in (1.3) can be proved. For more details, one is referred to Chong (2001).

To show the second and third parts of (1.3), it follows easily from the first result of (1.3) that $\hat{\beta}_1(\hat{\tau}_T)$ and $\hat{\beta}_1(\tau_0)$ have the same asymptotic distribution, and so do $\hat{\beta}_2(\hat{\tau}_T)$ and $\hat{\beta}_2(\tau_0)$. By Lemma 4.2, we have

$$\begin{aligned} T(\hat{\beta}_1(\hat{\tau}_T) - \beta_1) &\stackrel{d}{=} T(\hat{\beta}_1(\tau_0) - \beta_1) \\ &= \frac{\frac{1}{Tl(\eta_T)} \sum_{t=1}^{[\tau_0 T]} y_{t-1}\varepsilon_t}{\frac{1}{T^2 l(\eta_T)} \sum_{t=1}^{[\tau_0 T]} y_{t-1}^2} \\ &\Rightarrow \frac{\frac{1}{2} e^{2c(1-\tau_0)} G^2(W,c,\tau_0) + c\int_0^{\tau_0} e^{2c(1-t)} G^2(W,c,t)dt - \frac{\tau_0}{2}}{\int_0^{\tau_0} e^{2c(1-t)} G^2(W,c,t)dt}. \end{aligned}$$

Similarly, it follows from Lemma 4.3 that

$$\begin{aligned}
\sqrt{T}(\hat\beta_2(\hat\tau_T)-\beta_2) &\overset{d}{=} \sqrt{T}(\hat\beta_2(\tau_0)-\beta_2)\\
&= \frac{\frac{1}{\sqrt{T}l(\eta_T)}\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}\varepsilon_t}{\frac{1}{Tl(\eta_T)}\sum_{t=[\tau_0 T]+1}^{T} y_{t-1}^2}\\
&\Rightarrow \frac{\sqrt{1-\beta_2^2}\cdot \overline{W}(B(c,\tau_0))}{1-\tau_0+e^{2c(1-\tau_0)}G^2(W,c,\tau_0)}.
\end{aligned}$$

To derive the limiting distribution of $\hat\tau_T$ for shrinking break, we let $\beta_2=\beta_{2T}=\beta_{1T}-1/\sqrt{Tg(T)}$, where $g(T)>0$ with $g(T)\to\infty$ and $g(T)/\sqrt{T}\to 0$. Moreover, let ν be a constant. For $\tau=\tau_0+\nu g(T)/T$ and $\nu\le 0$, following Appendix K in Chong (2001), we have

$$\frac{RSS_T(\tau)-RSS_T(\tau_0)}{l(\eta_T)} \Rightarrow -2e^{c(1-\tau_0)}G(W_1,c,\tau_0)W_1(|\nu|)+|\nu|e^{2c(1-\tau_0)}G^2(W_1,c,\tau_0).$$

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

$$\frac{RSS_T(\tau)-RSS_T(\tau_0)}{l(\eta_T)} \Rightarrow -2e^{c(1-\tau_0)}G(W_1,c,\tau_0)W_2(\nu)+\nu e^{2c(1-\tau_0)}G^2(W_1,c,\tau_0).$$

For more details, one is referred to Chong (2001). Thus, by applying the continuous mapping theorem for argmax functionals, we have

$$\begin{aligned}
& (\beta_{1T}-\beta_{2T})^2T^2(\hat\tau_T-\tau_0)\\
=\ & \frac{T}{g(T)}(\hat\tau_T-\tau_0)=\hat\nu\\
=\ & \arg\min_{\nu\in R}\left\{RSS_T(\tau)-RSS_T(\tau_0)\right\}\\
=\ & \arg\min_{\nu\in R}\left\{\frac{RSS_T(\tau)-RSS_T(\tau_0)}{l(\eta_T)}\right\}\\
\Rightarrow\ & \arg\min_{\nu\in R}\left\{-2e^{2c(1-\tau_0)}G^2(W_1,c,\tau_0)\Big(\frac{B^*(\nu)}{e^{c(1-\tau_0)}G(W_1,c,\tau_0)}-\frac{|\nu|}{2}\Big)\right\}\\
=\ & \arg\max_{\nu\in R}\left\{\frac{B^*(\nu)}{e^{c(1-\tau_0)}G(W_1,c,\tau_0)}-\frac{|\nu|}{2}\right\},
\end{aligned}$$

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

Acknowledgements: The authors would like to thank an associate editor and two anonymous referees for their constructive and illuminating comments which have significantly improved the present paper.

References

[1] Andrews, D. W. K. (1987). Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica* **55** (6), 1465-1471.

[2] Andrews, D. W. K. and Guggenberger, P. (2008). Asymptotics for stationary very nearly unit root processes. *Journal of Time Series Analysis* **29** (1), 203-212.

[3] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd edn. John Wiley and Sons, New York.

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

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

[6] Csörgő, M., Szyszkowicz, B. and Wang, Q. Y. (2003). Donsker's theorem for self-normalized partial sums processes. *Annals of Probability* **31** (3), 1228-1240.

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

[8] Giné, E., Götze, F. and Mason, D. M. (1997). When is the Student t-statistic asymptotically standard normal? *Annals of Probability* **25** (3), 1514–1531.

[9] Hansen, B. E. (1992). Tests for parameter instability in regressions with I(1) processes. *Journal of Business and Economic Statistics* **10** (3), 321-335.

[10] Huang, S. H., Pang, T. X. and Weng, C. (2012). Limit theory for moderate deviations from a unit root under innovations with a possibly infinite variance. (*Methodology and Computing in Applied Probability*, forthcoming)

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

[12] Mankiw, N. G. and Miron, J. A. (1986). The changing behavior of the term structure of interest rates. *Quarterly Journal of Economics* **101** (2), 221-228.

[13] 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. *American Economic Review* **77** (3), 358-374.

[14] Pang, T. X. and Zhang, D. N. (2011). Asymptotic inferences for an AR(1) model with a change point and possibly infinite variance. (accepted by *Communications in Statistics: Theory and Methods*)

[15] Pang, T. X., Zhang D. N. and Chong, T. L. (2013). Asymptotic inferences for an AR(1) model with a change point: stationary and nearly non-stationary cases. (Available at http://arxiv.org/abs/1306.1294)

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

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