Joint Asymptotic Properties of Stopping Times and Sequential Estimators for Stationary First-order Autoregressive Models

Hitomi, Kohtaro; Nagai, Keiji; Nishiyama, Yoshihiko; Tao, Junfan

KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No. 1060

"Joint Asymptotic Properties of Stopping Times and Sequential Estimators for Stationary First-order Autoregressive Models"

Kohtaro Hitomi, Keiji Nagai, Yoshihiko Nishiyama, and Junfan Tao

June 2021

KYOTO UNIVERSITY

KYOTO, JAPAN

Joint Asymptotic Properties of Stopping Times and Sequential Estimators for Stationary First-order Autoregressive Models

Kohtaro Hitomi,[*] Keiji Nagai,[†] Yoshihiko Nishiyama,[‡] and Junfan Tao[§]

June 1, 2021

Abstract

Currently, because online data is abundant and can be collected more easily , people often face the problem of making correct statistical decisions as soon as possible. If the online data is sequentially available, sequential analysis is appropriate for handling such a problem. We consider the joint asymptotic properties of stopping times and sequential estimators for stationary first-order autoregressive (AR(1)) processes under independent and identically distributed errors with zero mean and finite variance. Using the stopping times introduced by Lai and Siegmund (1983) for AR(1), we investigate the joint asymptotic properties of the stopping times, the sequential least square estimator (LSE), and the estimator of σ^2. The functional central limit theorem for nonlinear ergodic stationary processes is crucial for obtaining our main results with respect to their asymptotic properties. We found that the sequential least square estimator and stopping times exhibit joint asymptotic normality. When σ^2 is estimated, the joint limiting distribution degenerates and the asymptotic variance of the stopping time is strictly smaller than that of the stopping time with a known σ^2. [1]

1 Introduction

Consider a scalar stationary first-order autoregressive (AR(1)) process $\{x_n\}$ on probability space $(\Omega, \mathcal{F}, P)$ with an initial value x_0 independent of ϵ_n, $n \geq 1$:

$$x_n = \beta x_{n-1} + \epsilon_n, \quad n = 1, 2, \ldots. \tag{1}$$

We assume that $\epsilon_1,\epsilon_2, \ldots$ are independent and identically distributed (i.i.d.) random variables with $E(\epsilon_1) = 0$, $E(\epsilon_1^2) = \sigma^2 \in (0, \infty)$ and that initial value $x_0 \in L^2$ is independent of $\{\epsilon_n\}$. The LSE is given by

$$\hat{\beta}_N = \sum_{n=1}^{N} x_n x_{n-1} / \sum_{n=1}^{N} x_{n-1}^2. \tag{2}$$

It is well known that when a process is stationary AR(1), the LSE $\hat{\beta}_N$ has asymptotic normality; that is, $\sqrt{N}\left(\hat{\beta}_N - \beta\right) \to N(0, 1-\beta^2)$ as $N \to \infty$. If $\epsilon_1, \epsilon_2, \ldots$ is normally distributed, the observed Fisher

[*]Kyoto Institute of Technology, Matsugasaki, Sakyo-ku, Kyoto, 606-8585, Japan, hitomi@kit.ac.jp

[†]Yokohama National University, 79-4 Tokiwadai, Hodogaya-ku, Yokohama, 240-8501, Japan, nagai-keiji-hs@ynu.ac.jp

[‡]Institute of Economic Research, Kyoto University, Yoshidahonmachi, Sakyo-ku, Kyoto, 606-8501, Japan, nishiyama.yoshihiko.3u@kyoto-u.ac.jp

[§]JSPS International Research Fellow (Kyoto University), Institute of Economic Research, Kyoto University, Yoshidahonmachi, Sakyo-ku, Kyoto, 606-8501, Japan, tao.junfan.46r@st.kyoto-u.ac.jp

[1]Supported by JSPS KAKENHI Grants Numbers JP17K03656, JP18K01543, JP19F19312, JP19H01473, JP19K21691, JP20K01589.

MSC 2000 subject classifications. Primary 62M10, 62L12, Secondary 60J70, 60F17

Key words and phrases. Observed Fisher information, joint asymptotic normality, functional central limit theorem in $D[0, \infty)$, Anscombe's Theorem

information of β is given by

$$I_N = -\frac{\partial^2}{\partial\beta^2}\left(\beta\sum_{n=1}^{N} x_{n-1}x_n - \frac{1}{2}\beta^2\sum_{n=1}^{N} x_{n-1}^2\right)/\sigma^2 = \sum_{n=1}^{N} x_{n-1}^2/\sigma^2. \tag{3}$$

Lai and Siegmund (1983) (LS83 hereafter) considered a sequentially observed AR(1) process and proposed to evaluate the LSE at stopping time τ_{1c} defined by

$$\tau_{1c} = \inf\left\{N > 1 : \sum_{n=1}^{N} x_{n-1}^2/\sigma^2 \geq c\right\}, \tag{4}$$

for some prescribed $c > 0$. They also defined a feasible stopping time, τ_{2c}, in (and stopping times. We call the above procedure the Anscombe-Donsker-Skorohod 9) by replacing σ^2 with its estimator in (4).

For the stopping time defined in (4), we define the sequential LSE by setting $N = \tau_{1c}$ in (2). In addition, the asymptotic normality of $\hat{\beta}_{\tau_{1c}}$ has been demonstrated by LS83. One of their main results is as $c \to \infty$,

$$P_\beta\left[I_{\tau_{1c}}^{1/2}\left(\hat{\beta}_{\tau_{1c}} - \beta\right) \leq z\right] \to \Phi(z) \quad \text{uniformly in } \beta \in [-1, 1], \tag{5}$$

where Φ is the standard normal distribution function.[2] For $|\beta| < 1$, they showed

$$\tau_{1c}/c \to \sigma^2/\gamma(0) = 1 - \beta^2 \quad a.s., \tag{6}$$

where γ is the covariance function of $\{x_n\}$.

After LS83, some studies on the sequential estimation in AR models have been conducted. Sriram proposed the asymptotically risk efficient sequential estimation using a stopping time defined by some loss function. See, for example, Sriram (1987, 1988) and Sriram and Iaci (2014). Konev and Pergamenshchikov (1986) and Pergamenshchikov (1991) considered the same estimation problem as LS83 and proved the asymptotic normality of the stopping time for stationary cases. Sriram, Basawa, and Huggins (1991) studied sequential estimation of the critical and subcritical offspring mean of a branching process with immigration. Shete and Sriram (1998) obtained the asymptotic normality of the stopping time for the critical branching process with immigration. Galtchouk and Konev (2003) considered sequential estimation of a stationary AR(1) process with a constant using a stopping time defined by the trace of the information matrix. Galtchouk and Konev (2004) extended it to AR(p) processes. Galtchouk and Konev (2005) proposed to use a stopping time depending on the minimum eigenvalue of the information matrix for stationary AR(p). Galtchouk and Konev (2006, 2011) respectively proved the uniform joint asymptotic normality of AR(2) and AR(p) parameter estimation for both stationary and unit-root cases.

The contributions of the present paper in sequential analysis are summarized as follows.

First, we prove the joint asymptotic normality of the sequential estimators for (β, σ^2) and stopping times. We call the above procedure the Anscombe-Donsker-Skorohod es τ_{1c} in (4) and τ_{2c} in (9). When σ^2 is estimated, the joint distribution degenerates and the asymptotic variance of τ_{1c} is strictly larger than that of τ_{2c}.

Second, we introduce the following new methodology to examine the statistical properties in sequential analysis. We consider the application of functional central limit theorems for dependent observations, including Donsker's theorem and Skorohod's representation theorem in $D[0,\infty)$: the set of right continuous functions on $[0,\infty)$ with left limits (details are presented in Billingsley(1999)). In traditional sequential analysis, Anscombe's theorem (Anscombe (1952) or Theorem 1.4 in Woodroofe (1982)) is used along with discrete-time weak convergence to obtain the asymptotic distributions of sequential statistics. Modifying Anscombe's theorem by combining Donsker's invariance principle and Skorohod's representation theorem (see also Billingsley(1999)) in $D[0,\infty)$, we derive the joint asymptotic distribution of sequential statistics and stopping times. We call the above procedure the Anscombe-Donsker-Skorohod approach.

[2]To obtain the uniform asymptotic normality, LS83 assumed that the initial value x_0 does not depend on β and $\sup_{|\beta|\leq 1} P_\beta\left\{x_n^2 > a\right\} \to 0 \quad as \quad a \to \infty$ for each fixed $n \geq 0$. These assumptions do not hold when $\{x_n\}$ is a strongly stationary process.

The remainder of this paper is organized as follows. In section 2, we derive the asymptotic properties of the sequential estimators and the stopping times. In section 3, simulation studies are conducted to verify our findings. Section 4 contains a summary and further remarks. Some proofs are given in the Appendix.

2 Asymptotic properties of sequential estimation

In this section, we derive the main results for an AR(1) process with $|\beta| < 1$. We obtain the strong limits of stopping times τ_{1c} in (4) and τ_{2c} in (9) and derive the joint asymptotic normality of the sequential estimators for (β, σ^2) and stopping times τ_{1c} and τ_{2c}.

Suppose we sequentially observe $\{x_n\}$ from the stationary AR(1) model in (1). When the initial value, x_0, has a stationary distribution, $\{x_n\}$ has covariance function given by

$$\gamma(m) = \beta^{|m|}\sigma^2 / \left(1 - \beta^2\right). \tag{7}$$

We assume that initial value x_0 is a L^2 random variable and independent of $\epsilon_1, \epsilon_2, \ldots$. The estimator of σ^2 is given by

$$s_N^2 = \sum_{n=1}^{N} \left(x_n - \hat{\beta}_N x_{n-1}\right)^2 / N. \tag{8}$$

Like τ_{1c} in (4), we set a feasible stopping time

$$\tau_{2c} = \inf\left\{N > 1 : \sum_{n=1}^{N} x_{n-1}^2 / s_N^2 \geq c\right\}. \tag{9}$$

Our purpose here is to study the asymptotic behavior of $\left(\hat{\beta}_{\tau_{1c}}, \tau_{1c}\right)$ and $\left(\hat{\beta}_{\tau_{2c}}, s^2_{\tau_{2c}}, \tau_{2c}\right)$.

According to LS83, $\tau_{1c}/c \to 1 - \beta^2$ almost surely as $c \to \infty$. We show that the same result holds for τ_{2c}/c. [3]

Theorem 1. *Suppose we sequentially observe $\{x_n\}_{n=1,2,\ldots}$ from the stationary AR(1) model in (1) with initial value x_0 independent of $\epsilon_1, \epsilon_2, \ldots$. Then, $\hat{\beta}_N \to \beta$ and $s_N^2 \to \sigma^2$ a.s. and and stopping times. We call the above procedure the Anscombe-Donsker-Skorohod*

$$\lim_{c\to\infty} \tau_{1c} = \lim_{c\to\infty} \tau_{2c} = \infty \ a.s. \tag{10}$$

$$\lim_{c\to\infty} \frac{\tau_{1c}}{c} = \lim_{c\to\infty} \frac{\tau_{2c}}{c} = \frac{\sigma^2}{\gamma(0)} = 1 - \beta^2 \quad a.s.. \tag{11}$$

Finally, the sequential estimates of β and σ^2 have strong consistency:

$$\lim_{c\to\infty} \hat{\beta}_{\tau_{1c}} = \lim_{c\to\infty} \hat{\beta}_{\tau_{2c}} = \beta \ a.s \quad \text{and} \quad \lim_{c\to\infty} s^2_{\tau_{2c}} = \sigma^2 \ a.s. \tag{12}$$

Proof. According to Lemma 12 in Appendix 5.3, $\sum_{n=1}^{N} x_{n-1}^2/N \to \gamma(0) \quad a.s.$. Since $\sum_{n=1}^{N} x_{n-1}^2 \uparrow \infty \quad a.s.$, the inequalities

$$\sum_{n=1}^{\tau_{1c}-1} x_{n-1}^2 < c\sigma^2 \leq \sum_{n=1}^{\tau_{1c}} x_{n-1}^2 \tag{13}$$

imply $\lim_{c\to\infty} \tau_{1c} = \infty \ a.s.$. Lemma 12 also gives $\sum_{n=1}^{N} x_{n-1}\epsilon_n/N \to 0 \quad a.s.$. Hence, $\hat{\beta}_N - \beta = \sum_{n=1}^{N} x_{n-1}\epsilon_n / \sum_{n=1}^{N} x_{n-1}^2 \to 0 \quad a.s.$ and and stopping times. We call the above procedure the Anscombe-Donsker-Skorohod

$$s_N^2 = \sum_{n=1}^{N} \left\{\left(\beta - \hat{\beta}_N\right) x_{n-1} + \epsilon_n\right\}^2 / N \to \sigma^2 \quad a.s..$$

[3] $\lim_{c\to\infty} \tau_{1c}/c = \lim_{c\to\infty} \tau_{2c}/c = \sigma^2/\gamma(0)$ also holds for any stationary AR(p) process $x_n = \beta_1 x_{n-1} + \cdots + \beta_p x_{n-p} + \epsilon_n$ with covariance function $\gamma(m)$.

As to τ_{2c}, according to Lemma 12 in Appendix 5.3, $s_N^2 \to \sigma^2 \quad a.s.$ Let $\Omega' = \left\{\lim_{N\to\infty} s_N^2 = \sigma^2\right\} \cap \{\lim_{c\to\infty} \tau_{1c} = \infty\}$, fix $\omega \in \Omega'$ and $\epsilon \in (0, \sigma^2)$, choose $N_0 = N_0(\omega, \epsilon)$ so that$\sigma^2 - \epsilon < s_N^2 < \sigma^2 + \epsilon$, for any $N \geq N_0$. Letting $c' = c(1 - \epsilon/\sigma^2)$ and $c'' = c(1 + \epsilon/\sigma^2)$,

$$c'\sigma^2 < cs_N^2 < c''\sigma^2. \tag{14}$$

Now, for any $m = 1, \ldots, N_0 - 1$, let us take a large c_0 to obtain $\sum_{n=1}^{m} x_{n-1}^2 < cs_m^2$ and $N_0 < \tau_{1c'}$ for any $c > c_0$. When m satisfies $N_0 \leq m < \tau_{1c'}$ with $c > c_0$,

$$\sum_{n=1}^{m} x_{n-1}^2 < c'\sigma^2 < cs_m^2.$$

Hence, for any $m = 1, \ldots, \tau_{1c'} - 1$, $\sum_{n=1}^{m} x_{n-1}^2 < cs_m^2$, which implies $\tau_{1c'} \leq \tau_{2c}$ for $c > c_0$. Then, $N_0 < \tau_{1c'} \leq \tau_{2c}$ and $\tau_{1c'} \leq \tau_{1c''}$. Thus,

$$\sum_{n=1}^{\tau_{2c}-1} x_{n-1}^2 < cs_{\tau_{2c}-1}^2 \leq c''\sigma^2 \leq \sum_{n=1}^{\tau_{1c''}} x_{n-1}^2,$$

which gives $\tau_{2c} \leq \tau_{1c''}$. Therefore we have $\tau_{1c'} \leq \tau_{2c} \leq \tau_{1c''}$ for sufficiently large c, which implies $\tau_{2c} \to \infty\ a.s.$

Next, we will prove $\lim_{c\to\infty} \tau_{1c}/c = \sigma^2/\gamma(0)$. Dividing (13) by $\tau_{1c}\sigma^2$, we obtain

$$\frac{1}{\tau_{1c}\sigma^2} \sum_{n=1}^{\tau_{1c}-1} x_{n-1}^2 < \frac{c}{\tau_{1c}} \leq \frac{1}{\tau_{1c}\sigma^2} \sum_{n=1}^{\tau_{1c}} x_{n-1}^2.$$

Letting $c \to \infty$ on Ω' we have (11). As for τ_{2c}, since

$$\frac{1}{\tau_{2c}} \sum_{n=1}^{\tau_{2c}-1} x_{n-1}^2 < \frac{cs_{\tau_{2c}-1}^2}{\tau_{2c}}, \quad \frac{cs_{\tau_{2c}}^2}{\tau_{2c}} \leq \frac{1}{\tau_{2c}} \sum_{n=1}^{\tau_{2c}} x_{n-1}^2$$

we have $\gamma(0) \leq \sigma^2 \liminf_{c\to\infty} c/\tau_{2c} \leq \sigma^2 \limsup_{c\to\infty} c/\tau_{2c} \leq \gamma(0)$. Hence, we conclude that $\lim_{c\to\infty} \tau_{2c}/c = \sigma^2/\gamma(0) \quad a.s.$. □

To obtain asymptotic results with respect to the sequential statistics, we apply the theory of convergence of random elements in $D[0,\infty)$, the set of right-continuous functions on $[0,\infty)$ with left limits. Because we consider stopping times with the unbounded range of integers, $D[0,\infty)$ is thought to be appropriate for characterizing the limiting behavior of sequential statistics. Although it is difficult to deal with the weak convergence of random elements in $D[0,\infty)$ with respect to such a metric d_∞°defined in (36), we utilize the fact that when the limit function is continuous, the convergence with respect to Skorohod's metric d_∞° is equivalent to the uniform convergence on compacta. To explain this, we define the metric of two functions $f$and g in $D[0,\infty)$ which gives the uniform convergence on compacta;

$$\rho(f, g) = \sum_{m=1}^{\infty} 2^{-m} (\|f - g\|_m \wedge 1). \tag{15}$$

where for $m > 0$

$$\|f\|_m = \sup_{t\in[0,m]} |f(t)|, \quad \|f\|_\infty = \sup_{t\in[0,\infty)} |f(t)|. \tag{16}$$

Lemma 2. *Let the metric d_∞° be defined as in (36) and ρ in (15). Then, $d_\infty^\circ(f_n, f) \to 0$ of if and only if $\lim_{n\to\infty} \rho(f_n, f) = 0$ for $f_n \in D[0,\infty)$, and $f \in C[0,\infty)$..*

The proof can be found in Appendix 5.1.

In a sequential sampling scheme, using Skorokhod's representation (Billingsley (1999), Theorem 6.7, p. 70) and Lemma 2, we obtain the following theorem, which provides a unifying method for deriving the joint asymptotic limits of stopping times and sequential statistics under stationary and nonstationary processes as in Lemma 6 and Theorem 7.

Theorem 3. *(Anscome-Donsker-Skorohod) Let random elements $X_c, X \in D[0,\infty)$ and random variables $U_c, U \in [0,\infty)$ for $c > 0$. If $X \in C[0,\infty)$ and $(X_c, U_c) \Rightarrow (X, U)$ in $D[0,\infty) \times [0,\infty)$ as $c \to \infty$, then $X_c(U_c) \Rightarrow X(U)$ as $c \to \infty$.*

Proof. Since $D[0,\infty)$ is separable, we can appply Skorohod's representation theorem and obtain a probabilityand stopping times. We call the above procedure the Anscombe-Donsker-Skorohod space $\left(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P}\right)$, which has the random elements $\tilde{X}_c, \tilde{X} \in D[0,\infty)$ and the random variables $\tilde{U}_c, \tilde{U} \in [0,\infty)$ such that $\left(\tilde{X}, \tilde{\tilde{U}}\right) \sim \left(X, \tilde{U}\right)$ and $\left(\tilde{X}_c, \tilde{\tilde{U}}_c\right) \sim \left(X_c, \tilde{U}_c\right)$ with $\lim_{c\to\infty}\left(\tilde{X}_c(\tilde{\omega}), \tilde{U}_c(\tilde{\omega})\right) \to \left(\tilde{X}(\tilde{\omega}), \tilde{U}(\tilde{\omega})\right)$ for any $\tilde{\omega} \in \tilde{\Omega}$. For a fixed $\tilde{\omega} \in \tilde{\Omega}$, letting $M = M(\tilde{\omega}) > \tilde{U}(\tilde{\omega})$, then

$$\begin{aligned}
&\left|\tilde{X}_c\left(\tilde{U}_c(\tilde{\omega}), \tilde{\omega}\right) - \tilde{X}\left(\tilde{U}(\tilde{\omega}), \tilde{\omega}\right)\right| \\
\leq &\left|\tilde{X}_c\left(\tilde{U}_c(\tilde{\omega}), \tilde{\omega}\right) - \tilde{X}\left(\tilde{U}_c(\tilde{\omega}), \tilde{\omega}\right)\right| + \left|\tilde{X}\left(\tilde{U}_c(\tilde{\omega}), \tilde{\omega}\right) - \tilde{X}\left(\tilde{U}(\tilde{\omega}), \tilde{\omega}\right)\right| \\
\leq &\sup_{t\leq M}\left|\tilde{X}_c(t, \tilde{\omega}) - \tilde{X}(t, \tilde{\omega})\right| + \left|\tilde{X}\left(\tilde{U}_c(\tilde{\omega}), \tilde{\omega}\right) - \tilde{X}\left(\tilde{U}(\tilde{\omega}), \tilde{\omega}\right)\right| \to 0 \quad \text{as } c \to \infty.
\end{aligned}$$

□

The following theorem allows us to derive the joint asymptotic normality of the sequential estimates and stopping times. The proof is given in Appendix 5.1. Let $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ and $\pi_n : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ be the n-th coordinate map; $\pi_n(\omega) = \omega_n$ for $n \in \mathbb{Z}$ and $T : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ be the shift operator; $\pi_n(T\omega) = \omega_{n+1}$. Here, $\|\cdot\|$ represents the L^2 norm $\|\xi\| = \sqrt{E(\xi^2)}$, $\lfloor a \rfloor$ the integer part of a for $a > 0$, and $\mathbb{Z}_- = \{\ldots, -2, -1, 0\}$.

Theorem 4. *Let $\Omega = \mathbb{R}^{\mathbb{Z}}$, $\mathcal{F} = \mathcal{B}\left(\mathbb{R}^{\mathbb{Z}}\right)$, and P be a probability measure on $(\Omega, \mathcal{F})$. For $\omega = (\omega_n) \in \Omega$, let π_n be the n-th coordinate map; $\pi_n(\omega) = \omega_n$ for $n \in \mathbb{Z}$ and $T : \Omega \to \Omega$ be a measure-preserving and ergodic left-shift operator; $\pi_n(T\omega) = \omega_{n+1}$. Let $\mathcal{F}_0 = \sigma[\pi_k : k \leq 0]$ and $\mathcal{F}_n = T^{-n}\mathcal{F}_0$. Suppose $\xi_0(\omega) = g(\cdots, \omega_{-1}, \omega_0)$ is a L^2 random variable with a Borel function $g : \mathbb{R}^{\mathbb{Z}_-} \to \mathbb{R}$ and $\xi_n(\omega) = \xi_0(T^n\omega)$. If*

$$\sum_{h=1}^{\infty} \|E[\xi_h|\mathcal{F}_0]\| < \infty, \tag{17}$$

then $E[\xi_n] = 0$, and the series

$$\nu^2 = E\left[\xi_0^2\right] + 2\sum_{h=1}^{\infty} E[\xi_0\xi_h] \tag{18}$$

converges absolutely. Assume $\nu > 0$ and define

$$\theta_0 = \sum_{h=1}^{\infty} E[\xi_{n+h}|\mathcal{F}_n] \tag{19}$$

and

$$\eta_n = \xi_n + \theta_n - \theta_{n-1}. \tag{20}$$

Then, η_k is a ergodic stationary $\mathcal{F}_n$-martingale difference sequence with $E\left[\eta_n^2\right] = \nu^2$. Let $S_n = \sum_{k=1}^{n} \xi_k$ and $M_n = \sum_{k=1}^{n} \eta_k$, then as $c \to \infty$, $S_{\lfloor ct \rfloor}/\sqrt{c} - M_{\lfloor ct \rfloor}/\sqrt{c} \to 0$ and

$$M_{\lfloor ct \rfloor}/\nu\sqrt{c} \Rightarrow W$$

in the sense of $D[0,\infty)$, where W is a Brownian motion. Consequently,

$$S_{\lfloor ct \rfloor}/\nu\sqrt{c} \Rightarrow W \quad (c \to \infty).$$

Remark 5. Note that Theorem 4 is a simple extension of Theorem 19.1 in Billingsley (1999, p. 197). Unlike Billingsley's theorem, the filtration is not necessarily generated by the process to which we apply the functional central limit theorem.

The above theorem leads to the following lemma.

Lemma 6. *Suppose* $(\epsilon_n)_{n\in\mathbb{Z}}$ *is an i.i.d. sequence with* $E(\epsilon_n) = 0, E(\epsilon_n^2) = \sigma^2$, *and*

$$\omega^2 = E\left[\left(\epsilon_1^2 - \sigma^2\right)^2\right] < \infty \tag{21}$$

For $x_n = \sum_{k=0}^{\infty} \beta^k \epsilon_{n-k}$ *with* $|\beta| < 1$, *let*

$$\nu^2 = E\left[\left(x_0^2 - \gamma(0)\right)^2\right] + 2\sum_{n=1}^{\infty} E\left[\left(x_0^2 - \gamma(0)\right)\left(x_n^2 - \gamma(0)\right)\right]. \tag{22}$$

As $c \uparrow \infty$, *we have*

$$\sum_{n=1}^{\lfloor ct \rfloor} x_{n-1}\epsilon_n/\sqrt{c} \Rightarrow \sqrt{\gamma(0)}\sigma W^{(1)}(t) \tag{23}$$

$$\sum_{n=1}^{\lfloor ct \rfloor} \left(\epsilon_n^2 - \sigma^2\right)/\sqrt{c} \Rightarrow \omega W^{(2)}(t) \tag{24}$$

$$\sum_{n=1}^{\lfloor ct \rfloor} \left(x_{n-1}^2 - \gamma(0)\right)/\sqrt{c} \Rightarrow \nu W^{(3)}(t) \tag{25}$$

in the sense of $D[0,\infty)$, *where* $W^{(1)}$ *and* $W^{(2)}$ *are independent standard Brownian motions,*

$$W^{(3)}(t) = \left(\frac{2\beta\gamma(0)^{3/2}}{\sigma}W^{(1)}(t) + \frac{\gamma(0)\omega}{\sigma^2}W^{(2)}(t)\right)/\nu$$

and

$$\nu^2 = \left(\frac{4\beta^2\gamma(0)^3}{\sigma^2} + \frac{\gamma(0)^2\omega^2}{\sigma^4}\right). \tag{26}$$

Proof. To prove (25), we apply Theorem 4. Let

$$\xi_n = x_{n-1}^2 - \gamma(0) \tag{27}$$

Denote $\|\cdot\|$ the L^2 norm and put $\mathcal{F}_n = \sigma[x_k : k \le n] = \sigma[\epsilon_k : k \le n]$ for (27). We first prove that the following condition holds;

$$\sum_{n=1}^{\infty} \|E[|\xi_n|\mathcal{F}_0|]\| < \infty. \tag{28}$$

We have

$$\begin{aligned} E[\xi_n|\mathcal{F}_0] &= E\left[\left(\sum_{k=0}^{n-2}\beta^k\epsilon_{n-1-k} + \beta^{n-1}x_0\right)^2 - \gamma(0)\,|\mathcal{F}_0\right] \\ &= \sum_{k=0}^{n-2}\beta^{2k}\sigma^2 + \left(\beta^{n-1}x_0\right)^2 - \sum_{k=0}^{\infty}\beta^{2k}\sigma^2 \\ &= \beta^{2(n-1)}\left(x_0^2 - \gamma(0)\right) \end{aligned} \tag{29}$$

thus, $E\left[\left(E[\xi_n|\mathcal{F}_0]\right)^2\right] = \beta^{4(n-1)}E\left[\left(x_0^2 - \gamma(0)\right)^2\right]$. Hence, (28) holds.

For θ_n and η_n defined in (19) and (27), η_n become martingale differences with respect to $\mathcal{F}_n$. Since

$$\theta_0 = \sum_{n=1}^{\infty} E\left[\xi_n \left|\mathcal{F}_0\right.\right] = \sum_{n=1}^{\infty} \beta^{2(n-1)} \left(x_0^2 - \gamma(0)\right) = \sum_{n=1}^{\infty} \beta^{2(n-1)} \xi_1 = \xi_1/(1-\beta^2)$$

and $\theta_1 = \xi_2/(1-\beta^2)$,

$$\begin{aligned} \eta_1 &= \left(\xi_2 - \beta^2 \xi_1\right)/(1-\beta^2) \\ &= \left(x_1^2 - \gamma(0) - \beta^2\left(x_0^2 - \gamma(0)\right)\right)/(1-\beta^2) \\ &= \left(\left(\beta x_0 + \epsilon_1\right)^2 - \gamma(0) - \beta^2\left(x_0^2 - \gamma(0)\right)\right)/(1-\beta^2) \\ &= \left(2\beta x_0 \epsilon_1 + \epsilon_1^2 - \sigma^2\right)/(1-\beta^2). \end{aligned}$$

Using Theorem 4, we can compute

$$\nu^2 = E\left[\eta_1^2\right] = \left(\frac{4\beta^2\gamma(0)^3}{\sigma^2} + \frac{\gamma(0)^2\omega^2}{\sigma^4}\right).$$

Hence,

$$\begin{aligned} \sum_{n=1}^{N}\left(x_{n-1}^2 - \gamma(0)\right) = \sum_{n=1}^{N}\xi_n &= \left(\theta_0 - \theta_N + \sum_{n=1}^{N}\eta_n\right) \\ &= \frac{1}{(1-\beta^2)}\left(x_0^2 - x_N^2 + 2\beta\sum_{n=1}^{N} x_{n-1}\epsilon_n + \sum_{n=1}^{N}\left(\epsilon_n^2 - \sigma^2\right)\right) \end{aligned} \tag{30}$$

Therefore, as $c \to \infty$, $W^{(3)}(t) = \left(\frac{2\beta\gamma(0)^{3/2}}{\sigma}W^{(1)}(t) + \frac{\gamma(0)\omega}{\sigma^2}W^{(2)}(t)\right)/\nu$. □

Combining the above lemma and lemma 12 in Appendix 5.3, we obtain the following theorem, which indicates that the asymptotic variance of τ_{1c} is strictly greater than that of τ_{2c} and there exists perfect negative correlation between $\hat{\beta}_{\tau_{2c}}$ and τ_{2c}.

Theorem 7. *(Sequential asymptotic normality) Let x_0 be an arbitrary L^2 random variable and independent of $\epsilon_1, \epsilon_2, \ldots$ with $\omega^2 = E\left[\left(\epsilon_1^2 - \sigma^2\right)^2\right] < \infty$. As $c \to \infty$, in the sense of $D[0,\infty)$,*

$$\begin{aligned} \begin{pmatrix} \sqrt{c}\left(\hat{\beta}_{\tau_{1c}} - \beta\right) \\ \sqrt{c}\left(\frac{\tau_{1c}}{c} - \frac{\sigma^2}{\gamma(0)}\right) \end{pmatrix} &\Rightarrow \begin{pmatrix} \frac{\sqrt{\gamma(0)}}{\sigma}W^{(1)}\left(\frac{\sigma^2}{\gamma(0)}\right) \\ -\frac{2\beta\gamma(0)^{1/2}}{\sigma}W^{(1)}\left(\frac{\sigma^2}{\gamma(0)}\right) - \frac{\omega}{\sigma^2}W^{(2)}\left(\frac{\sigma^2}{\gamma(0)}\right) \end{pmatrix} \\ &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -2\beta \\ -2\beta & 4\beta^2 + (1-\beta^2)\omega^2 \end{pmatrix}\right), \end{aligned} \tag{31}$$

and

$$\begin{pmatrix} \sqrt{c}\left(\hat{\beta}_{\tau_{2c}} - \beta\right) \\ \sqrt{c}\left(\frac{\tau_{2c}}{c} - \frac{\sigma^2}{\gamma(0)}\right) \end{pmatrix} \Rightarrow \begin{pmatrix} \frac{\sqrt{\gamma(0)}}{\sigma}W^{(1)}\left(\frac{\sigma^2}{\gamma(0)}\right) \\ -\frac{2\beta\gamma(0)^{1/2}}{\sigma}W^{(1)}\left(\frac{\sigma^2}{\gamma(0)}\right) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -2\beta \\ -2\beta & 4\beta^2 \end{pmatrix}\right). \tag{32}$$

Proof. In the following argument, we use Skorohod's representation (Billingsley(1999), Theorem 6.7, p.70) for the almost sure convergences (11) and the weak convergence (23). Theorem 1 and Lemma 6 imply

$$\sqrt{c}\left(\hat{\beta}_{\tau_{1c}} - \beta\right) = \frac{\sum_{n=1}^{\tau_{1c}} x_{n-1}\epsilon_n/\sqrt{c}}{\sum_{n=1}^{\tau_{1c}} x_{n-1}^2/c} \Rightarrow \frac{\sqrt{\gamma(0)}}{\sigma}W^{(1)}\left(\frac{\sigma^2}{\gamma(0)}\right) \sim W_1^{(1)}. \tag{33}$$

As for τ_{1c}, using the following inequality;

$$\frac{1}{\sqrt{c}}\sum_{n=1}^{\tau_{1c}-1}\left(x_{n-1}^2-\gamma(0)\right)<\sqrt{c}\gamma(0)\left(\frac{\sigma^2}{\gamma(0)}-\frac{\tau_{1c}}{c}\right)+\frac{\gamma(0)}{\sqrt{c}}\leq\frac{1}{\sqrt{c}}\sum_{n=1}^{\tau_{1c}}\left(x_{n-1}^2-\gamma(0)\right)+\frac{\gamma(0)}{\sqrt{c}},$$

and Theorem 4, we have $\sum_{n=1}^{\tau_{1c}}\left(x_{n-1}^2-\gamma(0)\right)/v\sqrt{c}\Rightarrow W^{(3)}\left(\gamma(0)/\sigma^2\right)$ which implies (31). Next,

$$\begin{aligned}\sqrt{c}\left(s_{\tau_{2c}}^2-\sigma^2\right)&=\frac{\sqrt{c}}{\tau_{2c}}\sum_{n=1}^{\tau_{2c}}\left(\left(x_n-\hat{\beta}_{\tau_{2c}}x_{n-1}\right)^2-\sigma^2\right)\\&=\frac{\sqrt{c}}{\tau_{2c}}\mathcal{F}_n\sum_{n=1}^{\tau_{2c}}\left[\left\{\left(\beta-\hat{\beta}_{\tau_{2c}}\right)x_{n-1}+\epsilon_n\right\}^2-\sigma^2\right]\\&=\sqrt{c}\left(\hat{\beta}_{\tau_{2c}}-\beta\right)^2\frac{1}{\tau_{2c}}\sum_{n=1}^{\tau_{2c}}x_{n-1}^2\\&-2\sqrt{c}\left(\hat{\beta}_{\tau_{2c}}-\beta\right)\frac{1}{\tau_{2c}}\sum_{n=1}^{\tau_{2c}}x_{n-1}\epsilon_n\\&+\frac{c}{\tau_{2c}}\frac{1}{\sqrt{c}}\sum_{n=1}^{\tau_{2c}}\left(\epsilon_n^2-\sigma^2\right).\end{aligned}$$

Then, as $c\to\infty$, for the last line of the equation, the first and second elements converge to 0. For the third element, applying Theorem 1 and Lemma 6, in the sense of $D[0,\infty)$,

$$\frac{\sqrt{c}}{\omega}\left(s_{\tau_{2c}}^2-\sigma^2\right)\Rightarrow\frac{\gamma(0)}{\sigma^2}W^{(2)}\left(\frac{\sigma^2}{\gamma(0)}\right).$$

The definition of τ_{2c} implies $\sum_{n=1}^{\tau_{2c}-1}x_{n-1}^2<s_{\tau_{2c}-1}^2c$, then

$$\frac{1}{\sqrt{c}}\sum_{n=1}^{\tau_{2c}-1}\left(x_{n-1}^2-\gamma(0)\right)<\frac{1}{\sqrt{c}}\left(s_{\tau_{2c}-1}^2c-\tau_{2c}\gamma(0)\right)+\frac{\gamma(0)}{\sqrt{c}}. \tag{34}$$

Since $s_{\tau_{2c}}^2c\leq\sum_{n=1}^{\tau_{2c}}x_{n-1}^2$, the right side of (34) can be written as

$$\frac{1}{\sqrt{c}}\left(s_{\tau_{2c}-1}^2c-\tau_{2c}\gamma(0)\right)+\frac{\gamma(0)}{\sqrt{c}}\leq\sqrt{c}\left(s_{\tau_{2c}-1}^2-s_{\tau_{2c}}^2\right)+\frac{1}{\sqrt{c}}\sum_{n=1}^{\tau_{2c}}\left(x_{n-1}^2-\gamma(0)\right)+\frac{\gamma(0)}{\sqrt{c}}. \tag{35}$$

For the left side of (35), we have

$$\frac{1}{\sqrt{c}}\left(s_{\tau_{2c}-1}^2c-\tau_{2c}\gamma(0)\right)=\sqrt{c}\left(s_{\tau_{2c}-1}^2-\sigma^2\right)+\sqrt{c}\gamma(0)\left(\frac{\sigma^2}{\gamma(0)}-\frac{\tau_{2c}}{c}\right).$$

Combining these inequalities and subtracting $\left(s_{\tau_{2c}-1}^2-\tau_{2c}\gamma(0)\right)/\sqrt{c}$ to each side, we get

$$\begin{aligned}&\frac{1}{\sqrt{c}}\sum_{n=1}^{\tau_{2c}-1}\left(x_{n-1}^2-\gamma(0)\right)-\sqrt{c}\left(s_{\tau_{2c}-1}^2-\sigma^2\right)\\&<\sqrt{c}\gamma(0)\left(\frac{\sigma^2}{\gamma(0)}-\frac{\tau_{2c}}{c}\right)+\frac{\gamma(0)}{\sqrt{c}}\\&\leq\frac{1}{\sqrt{c}}\sum_{n=1}^{\tau_{2c}}\left(x_{n-1}^2-\gamma(0)\right)+\frac{\gamma(0)}{\sqrt{c}}-\sqrt{c}\left(s_{\tau_{2c}}^2-\sigma^2\right).\end{aligned}$$

Hence,

$$\sqrt{c}\left(\frac{\tau_{2c}}{c}-\frac{\sigma^2}{\gamma(0)}\right)\Rightarrow-\frac{\nu}{\gamma\mathcal{F}_n(0)}W^{(3)}\left(\frac{\sigma^2}{\gamma(0)}\right)+\frac{\omega}{\sigma^2}W^{(2)}\left(\frac{\sigma^2}{\gamma(0)}\right)$$
$$=-\frac{2\beta\gamma(0)^{1/2}}{\sigma}W^{(1)}\left(\frac{\sigma^2}{\gamma(0)}\right).$$

□

Often it is regarded as desirable to set the maximum sample size, say M, and truncate the stopping time at M. We have the following results.

Corollary 8. *Let $M=\lfloor cm\rfloor$ be the maximum sample size and $\tau'_{ic}=\tau_{ic}\wedge M$ $(i=1,2)$, as $c\to\infty$,we have*

$$\lim_{c\to\infty}\frac{\tau'_{ic}}{c}=\frac{\tau_{ic}}{c}\wedge m\to\frac{\sigma^2}{\gamma(0)}\wedge m\quad a.s.(i=1,2)$$

and

$$\sqrt{c}\left(\hat{\beta}_{\tau'_{ic}}-\beta\right)_{i=1,2}\Rightarrow\frac{\sqrt{\gamma(0)}\sigma W^{(1)}\left(\frac{\sigma^2}{\gamma(0)}\wedge m\right)}{\left(\frac{\sigma^2}{\gamma(0)}\wedge m\right)\gamma(0)}\sim\begin{cases}N(0,1) & \frac{\sigma^2}{\gamma(0)}\le m\\ N\left(0,\frac{\sigma^2/\gamma(0)}{m}\right) & \frac{\sigma^2}{\gamma(0)}>m.\end{cases}$$

The proof can be provided in the same way as that of (33).

Remark 9. In theorem 7, we derive the asymptotic properties of sequential LSE $\hat{\beta}_{\tau_{ic}}$ and stopping time τ_{ic} $(i=1,2)$. For a finite c, the initial value x_0 does not affect the properties of sequential LSE, while the expectation and variance of τ_{1c} and τ_{2c} decreases as follows, $E(\tau_{ic}/c)\approx\frac{\sigma^2}{\gamma(0)}-\frac{x_0^2}{c\sigma^2}$ $(i=1,2)$, $Var(\tau_{1c})\approx\frac{\nu^2\sigma^2}{\gamma(0)^3}c-\frac{4\beta^2\gamma(0)^2}{\sigma^4}-\frac{\omega^2x_0^2}{\sigma^6}$, and $Var(\tau_{2c})\approx Var(\tau_{1c})-\frac{c\omega^2}{\gamma(0)^2}\left(\frac{2}{\sigma^2/\gamma(0)}-\frac{1}{\sigma^2/\gamma(0)-x_0^2/c\sigma^2}\right)$. Using the same argument which derives (34), (35) and (30), we have

$$\sqrt{c}\gamma(0)\left(\frac{\sigma^2}{\gamma(0)}-\frac{x_0^2}{c\sigma^2}-\frac{\tau_{2c}}{c}\right)$$
$$=\frac{2\beta}{1-\beta^2}\frac{1}{\sqrt{c}}\sum_{n=1}^{\tau_{2c}}x_{n-1}\epsilon_n+\left(\frac{\gamma(0)}{\sigma^2}-\frac{c}{\tau_{2c}}\right)\frac{1}{\sqrt{c}}\sum_{n=1}^{\tau_{2c}}\left(\epsilon_n^2-\sigma^2\right)+o_p(1).$$

If we ignore the effect of initial value x_0 to the functional C.L.T., we consider a rough approximation

$$\sqrt{c}\gamma(0)\left(\frac{\sigma^2}{\gamma(0)}-\frac{x_0^2}{c\sigma^2}-\frac{\tau_{2c}}{c}\right)$$
$$\sim\frac{2\beta}{1-\beta^2}\sqrt{\gamma(0)}\sigma W^{(1)}\left(\frac{\sigma^2}{\gamma(0)}-\frac{x_0^2}{c\sigma^2}\right)+\left(\frac{\gamma(0)}{\sigma^2}-\frac{1}{\frac{\sigma^2}{\gamma(0)}-\frac{x_0^2}{c\sigma^2}}\right)\omega W^{(2)}\left(\frac{\sigma^2}{\gamma(0)}-\frac{x_0^2}{c\sigma^2}\right).$$

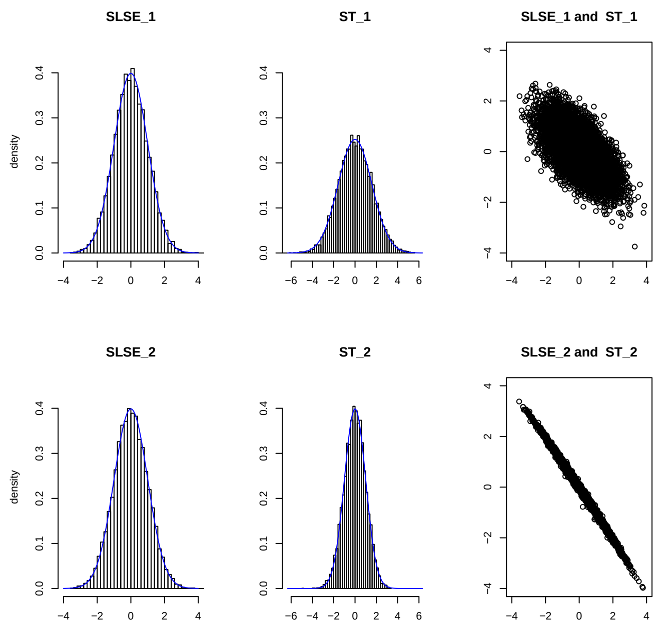
3 Simulation study

We conducted a simulation study to examine the joint asymptotic properties of stopping times and sequential estimators for both stationary AR(1) processes.

For a stationary AR(1) process, we prove the joint asymptotic normality of the stopping times and sequential estimators in Lemma 6 and Theorem 7. The simulation parameters were $\beta=0.5$, $\epsilon_n\sim_{i.i.d.}N(0,1)$, $c=2000$, and replication number of 10,000. Put $SLSE_i=\sqrt{c}\left(\hat{\beta}_{\tau_{ic}}-\beta\right), ST_i=\sqrt{c}\left(\tau_{ic}/c-\sigma^2/\gamma(0)\right)\quad(i=1,2)$.

Figure 1 presents the simulation results. The first and second columns show the simulated histograms of $\hat{\beta}_{\tau_{1c}}$ and $\hat{\beta}_{\tau_{2c}}$ and stopping times τ_{1c} and τ_{2c} after centralization, and the curves represent the densities

Figure 1: Simulation results for AR(1) model ($c = 5000$)



of the corresponding normal distributions. We can see that both sequential estimators $\hat{\beta}_{\tau_{1c}}$ and $\hat{\beta}_{\tau_{2c}}$ and stopping times τ_{1c} and τ_{2c} are well approximated by normal distributions and the variance of τ_{1c} is larger than that of τ_{2c}. The third column is the scatter plot of $\left(\hat{\beta}_{\tau_{1c}}, \tau_{1c}\right)$ and $\left(\hat{\beta}_{\tau_{2c}}, \tau_{2c}\right)$. The figure indicates that as c increases for $\left(\hat{\beta}_{\tau_{2c}}, \tau_{2c}\right)$, the joint distribution degenerates. There exists a nearly perfect negative correlation for $\left(\hat{\beta}_{\tau_{2c}}, \tau_{2c}\right)$.

4 Conclusion

We investigate the joint asymptotic properties of stopping times and sequential least-squares estimator of the autoregressive parameters for stationary AR(1) processes.

Anscombe Donsker-Skorohod approach (Theorem 3 and Theorem 4) in $D[0, \infty)$ play a crucial role to analyze the asymptotic properties of linear or nonlinear processes in sequential analysis of an AR(1) process. Using these theorems, for a stationary AR(1) process, we prove the joint asymptotic normality of the stopping time and sequential estimators. We find that when the disturbance variance σ^2 is estimated, the joint distribution degenerates and the asymptotic variance of the stopping time is strictly smaller than that of the stopping time with the known σ^2. A similar investigation can be extended to p-th order autoregressive process (AR(p)).

We present the ideas and the theoretical results here. The application should be also considered. For example, combing the results of companion paper (K.Nagai, Y. Nishiyama, and K. Hitomi (2018)), the sequential detection for the order d of Integrated AR(p) process is developed and examined by simulation.

Based on our method, sequential analysis and statistical process monitoring can be developed for linear and nonlinear time series, such as ARMA (autoregressive moving average), ARCH (autoregressive conditional heteroscedasticity), and GARCH (generalized autoregressive conditional heteroscedasticity) models. Research on some of these approaches is currently under way.

5 Appendix

5.1 The Geometry of the Space $D[0,\infty)$

In this subsection we briefly describe some characteristics of the space $D[0,\infty)$. Essentially, we follow the argument in Billingsley (1999), but we do not seek mathematically perfect explanations. According to Billingsley (1999), $D[0,\infty)$ becomes a Polish space, i.e. a complete separable metric space with a suitable metric.

Let Λ_m denote the class of strictly increasing, continuous mappings of $[0,m]$ onto itself. If $\lambda \in \Lambda_m$, then $\lambda 0 = 0$ and $\lambda m = m$. Let

$$\|\lambda\|_m^\circ = \sup_{0\le s<t\le m}\left|\log\frac{\lambda t-\lambda s}{t-s}\right| \text{ and } d_m^\circ(x,y) = \inf_{\lambda\in\Lambda_m}\left\{\|\lambda\|_m^\circ \vee \|x-y\lambda\|_m\right\}.$$

and define

$$h_m(t) = \begin{cases} 1 & t\le m-1, \\ m-t & m-1\le t\le m, \\ 0 & t\ge m. \end{cases}$$

For $f\in D[0,\infty)$, let f^m be the element of $D[0,\infty)$ defined by$f^m(t) = h_m(t)f(t), \quad t\ge 0$. Now we define the metric on $D[0,\infty)$ for $f,g\in D[0,\infty)$; and stopping times. We call the above procedure the Anscombe-Donsker-Skorohod

$$d_\infty^\circ(f,g) = \sum_{m=1}^\infty 2^{-m}\left(1\wedge d_m^\circ(f^m,g^m)\right). \tag{36}$$

One can recognize $d_\infty^\circ(f,g)\le \rho(f,g)$. Let Λ_∞ be the set of strictly increasing, continuous maps of $[0,\infty)$ onto itself.

Theorem 10. *(Billingsley (1999) p.168 , Th16.1). Let $I:[0,\infty)\to[0,\infty)$ be the identity map; $I(t)=t$. Then, $d_\infty^\circ(f_n,f)\to 0$ in $D[0,\infty)$ as $n\to\infty$ if and only if there exist elements λ_n of Λ_∞ such that*

$$\lim_{n\to\infty}\|\lambda_n - I\|_\infty = 0 \tag{37}$$

and for each m

$$\lim_{n\to\infty}\|f_n\circ\lambda_n - f\|_m \to 0. \tag{38}$$

Based on this theorem, we provide the proof of Lemma 2.

Proof. Suppose (37) and (38) hold for $f_n\in D[0,\infty)$, $f\in C[0,\infty)$, and $\lambda_n\in\Lambda_\infty$. Fix $m\in\mathbb{N}$ and $\epsilon\in(0,m)$. Since f is uniformly continuous on any compact set, there exists $0<\delta<m$ such that $|f(s)-f(t)|<\epsilon/2$ for any $s,t\in[0,2m]$ satisfying $|s-t|<\delta$. Choose n_0 such that $\|\lambda_n - I\|_\infty<\delta$ and $\|f_n\circ\lambda_n - f\|_{2m}<\epsilon/2$ for any $n\ge n_0$. For any $t\in[0,m]$, let $s=\lambda_n^{-1}t$. Then, we can obtain $\left|t-\lambda_n^{-1}t\right| = \left|\lambda_n s - \lambda_n^{-1}\lambda_n s\right| \le \|\lambda_n - I\|_\infty < \delta$. Hence, $\lambda_n^{-1}t\in[0,2m]$ and

$$\begin{aligned} \|f_n - f\|_m &= \sup_{t\le m}\left|f_n(\lambda_n\lambda_n^{-1}t) - f(\lambda_n^{-1}t) + f(\lambda_n^{-1}t) - f(t)\right| \\ &\le \|f_n\circ\lambda_n - f\|_{2m} + \sup_{t\le m}\left|f(\lambda_n^{-1}t)-f(t)\right| \\ &\le \frac{\epsilon}{2}+\frac{\epsilon}{2}. \end{aligned}$$

The necessity follows by letting $\lambda_n = I$ for each n. □

Lemma 11. *Suppose that for some random sequence w_n $n=1,2,\ldots$ and $\theta\in\mathbb{R}$*

$$\frac{1}{N}\sum_{n=1}^N w_n \to \theta \text{ a.s.} \quad (\textit{resp. in probability})$$

Then, as $\kappa \uparrow \infty$, for any $m > 0$,

$$\sup_{t \le m} \left| \frac{1}{\kappa} \sum_{n=1}^{\lfloor \kappa t \rfloor} w_n - \theta t \right| \to 0 \ a.s. \quad (resp.\ in\ probability)\,.$$

Proof. Fix $\omega \in \left\{ \sum_{n=1}^{N} w_n/N \to \theta \right\}$, $\forall m > 0$ and $\forall \varepsilon > 0$ find N_0 so that for any $N > N_0$,

$$\left| \frac{1}{N} \sum_{n=1}^{N} w_n - \theta \right| < \frac{\varepsilon}{2m}.$$

Then, for sufficiently large $\kappa > 0$,

$$\begin{aligned}
&\sup_{t \le m} \left| \frac{1}{\kappa} \sum_{n=1}^{\lfloor \kappa t \rfloor} w_n - \theta t \right| \\
&= \sup_{t \le m} \left| \frac{1}{\kappa} \sum_{n=1}^{\lfloor \kappa t \rfloor} (w_n - \theta) + \frac{(\lfloor \kappa t \rfloor - \kappa t)\theta}{\kappa} \right| \\
&\le \max_{N \le N_0 + 1} \left| \frac{1}{\kappa} \sum_{n=1}^{N} (w_n - \theta) \right| \vee \sup_{N_0 + 1 \le \kappa t \le \kappa m} \left| \frac{t}{\kappa t} \sum_{n=1}^{\lfloor \kappa t \rfloor} (w_n - \theta) \right| + \frac{|\theta|}{\kappa} \\
&\le \frac{1}{4}\varepsilon + m\frac{\varepsilon}{2m} + \frac{1}{4}\varepsilon = \varepsilon. \qquad (39)
\end{aligned}$$

□

To show the central limit theorem (Theorem 7) of the sequential estimator and the stopping times τ_{1c} and τ_{2c}, we apply the following theorem shown in Billingsley(1999). Denote $\|\cdot\|$ the L^2 norm and $\mathbb{Z}_- = \{\ldots, -2, -1, 0\}$.

5.2 Proof of Theorem 4

Proof. Observe that ξ_n are stationary and ergodic. First, by Lapounov's inequality, we have

$$|E[\xi_0]| = |E[\xi_n]| = |E\left[E[\xi_n|\mathcal{F}_0]\right]| \le \|E[\xi_n|\mathcal{F}_0]\| \to 0.$$

which implies $E[\xi_n] = 0$. Next, we show (18) is absolutely summable. From Schwarz's inequality, we have

$$|E\left[\xi_0 \xi_n\right]| = |E\left[\xi_0 E\left[\xi_n \,|\mathcal{F}_0\right]\right]| \le \|\xi_0\| \cdot \|E\left[\xi_n \,|\mathcal{F}_0\right]\|\,.$$

From (17) and $\|\xi_0\| < \infty$, (18) is absolutely summable.

Letting $\gamma_h = E\left[\xi_0 \xi_h\right]$, by stationarity, for S_n defined in Theorem 4, the dominated convergence theorem gives

$$\begin{aligned}
\frac{1}{n} E\left[S_n^2\right] &= \frac{1}{n} \left(\sum_{k=1}^{n} E\left[\xi_k^2\right] + 2 \sum_{k<l}^{n} E\left[\xi_k \xi_l\right] \right) \\
&= \gamma_0 + 2 \sum_{h=1}^{n-1} (\frac{n-h}{n}) \gamma_h \\
&\to \gamma_0 + 2 \sum_{h=1}^{\infty} \gamma_h = \nu^2.
\end{aligned}$$

Next, we will check By Lapounov's inequality again, together with (17),

$$E\left[\sum_{i=1}^{\infty}\left|E\left[\xi_i\left|\mathcal{F}_0\right.\right]\right|\right]\leq\sum_{i=1}^{\infty}\left\|E\left[\xi_i\left|\mathcal{F}_0\right.\right]\right\|<\infty. \tag{40}$$

Let $\pi_n:\Omega\to\mathbb{R}$ be the coordinate map $\pi_n(\omega)=\omega_n$ and $T:\Omega\to\Omega$ be the left shift operator $\pi_n(T\omega)=\omega_{n+1}$. Then T is measure-preserving and ergodic by the assumption. For a $\mathcal{F}$-measurable L^1-random variable θ and a σ-field $\mathcal{G}\subset\mathcal{F}$, we have

$$E\left[\theta T\left|T^{-1}\mathcal{G}\right.\right](\omega)=E\left[\theta\left|\mathcal{G}\right.\right](T\omega)\quad a.s. \tag{41}$$

Define

$$\theta_k(\omega)=\sum_{i=1}^{\infty}E\left[\xi_{k+i}\left|\mathcal{F}_k\right.\right](\omega).$$

Then, from equation (41) and $\mathcal{F}_k=T^{-k}\mathcal{F}_0$, $\theta_k(\omega)=\sum_{i=1}^{\infty}E\left[\xi_i\left|\mathcal{F}_0\right.\right]\left(T^k\omega\right)=\theta_0\left(T^k\omega\right)$. Hence, θ_k are stationary and ergodic. For η_k in (20), it is obvious that η_k are ergodic, stationary process. Since, by the tower property, $E\left[\theta_k-\theta_{k-1}\left|\mathcal{F}_{k-1}\right.\right]=\sum_{i=1}^{\infty}E\left[\xi_{k+i}\left|\mathcal{F}_{k-1}\right.\right]-\sum_{i=1}^{\infty}E\left[\xi_{k-1+i}\left|\mathcal{F}_{k-1}\right.\right]-E\left[\xi_k\left|\mathcal{F}_{k-1}\right.\right]$, therefore $E\left[\eta_k\left|\mathcal{F}_{k-1}\right.\right]=0$; η_k is a martingale difference. To apply the results of C.L.T. for ergodic martingale differences (Theorem 18.3 in Billingsley (1999)), we show η_n in L^2. Let

$$\theta_0=\sum_{i=1}^{\infty}E\left[\xi_i\left|\mathcal{F}_0\right.\right].$$

By (17), we have $E[\theta_0^2]\leq\sum_{i,j=1}^{\infty}E\left[\left|E[\xi_i\left|\mathcal{F}_0\right.]\right|\cdot\left|E[\xi_j\left|\mathcal{F}_0\right.]\right|\right]\leq\left(\sum_{i=1}^{\infty}\left\|E\left[\xi_i\left|\mathcal{F}_0\right.\right]\right\|\right)^2<\infty$. Since $\|\theta_n\|=\|\theta_0\|<\infty$ and $\|\xi_n\|<\infty$, thus, η_n have finite second moments. For M_n defined in in Theorem 4, let $E\left[\eta_i^2\right]=\tau^2$. Using C.L.T. for ergodic martingale differences (Theorem 18.3 in Billingsley (1999)), we have

$$\frac{M_{\lfloor ct\rfloor}}{\sqrt{c}\tau}\Rightarrow W_t$$

in the sense of $D[0,\infty)$. We can write M_n as

$$M_n=\sum_{i=1}^{n}\left(\xi_i+\theta_i-\theta_{i-1}\right)=S_n+\theta_n-\theta_0, \tag{42}$$

since$\|\theta_n\|=\|\theta_0\|<\infty$, as $n\to\infty$,

$$\frac{1}{\sqrt{n}}\left\|M_n-S_n\right\|=\frac{1}{\sqrt{n}}\left\|\theta_n-\theta_0\right\|\leq\frac{2}{\sqrt{n}}\left\|\theta_0\right\|\to 0.$$

Thus, we obtain

$$\frac{S_n}{\sqrt{n}\tau}\Rightarrow N.$$

We have

$$\begin{aligned}\left\|\frac{M_n}{\sqrt{n}}\right\|^2&=E\left(\frac{\eta_1+\cdots+\eta_n}{\sqrt{n}}\right)^2\\&=\frac{1}{n}\left(nE\left[\eta_1^2\right]+2\sum_{k<l}^{n}E\left[\eta_k\eta_l\right]\right)\\&=\tau^2,\end{aligned}$$

since $E\left[\eta_k\eta_l\right]=E\left[E\left[\eta_k\eta_l\left|\mathcal{F}_{l-1}\right.\right]\right]=E\left[\eta_kE\left[\eta_l\left|\mathcal{F}_{l-1}\right.\right]\right]=0$. From this equality, $E\left[S_n^2\right]/n\to\nu^2$.and

$$\left|\left\|\frac{M_n}{\sqrt{n}}\right\|-\left\|\frac{S_n}{\sqrt{n}}\right\|\right|\leq\left\|\frac{M_n-S_n}{\sqrt{n}}\right\|\to 0,$$

we can obtain $\tau = \nu$. Therefore, we have

$$\frac{S_n}{\sqrt{n}\nu} \Rightarrow N(0,1).$$

Now, define

$$X_t^c = \frac{S\lfloor ct \rfloor}{\nu\sqrt{c}},\ Y_t^c = \frac{M\lfloor ct \rfloor}{\nu\sqrt{c}},$$

where $Y_t^c = M\lfloor ct \rfloor / \nu\sqrt{c} \Rightarrow W_t$. For each $m > 0$, as $n \to \infty$

$$\begin{aligned} \|X^c - Y^c\|_m &= \sup_{t \le m} |X_t^c - Y_t^c| \\ &= \frac{1}{\sqrt{c}\nu} \sup_{t \le m} \left|\theta_{\lfloor ct \rfloor} - \theta_0\right| \\ &= \frac{1}{\sqrt{c}\nu} \max_{k \le cm} |\theta_k - \theta_0| \end{aligned}$$

Now, we will show $\max_{k \le cm} |\theta_k - \theta_0| / \sqrt{c}\nu \to 0 \quad a.s.$ as $c \uparrow \infty$ Since $E\left[\theta_0^2\right] < \infty$, we have

$$\sum_{n=1}^{\infty} P\left\{\frac{\theta_n^2}{n} \ge \epsilon\right\} = \sum_{n=1}^{\infty} P\left\{\frac{\theta_0^2}{n} \ge \epsilon\right\} \le E\left[\frac{\theta_0^2}{\epsilon}\right] < \infty.$$

The first inequality is from $\sum_{n=1}^{\infty} P\{X \ge n\} \le E[X]$ for a nonnegative r.v. X. By the Borel–Cantelli lemma, we have $\theta_n^2/n \to 0 \quad a.s.$ For fixed $m > 0$ and $\epsilon > 0$, choose n_0 such that $\theta_k^2/k < \epsilon/m$ for $k > n_0$.Then, we have, for sufficiently large c,

$$\frac{1}{c} \max_{k \le cm} \theta_k^2 \le \max_{k \le n_0} \frac{\theta_k^2}{c} \bigvee \max_{n_0 < k \le cm} \frac{\theta_k^2}{c} \le \epsilon,$$

since when $n_0 < k \le cm$,

$$\frac{\theta_k^2}{c} = \frac{k}{c}\frac{\theta_k^2}{k} \le \frac{cm}{c}\frac{\theta_k^2}{k} = m\frac{\theta_k^2}{k} \le \epsilon.$$

Hence,

$$\frac{1}{\sqrt{c}\nu} \max_{k \le cm} |\theta_k - \theta_0| \to 0 \quad a.s.$$

$Therefore, \|X^c - Y^c\|_m \to 0 \quad a.s.$ and $d_\infty^\circ(X^c, Y^c) \le \rho(X^c, Y^c) \to 0$. By Slutzky's theorem, we can obtain $X^c \Rightarrow W$. □

5.3 Some preliminary asymptotics for stationary AR(1) process

We now consider another probability space $(\Omega', \mathcal{F}', P')$ on which $\epsilon_0, \epsilon_{-1}, \epsilon_{-2}\dots$ are independent and have the same distribution as ϵ_1. Let $(\Omega'', \mathcal{F}'', P'')$ be the completion of the product space $(\Omega \times \Omega', \mathcal{F} \times \mathcal{F}', P \times P')$. Let $\tilde{x}_n$ be the ergodic, stationary AR(1) process in $(\Omega'', \mathcal{F}'', P'')$;

$$\tilde{x}_n = \sum_{k=0}^{\infty} \beta^k \epsilon_{n-k} \tag{43}$$

for $n = 1, 2, \dots$. Then, the following relation holds between x_n and $\tilde{x}_n$.

$$x_n = \tilde{x}_n + \beta^n (x_0 - \tilde{x}_0). \tag{44}$$

Lemma 12. . *Let x_n be an AR(1) process defined in (1) with the initial value x_0 independent of $\epsilon_1, \epsilon_2, \dots$. Then, as $N \to \infty$,*

$$\frac{1}{N}\sum_{n=1}^{N} x_{n-1}\epsilon_n \to 0 \quad a.s. \quad \text{and} \quad \frac{1}{N}\sum_{n=1}^{N} x_{n-1}^2 \to \gamma(0) \quad a.s. \tag{45}$$

where $\gamma(m)$ is defined in (7).Furthermore, assuming $x_0 \in L^2(\Omega, \mathcal{F}, P)$, then as $c \to \infty$

$$\frac{1}{\sqrt{c}} \sum_{n=1}^{\lfloor ct \rfloor} x_{n-1}\epsilon_n - \frac{1}{\sqrt{c}} \sum_{n=1}^{\lfloor ct \rfloor} \tilde{x}_{n-1}\epsilon_n \to_P \mathbf{0} \tag{46}$$

$$\frac{1}{\sqrt{c}} \sum_{n=1}^{\lfloor ct \rfloor} x_{n-1}^2 - \frac{1}{\sqrt{c}} \sum_{n=1}^{\lfloor ct \rfloor} \tilde{x}_{n-1}^2 \to_P \mathbf{0} \tag{47}$$

where $\to_P$ indicates convergence in probability in $D[0,\infty)$ and $\mathbf{0}$ represents the zero function.

Proof. There is no harm in using $(\Omega'', \mathcal{F}'', P'')$ instead of $(\Omega, \mathcal{F}, P)$. See

$$P\left(\lim_{N\to\infty} \sum_{n=1}^{N} x_{n-1}\epsilon_n / N = 0\right) = P''\left(\lim_{N\to\infty} \sum_{n=1}^{N} x_{n-1}\epsilon_n / N = 0\right).$$

By the ergodic theorem (See Shiryaev (1984)) and Schwartz's inequality,

$$\frac{1}{N}\sum_{n=1}^{N} x_{n-1}\epsilon_n = \frac{1}{N}\sum_{n=1}^{N} \left(\tilde{x}_{n-1} + \beta^{n-1}(x_0 - \tilde{x}_0)\right)\epsilon_n \to 0 \quad a.s.P''$$

and

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N}\left(x_{n-1}^2 - \tilde{x}_{n-1}^2\right) &= \frac{2}{N}\sum_{n=1}^{N}\beta^{n-1}\tilde{x}_{n-1}(x_0 - \tilde{x}_0) + \frac{1}{N}\sum_{n=1}^{N}\beta^{2n-2}(x_0 - \tilde{x}_0)^2 \\ &\to 0 \quad a.s.P''.\end{aligned}$$

Define the sup norms $\|\cdot\|_m$ for $m > 0$ as in (16), the metric ρ for uniform convergence of compacta in (15), and the Skorohod metric d_∞° as in (36). Since x_n and $\tilde{x}_n$ have the relation (44),

$$\frac{1}{\sqrt{c}} \sum_{n=1}^{\lfloor ct \rfloor} x_{n-1}\epsilon_n - \frac{1}{\sqrt{c}} \sum_{n=1}^{\lfloor ct \rfloor} \tilde{x}_{n-1}\epsilon_n = \frac{1}{\sqrt{c}}(x_0 - \tilde{x}_0)\sum_{n=1}^{\lfloor ct \rfloor}\beta^{n-1}\epsilon_n. \tag{48}$$

Let

$$\Delta_1^c(t) = \frac{1}{\sqrt{c}}\sum_{n=1}^{\lfloor ct \rfloor}\beta^{n-1}\epsilon_n.$$

To obtain $d_\infty^\circ(\Delta_1^c, \mathbf{0}) \to_P 0$, it suffices to show that $\rho(\Delta_1^c, \mathbf{0}) \to_P 0$ by Lemma 2. Hence, we show $\|\Delta_1^c\|_m \to_P 0$ for each $m > 0$. In fact, for any $\epsilon > 0$, Kolmogorov's inequality implies

$$P\left\{\max_{t\le m}\left|\sum_{n=1}^{\lfloor ct \rfloor}\beta^{n-1}\epsilon_n\right| \ge \sqrt{c}\epsilon\right\} \le \frac{E\left[\left(\sum_{n=1}^{\lfloor cm \rfloor}\beta^{n-1}\epsilon_n\right)^2\right]}{c\epsilon^2} \to 0 \quad (c\to\infty).$$

Since $x_{\lfloor ct \rfloor}/c^{1/4} = \sum \beta^k \epsilon_{\lfloor ct \rfloor - k}/c^{1/4} + \beta^{\lfloor ct \rfloor}x_0/c^{1/4}$, we have

$$P\left\{\max_{t\le m}\left|x_{\lfloor ct \rfloor}/c^{1/4}\right| \ge \epsilon\right\} \to 0 \quad (c\to\infty).$$

From the equation (30),

$$\sum_{n=1}^{\lfloor ct \rfloor}\left(x_{n-1}^2 - \gamma(0)\right) = \frac{1}{(1-\beta^2)}\left(x_0^2 - x_{\lfloor ct \rfloor}^2 + 2\beta\sum_{n=1}^{\lfloor ct \rfloor}x_{n-1}\epsilon_n + \sum_{n=1}^{\lfloor ct \rfloor}\left(\epsilon_n^2 - \sigma^2\right)\right)$$

Therefore,

$$\begin{aligned}&\frac{1}{\sqrt{c}}\sum_{n=1}^{\lfloor ct\rfloor}x_{n-1}^2-\frac{1}{\sqrt{c}}\sum_{n=1}^{\lfloor ct\rfloor}\tilde{x}_{n-1}^2\\ =&\frac{1}{\sqrt{c}(1-\beta^2)}\left(x_0^2-\tilde{x}_0^2-\left(x_{\lfloor ct\rfloor}^2-\tilde{x}_{\lfloor ct\rfloor}^2\right)+2\beta\sum_{n=1}^{\lfloor ct\rfloor}\left(x_{n-1}-\tilde{x}_{n-1}\right)\epsilon_n\right)\to_P \mathbf{0}.\end{aligned}$$

□

References

[1] Anscombe, J. (1952). Large sample theory of sequential estimation. *Proc. Camb. Phil. Soc.* **48**(4) 600-607.

[2] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd edition, Wiley, New York.

[3] Galtchouk, L. and Konev, V. (2003). On uniform asymptotic normality of sequential least squares estimators for the parameters in a stable AR(1). *Sequential Anal.* **22** 31-54.

[4] Galtchouk, L. and Konev, V. (2004). On uniform asymptotic normality of sequential least squares estimators for the parameters in a stable AR(p). *J. Multivariate Anal.* **91** 119-142.

[5] Galtchouk, L. and Konev, V. (2005). On Sequential Least Squares Estimates of Autoregressive Parameters. *Sequential Anal.* **24** 335-364.

[6] Galtchouk, L. and Konev, V. (2006). Sequential estimation of the parameters in unstable AR(2). *Sequential Anal.* **25** 25-43.

[7] Galtchouk, L. and Konev, V. (2011). On asymptotic normality of sequential LS-estimates of unstable autoregressive processes. *Sequential Anal.* **30** 117-144 .

[8] Konev, V. and Pergamenshchikov, S. M. (1986). On the Duration of Sequential Estimation of Parameters of Stochastic Processes in Discrete Time. *Stochastics*, **18** 133-154

[9] Lai, T. L. and Siegmund, D. (1983). Fixed accuracy estimation of an autoregressive parameter. *Ann. Statist.* **11** 478-485.

[10] Lee, S. and Sriram T. N. (1999). Sequential point estimation of parameters in a threshold AR(1) model. *Stochastic Process. Appl.* **84** 343-355

[11] Nagai, K., Nishiyama, Y. and Hitomi, K. (2018). Sequential test for unit root in AR(1) model, *Kyoto Institute of Economic Research Discussion Paper*, No. 1003, Kyoto University

[12] Pergamenshchikov, S. M. (1991). Asymptotic properties of sequential design for estimating the parameter of a first-order autoregression, *Theory Probab. Appl.* **36** 36-49.

[13] Shiryaev, A. N. (1984). *Probability,* Springer-Verlag New York

[14] Shete, S. and Sriram, T. N. (1998). Fixed precision estimator of the offspring mean in branching processes. *Stochastic Process. Appl.* **77** 17-33

[15] Sriram, T. N. (1987). Sequential estimation of the mean of a first order stationary autoregressive process. *Ann.Statist.* **15** 1079-1090.

[16] Sriram, T. N. (1988). Sequential estimation of the autoregressive parameter in a first order autoregressive process. *Sequential Anal.* **7** 53-74.

[17] Sriram,T. N., Basawa. I. V. and Huggins. R. M. (1991). Fixed precision estimator of the offspring mean in branching processes. *Stochastic Process. Appl.* **77** 17-33

[18] Sriram, T. N., and Iaci, R. (2014). Sequential Estimation for Time Series Models. *Sequential Anal.* **33** 136–157

[19] Woodroofe, M. (1982). Nonlinear Renewal Theory in Sequential Analysis. *Society for Industrial and Applied Mathematics*, Philadelphia.