KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE OF ECONOMIC RESEARCH

Discussion Paper No.1003

"Sequential test for unit root in AR(1) model"

K. Nagai, Y. Nishiyama, AND K. Hitomi

October 2018

KYOTO UNIVERSITY

KYOTO, JAPAN

Sequential test for unit root in AR(1) model

By K. Nagai, Y. Nishiyama, AND K. Hitomi
Yokohama National University, Kyoto University and Kyoto Institute of Technology

October 10, 2018

Abstract

We consider unit root tests under sequential sampling for an AR(1) process against both stationary and explosive alternatives. We propose three kinds of test, or t type, stopping time and Bonferroni tests, using the sequential coefficient estimator and the stopping time of Lai and Siegmund (1983). To examine the statistical properties, we obtain their weak joint limit by approximating the processes in $D[0,\infty)$ and using time change and a DDS (Dambis and Dubins-Schwarz) Brownian motion. The distribution of the stopping time is characterized by a Bessel process of dimension 3/2 with and without drift, while the esitimator is asymptotically normally distributed. We implement Monte Carlo simulations and numerical computations to examine their small sample properties.

1 Introduction

Consider a scalar first order autoregressive process (AR(1) herafter),

$$x_n = \beta x_{n-1} + \epsilon_n, \tag{1}$$

where ϵ_n's are independently and identically distributed random variables with mean zero and finite variance σ^2. We write it as $\epsilon_n \sim i.i.d.(0, \sigma^2)$. When $|\beta| = 1$ holds, $\{x_n\}$ is a unit root process. If $|\beta| < 1$, it is stationary, while we shall say that it is explosive if $|\beta| > 1$. Unit root processes are nonstationary that behave in a quite different manner from stationary series. The most notable feature of unit root processes is that they are not mean reverting unlike stationary processes. Many macroeconomic time series, for example, obviously do not look like explosive processes, but they also do not appear to be stationary. Therefore, it has been one of the main issues of interest in time series econometrics to examine if they are unit root processes. The case of $\beta = -1$ is certainly of some theoretical interest. In the unit root literature, however, researchers have focused only on the case of $\beta = 1$, because we can hardly find empirical data with $\beta = -1$.

Given a sample $\{x_n\}$, $n = 0, 1, \cdots, T$, the ordinary least squares (OLS) estimator $\hat{\beta}_T = \left(\sum_{n=1}^T x_{n-1}^2\right)^{-1} \sum_{n=1}^T x_{n-1}x_n$ is known to be consistent and have a non-standard asymptotic distribution when the $\beta = 1$, namely, as $T \to \infty$,

$$T(\hat{\beta}_T - 1) \Rightarrow \frac{\int_0^1 W_t dW_t}{\int_0^1 W_t^2 dt} = \frac{\frac{1}{2}(W_1^2 - 1)}{\int_0^1 W_t^2 dt}$$

where W_t is a standard Brownian motion and $\Rightarrow$ indicates the weak convergence. See White (1956) among others. A variety of procedures testing for the existence of unit root has been proposed since the middle of 1970's. The most widely used unit root tests are Dicky-Fuller test (Dickey and Fuller (1979)) and its extensions such as Chan and Wei (1988), Nabeya and Tanaka (1988), Phillips (1987a) and Phillips and Perron (1988). To examine the statistical properties and performance of the test under a near unit root process, many authors have considered local altenatives of $\beta = 1 - \delta/T$. See e.g. Bobkoski (1983), Cavanagh (1985), Chan (1988) ,Chan and Wei (1987), Phillips (1987b) ****** among others.

The present paper considers the same testing problem under a sequential sampling scheme. Sequential analysis was originally considered by Wald (1947). Its principle is that we collect observations such that we can ensure a predetermined accuracy in statistical decision making. This approach is particularly useful when sampling entails certain cost including opportunity cost. In a time series setting, suppose a fund manager would like to know whether a stock market is a bubble or not. She must want to know it as soon as possible to make decisions on her portfolio position. For simplicity, suppose she makes a decision based on, say, daily closing price. She collects an observation everyday and stop sampling when she has accumulated "sufficient" amount of information for a decision. The time when she stops sampling is called the stopping time. How "sufficient" depends on her preference to the decision accuracy and the cost of sampling. If she waits too long, she may possibly loose a lot of money. In general, the accuracy is typically measured by the standard error of the estimator or the power in testing. There exists a trade-off between the accuracy and the cost of sampling.

A number of papers examine the statistical properties of sequential OLS estimation of parameters for both stationary and nonstationary time series. Using a stopping time $\tau_c = \inf\left\{N > 1 : \sum_{n=1}^{N} x_{n-1}^2 \geq c\sigma^2\right\}$ for some predetemined constant $c > 0$, Lai and Siegmund (1983; LS83 hereafter) show that the sequential OLS estimator of (1) is asymptotically uniformly normal for $|\beta| \leq 1$ as $c \to \infty$. This sampling scheme collects observations until the observed Fisher information exceeds c, or equivalently the standard error of estimation becomes smaller than $1/c$. Shiryaev and Spokoiny (1997) prove the same result under the condition of normal disturbances when the absolute value of β is greater than unity. These results are extended to the case of AR(p) for $p \geq 1$ by ***. Dzhaparidze et.al. (1994) study the statisical properties of an sequential OLS estimator when the DGP is a near unit root process of $\beta = 1 - \delta/T$. They use a diffusion approximation on $D[0, 1]$ to study the asymptotic properties. ***other references?***

To the best of our knowledge, there exist only a few papers dealing with sequential testing for unit root. The unit root test using $T(\hat{\beta}_T - 1)$ is incovenient because the null distribution of $\hat{\beta}_T$ is non-standard as mentioned above. However, sequential OLS estimator is asymptotically normally distributed by LS83, and thus one can test the null of unit root comparing the the estimate with a quantile of normal distribution. Given a sample $\{x_0, \cdots, x_T\}$, Chang and Park (2004) and Chang (2012) take a similar approach to propose a unit root test using a part of observations, not all of them, such that the resulting estimator of β is asympototically normally distributed. They use a subsample of $\{x_0, \cdots, x_{m_T}\}$ to estimate β where $m_T = \inf\left\{k > 1 : T^{-2} \sum_{n=1}^{k} x_{n-1}^2 \geq c\right\}$ for an arbitrarily chosen constant c. This estimator is shown to be asymptotically normally distributed when $T \to \infty$. This is different from LS83 which considers asymptotic theory under $c \to \infty$. They also investigate its statistical properties under the local altenative considered in Dzhaparidze et.al. (1994) where it is shown that the test statistic is asymptotically normal with a mean shift and the proof uses a diffusion approximation of the process on $D[0, 1]$.

An important issue in sequential analysis is the operating characteristics such as the expectation and variance of stopping times. It is because researchers would like to know how much cost they need to pay for sampling in average. Many researchers have studied it by nonlinear renewal theory which plays a crucial role. See for example Woodroofe (1976) and Lai and Siegmund (1977, 1979). We also refer to the books by Woodroofe (1982) and Siegmund (1985). However, nonlinear renewal theory strongly depends on Wald's identity and thus works only for i.i.d. observations in principle. Without relying on nonlinear renewal theory, LS83 obtain the marginal stochastic limit of the stopping time using diffusion approximation of the process.

This paper proposes a sequential counterpart of Dicky-Fuller test possessing a standard normal limits. We also show that one can use the stopping time for the test. We derive the joint stochastic limit of the sequential OLS estimator by LS83 and the stopping time, where we approximate the unit root AR(1) time series with a diffusion process on $D[0, \infty)$ plays a crucial role. To prove the results, using the diffusion approximation on $D[0, \infty)$, we apply DDS (Dambis and Dubins-Schwarz hereafter) time change and Itô's Lemma. The joint limiting distribution is characterized by a DDS Brownian motion and a Bessel diffusion of dimension $3/2$ with driven by the same Brownian motion under the null of unit root. We consider the local alternative of $\beta = 1 - \delta/c$ when the process is approximated by an Ornstein-Uhlenbeck (OU) process, the asymptotic distribution of estimator is a Brownian motion subtracted by δ and the stopping time has a limit of Bessel process of the same dimension with a shift.

The following section explains the model and the testing hypotheses. Then we also define the stopping time and parameter estimation, and briefly discuss testing procedures based on them. We provide the joint asymptotic distribution of the estimator and the stopping time under the null in Section 3, while Section 4 gives the results under local alternatives. In Section 5, we report Monte

Carlo results to see the size and power of the tests. Section 5 concludes.

2 Sequential unit root test for AR(1) process

2.1 Model and hypotheses

Suppose a time series $\{x_n,\, n = 1, 2, \cdots\}$ is generated from the following AR(1) process with initial value x_0;

$$(1 - \beta L)x_n = \epsilon_n \tag{2}$$

where $\epsilon_n \sim i.i.d(0, \sigma^2)$. x_0 can be any value for now, but we shall discuss about ot later. We would like to test the null hypothesis of

$$H_0 : \beta = 1.$$

As the alternative hypothesis, we can consider two possibilities $|\beta| < 1$ and $\beta \neq 1$. The former is the stationary cases that most unit root tests consider in econometrics, while the latter includes the exclusive cases $|\beta| > 1$ as well as stationarity. Most literature do not look at the exclusive cases simply because such series are hardly found in practice. We also take this framwork in this paper. In view that many practical examples obtain parameter esimates close to unity, we also deal with the near unit root case. It is formally written as local alternatives,

$$\begin{aligned} H_0 &: \beta = 1 \\ H_1 &: \beta = 1 - \frac{\delta}{\sqrt{c}}, \end{aligned}$$

for $\delta > 0$ or $\delta \neq 0$ and some positive $c \to \infty$. In standard sampling theory, c is typically the sample size, then the H_1 is called Pitman local alternatives. In the present sequential framework, the sample size turns out to be random so that it is inappropriate. We shall explain about c in the following section. This local alternative setting is also useful to scrutinize the statistical properties of the test.

For the purpose of exposition, we consider the following null and alternative hypotheses for now.

$$\begin{aligned} H_0 &: \beta = 1 \\ H_1 &: |\beta| < 1, \end{aligned}$$

namely the null is a unit root process and the alternative is stationarity. A natural approach is to estimate β and compare it with 1. As in Dickey=Fuller unit root test, we estimate the following tranformed model,

$$\Delta x_n = \phi x_{n-1} + \epsilon$$

instead of directly estimating β, where $\phi \equiv \beta - 1$. Correspondingly, the testing hypotheses in terms of ϕ are

$$\begin{aligned} H_0 &: \phi = 0 \\ H_1 &: -2 < \phi < 0. \end{aligned}$$

2.2 Stopping time and sequential parameter estimation

We now explain how we stop sampling and test the hypothesis using the observations. Suppose we observe $x_0, x_1, x_2, \ldots$ sequentially and σ^2 is known for now. We propose to stop sampling at time

$$\tau_c = \inf\left\{ N > 1 : \frac{1}{\sigma^2} \sum_{n=1}^{N} x_{n-1}^2 \geq c \right\}, \tag{3}$$

for some predetermined $c > 0$. This is the same stopping time considered in LS83. If ϵ_n are normally distributed, the left side of the inequality in the wave brackets coincides with the observed Fisher information for ϕ. Therefore we can interprete that this stopping time guarantees the estimation accuracy c. Given a sample $x_0, x_1, x_2, \cdots .x_{\tau_c}$ in hand, we obtain its ordinary least squares (OLS hereafter) estimator $\hat{\phi}_{\tau_c}$ of ϕ, where

$$\hat{\phi}_N = \frac{\sum_{n=1}^{N} x_{n-1}\Delta x_n}{\sum_{n=1}^{N} x_{n-1}^2} \tag{4}$$

for $N \geq 1$. LS83 show that this estimator is asymptotically normally distributed as $c \to \infty$ uniformly on $[0, 2]$. It can be used to test if $\phi = 0$. Because σ^2 is unknown in practice, we need to estimate it for a feasible stopping time. Letting

$$s_N^2 = \frac{1}{N} \sum_{n=1}^{N} \left(\Delta x_n - \hat{\phi}_N x_{n-1} \right)^2, \tag{5}$$

and

$$\hat{\tau}_c = \inf \left\{ N > 1 : \frac{1}{s_N^2} \sum_{n=1}^{N} x_{n-1}^2 \geq c \right\}, \tag{6}$$

we obtain a feasible OLS estimator

$$\hat{\phi}_{\hat{\tau}_c} = \frac{\sum_{n=1}^{\hat{\tau}_c} x_{n-1}\Delta x_n}{\sum_{n=1}^{\hat{\tau}_c} x_{n-1}^2}. \tag{7}$$

3 Sequential unir root tests and asymptotic properties of the estimator and stopping time under H_0

We show the asymptotic properties of the estimator $\hat{\phi}_{\hat{\tau}_c}$ and stopping time $\hat{\tau}_c$ under the null of $\beta = 1$, namely

$$(1 - L)x_n = \epsilon_n \quad n = 1, 2, \ldots, \tag{8}$$

where $\epsilon_n \sim i.i.d.(0, \sigma^2)$ with $\sigma^2 \in (0, \infty)$. We first show in section 3.1 a result of diffusion approximation of the series for a unit root process. The proof is provided in Appendix 1. Section 3.2 states the main results, whose proofs are collected in Appendix 2.

3.1 Convergence to a Brownian motion on $D[0, \infty)$

Diffusion approximation on $D[0, 1]$, the space of right continuous functions on $[0, 1]$ with left limits, is a common approach to study the statistical properties of estimation and testing procedures associated with unit root processes. Some authors such as Chang (2012) and Chang and Park (2004) approximate the process on $D[0, 1]$, where they consider statistical procedures given a sample of size T but use a part of the data such that the information of observations used exceeds certain predetermined level. There, they obtain the asymptotic properties when $T \to \infty$. For the present purpose, however, we believe that it is more appropriate to approximate the process on the space of $D[0, \infty)$, the set of the right continuous functions on $[0, \infty)$ with left limits, to characterize the limiting behavior of the sequential estimator and the stopping time. We do not assume that we have a sample of size T in our hand and use a part of it, but we observe data online and stop sampling when we obtain sufficient information.

Approximation on $D[0, 1]$ is the most frequently used approach in discussing the asymptotic properties of unit root processes in standard sampling. However in a sequential sampling scheme, it is appropriate to consider the functional space $D[0, \infty)$ to characterize the limiting behavior since the sample size is determined by a stopping time whose value is arbitrary in $\mathbb{N} = \{1, 2, 3, \ldots\}$.

********* Do we need more discussion why we have to consider D[0,infty) instead of D[0,1]? ***************

Suppose $x_1, x_2, \dots$ are generated by the model (8) with an initial value x_0 independent of ϵ_n, $n \geq 1$. Let

$$X_c(t) = \frac{x_{\lfloor \sqrt{c}t \rfloor}}{c^{1/4}\sigma}, \tag{9}$$

with W being a standard Brownian motion. Then, Theorem 18.3 of Billingsley (1999) yields the following diffusion approximation in $D[0, \infty)$ under the null.

$$X_c \Rightarrow W, \tag{10}$$

as $c \uparrow \infty$ in the sense of $D[0, \infty)$ where $\Rightarrow$ indicates weak convergence.

This result plays an essential role to prove the asymptotic properties of the estimator and stopping time under the null.

3.2 Asymptotic properties of $\hat{\tau}_c$, $\hat{\phi}_{\hat{\tau}_c}$, and $s^2_{\hat{\tau}_c}$.

In this section we show the asymptotic properties of $\hat{\tau}_c$, $\hat{\phi}_{\hat{\tau}_c}$and $s^2_{\hat{\tau}_c}$ under the null of (8) when c goes to ∞. We use the diffusion approximation shown in (10). We omit the parentheses of $W(t)$ to write W_t for the brevity of expressions.

To state the main theorem of this section, we define a martingale M_t and its quadratic variation as

$$M_t = \int_0^t W_u dW_u, \tag{11}$$

$$\langle M \rangle_t = \int_0^t W_u^2 du \tag{12}$$

and let

$$U_s = \langle M \rangle_s^{-1} = \inf\left\{ t \geq 0 : \int_0^t W_u^2 du = s \right\}, \tag{13}$$

where W_u and X_u are respectively Brownian motion and Brownian motion with an initial value in (**??**). By DDS Theorem (Theorem 1.6 in Revuz and Yor (1999), pp181),

$$B_s = M_{U_s} \tag{14}$$

is a Brownian motion with respect to the filtration $\mathcal{G}_s = \mathcal{F}_{U_s}$ and this is called a DDS Brownian motion.

Remark 1. LS83 prove the same result but the proofs are substantially different. We use a diffusion approximation and the DDS theorem, which can be directly extended to the case of local alternatives as we show later, but it is not clear if their approach is applicable to the local alternative cases.

Theorem 2. *Suppose x_n is generated by the model (8) with an initial value x_0 independent of ϵ_n, $n \geq 1$. Then, if we put*

$$\rho_s = W_{U_s}^2/2,$$

ρ_t is a 3/2-dimensional Bessel process;

$$\rho_t = B_t + \int_0^t \frac{1}{4\rho_s} ds. \tag{15}$$

The asymptotic behavior of the stopping time $\hat{\tau}_c$ in (6) and the sequential estimators $\hat{\phi}_{\hat{\tau}_c}$ in (7) is given as follows:

$$\begin{aligned}
\hat{\tau}_c &\to_p \infty, \\
\hat{\phi}_{\hat{\tau}_c} - \phi &\to_p 0, \\
s^2_{\hat{\tau}_c} &\to_p \sigma^2, \\
\left(\sqrt{c}\hat{\phi}_{\hat{\tau}_c}, \frac{\hat{\tau}_c}{\sqrt{c}} \right) &\Rightarrow \left(\int_0^{U_1} W_u dW_u, U_1 \right) \qquad (16) \\
&= \left(B_1, \int_0^1 \frac{1}{2\rho_s} ds \right).
\end{aligned}$$

Remark 3. We remark that LS83 also examine the limiting behavior of $\hat{\tau}_c/\sqrt{c}$, where they obtain the limit as $\inf\{t : \int_0^t W_s^2 ds = 1\}$. In the theorem above, we derive an alternative representation of the limit of the stopping time using a Bessel process. In the following section, we will see that pallarel results hold under the local alternatives.

Corollary 4. *Under the same assumptions as in Theorem 2, we get the asymptotic expectation*

$$E(\hat{\tau}_c/\sqrt{c}) \to 2E(\rho_1) = 2\frac{\sqrt{2}\Gamma(5/4)}{\Gamma(3/4)} = 2.0921,$$

using the following density of the Bessel process,

$$p_t^{\alpha}(0, y) = 2^{-\nu} t^{-(\nu+1)} \Gamma(\nu+1)^{-1} y^{2\nu+1} \exp(-y^2/2t),$$

where $t = 1$, $\alpha = 3/2$,$\nu = (\alpha/2) - 1 = -1/4$.

Next we consider the asymptotic behavior of $s_{\hat{\tau}_c}^2$.

Theorem 5. *Suppose* x_n *is generated by the model (8) with* ϵ_n *having a finite fourth moment. Put* $\mu_4 = E(\epsilon_n^4)$, $\mu_3 = E(\epsilon_n^3)$ *and*

$$\sigma_{\epsilon^2}^2 = E\left[\left(\epsilon_n^2 - \sigma^2\right)^2\right] = \mu_4 - \sigma^4. \tag{17}$$

Then as $c \uparrow \infty$,

$$\frac{c^{1/4}}{\sigma_{\epsilon^2}}\left(s_{\hat{\tau}_c}^2 - \sigma^2\right) \Rightarrow \frac{1}{U_1} W'(U_1)$$

where W' *is a Brownian motion satisfying* $\langle W', W\rangle_1 = \mu_3/\sigma_{\epsilon^2}\sigma$ *with* W *being the Brownian motion in (??). Furthmore, if* $\mu_3 = 0$, *then* W *and* W' *are independent and*

$$\frac{c^{1/2}}{\sigma_{\epsilon^2}\hat{\tau}_c^{1/2}}\left(s_{\hat{\tau}_c}^2 - \sigma^2\right) \Rightarrow N(0,1).$$

3.3 Testing procedures

Using the results above, we propose three testing procedures of $\beta = 1$ or $\phi = 0$, namely, t test, stopping time test, and Bonferroni test combining them described in the following. We explain the procedures only for one sided test considering the alternatives of $\beta < 1$ ($\phi < 0$) or $\beta > 1$ ($\phi > 0$). Obviously, we straightforwardly change the critical region when we would like to implement a two sided test.

First, t test (T hereafter) uses the asymptotic normality of $\hat{\phi}_{\hat{\tau}_c}$. Since the null hypothesis is $\phi = 0$ and its asymptotic variance equals to unity, the T simply looks at $\sqrt{c}\hat{\phi}_{\hat{\tau}_c}$. Let z_α be the α quantile of the standard normal distribution. We reject the null if

$$\sqrt{c}\hat{\phi}_{\hat{\tau}_c} < z_\alpha$$

for left sided test with the alternative of $\beta < 1$, and if

$$\sqrt{c}\hat{\phi}_{\hat{\tau}_c} > z_{1-\alpha}$$

for the right sided test. Second, it is also possible to construct a test using the stopping time $\hat{\tau}_c/\sqrt{c}$ whose asympototic limit is U_1 under the null. We call it ST. Its distribution is not standard, but we can easily obtain its quantiles from (13) by numerical computation or simulation. Under the stationary alternatives, the stopping time tends to be large as shown in the section 5, while if the GDP is explosive or $\beta > 1$, the sequential procedure stops earlier than unit root case. Therefore, letting u_α be the α quantile of U_1, we rejct the null if

$$\hat{\tau}_c/\sqrt{c} > u_{1-\alpha}$$

against the alternative of $\beta < 1$. If

$$\hat{\tau}_c/\sqrt{c} < u_\alpha$$

we reject the null against $\beta > 1$. Third, we can combine both statistics by Bonferroni test as follows, which we call BON. When we want to test the existence of unit root against the alternative of stationarity, we reject the null when

$$\sqrt{c}\hat{\phi}_{\hat{\tau}_c} < z_{\alpha/2} \; or \; \hat{\tau}_c/\sqrt{c} > u_{1-\alpha/2}.$$

Obviously, it will be a conservative test as is always the case with Bonferroni tests. We shall show some empirical performances by simulation in Section 5.

4 Asymptotic properties of the test statistics under a near unit root AR(1)

In this section, we examine the properties of the test statistics under the local alternatives

$$\left(1-(1-\frac{\delta}{\sqrt{c}})L\right)x_n = \epsilon_n \tag{18}$$

$n = 1, 2, \ldots$, where $c > 0$, $\epsilon_n \sim i.i.d.(0, \sigma^2)$ and are independent of the initial value x_0. We obtain parallel results to the null case shown in the previous section. We use a diffusion approximation on $D[0, \infty)$ again.

In this setup, we test $H_0 : \delta = 0$ vs $H_1 : \delta > 0$ or $H_0 : \delta = 0$ vs $H_1 : \delta < 0$. The alternative hypothesis we consider here is stationarity in the former case, while explosive in the latter. Let $\phi^c = -\delta/\sqrt{c}$, then (18) can be rewritten as

$$\Delta x_n = \phi^c x_{n-1} + \epsilon_n. \tag{19}$$

The stopping time and estimator are as in (5), (6) and (7). The only difference from the null case is the data generating process.

4.1 Convergence to an Orstein=Uhlenbeck process on $D[0, \infty)$

Now we provide a result on the diffusion approximation under the null. Bobkoski (1983) proved the following theorem in $C[0, 1]$, but we provide the theorem in $D[0, \infty)$.

Theorem 6. *Suppose $x_1, x_2, \ldots$ are generated by the model (18);*

$$\left(1-(1-\frac{\delta}{\sqrt{c}})L\right)x_n = \epsilon_n \quad n = 1, 2, \ldots \tag{20}$$

where $\epsilon_n \sim i.i.d.(0, \sigma^2)$ with $\sigma^2 \in (0, \infty)$. We assume that the initial value x_0 is independent of ϵ_n, $n \geq 1$. Let

$$X_c(t) = \frac{x_{\lfloor \sqrt{c}t \rfloor}}{c^{1/4}\sigma}, \tag{21}$$

and X^δ be the Ornstein-Uhlenbeck (OU) process;

$$X^\delta(t) = -\delta \int_0^t X^\delta(s)ds + W(t), \tag{22}$$

where $W(t)$ is a Brownian motion. Then,

$$X_c \Rightarrow X^\delta,$$

as $c \uparrow \infty$ in the sense of $D[0, \infty)$.

Remark 7. X of 10 in the previous section coincides with X^0, or X^δ with $\delta = 0$. Comparing the result 10 and 6, we immediately know that the contiguity holds in the present setting. That is, 10 is a special case of this theorem with $\delta = 0$.

4.2 Asymptotic properties of τ_c, $\hat{\tau}_c$, $\hat{\phi}_{\hat{\tau}_c}$, and $s^2_{\hat{\tau}_c}$ under the local alternatives

To present the main theorem, write $X^\delta_t = X^\delta(t)$ for brevity hereafter. Define the following martingale M_t and its quadratic variation,

$$M_t = \int_0^t X^\delta_u dW_u \tag{23}$$

$$\langle M \rangle_t = \int_0^t (X^\delta_u)^2 du. \tag{24}$$

and put

$$U^\delta_s = \langle M \rangle_s^{-1} = \inf \left\{ t \geq 0 : \int_0^t (X^\delta_u)^2 du = s \right\} \tag{25}$$

where W and X^δ are respectively the Brownian motion and OU process in (22). By DDS Theorem (Theorem 1.6 in Revuz and Yor (1999) pp181) again,

$$B_s = M_{U^\delta_s} \tag{26}$$

is a Brownian motion with respect to the filtration $\mathcal{G}_s = \mathcal{F}_{U^\delta_s}$ and this is also a DDS Brownian motion.

Here is the main theorem of this section.

Theorem 8. *Suppose x_n is generated by the model (18) with an initial value x_0 independent of ϵ_n, $n \geq 1$. Then*

$$\rho^\delta_s = (X^\delta_{U_s})^2/2,$$

ρ^δ_t is a 3/2-dimensional Bessel process with drift $-\delta$;

$$\rho^\delta_t = B_t + \int_0^t \left(\frac{1}{4\rho^\delta_s} - \delta \right) ds. \tag{27}$$

The asymptotic behavior of the stopping time $\hat{\tau}_c$ in (6) and the sequential estimators $\hat{\phi}_{\hat{\tau}_c}$ in (7) is given as follows:

$$\begin{aligned} \hat{\tau}_c &\to_p \infty, \\ \hat{\phi}_{\hat{\tau}_c} - \phi^c &\to_p 0, \\ s^2_{\hat{\tau}_c} &\to_p \sigma^2, \\ \left(\sqrt{c}\hat{\phi}_{\hat{\tau}_c}, \frac{\hat{\tau}_c}{\sqrt{c}} \right) &\Rightarrow \left(-\delta + \int_0^{U^\delta_1} X^\delta_u dW_u, U^\delta_1 \right) \\ &= \left(-\delta + B_1, \int_0^1 \frac{1}{2\rho^\delta_s} ds \right) \end{aligned} \tag{28}$$

as $c \uparrow \infty$.

Remark 9. Here again, we immediately know that this theorem includes Theorem 2 in the previous section as a special case when $\delta = 0$.

Remark 10. Figure ** shows the contour of the joint distribution of $(\sqrt{c}\hat{\phi}_{\hat{\tau}_c}, \hat{\tau}_c/\sqrt{c})$ obtained by simulation when the DGP is

$$x_i = (1 - \frac{delta}{10000})x_{i-1} + \epsilon_i$$

for $delta = 0, 1$. Obviously, $delta = 0$ and 1 respectively correspond to the null and the alternative of $\beta = 0.9999$. The red line is the null distribution and blue line is the stationary alternative. We see that $\hat{\phi}_{\hat{\tau}_c}$ and $\hat{\tau}_c$ are obviously dependent. The stopping time tends to be larger for stationary case than the unit root case.

Remark 11. Local asymptotic normality (LAN).

Suppose that the disturbances are normally independently distributed, $\varepsilon_n \sim i.i.d.N(0, \sigma^2)$. Then the log likelihood ratio of the observations is

$$\Lambda(x_0, x_1, \ldots, x_{\tau_c}; \delta/\sqrt{c}) = -\frac{\delta}{\sqrt{c}\sigma^2} \sum_{n=1}^{\tau_c} (x_n - x_{n-1})x_{n-1} - \frac{\delta^2}{2} + \frac{\delta^2}{2c}(c - \frac{1}{\sigma^2} \sum_{n=1}^{\tau_c} x^2_{n-1}).$$

Figure 1: Contour of the joint distribution of $(\sqrt{c}\hat{\phi}_{\hat{\tau}_c}, \hat{\tau}_c/\sqrt{c})$

../Dropbox/HNN/surt/draft/Local-AR(1)/joint-density-d0-d1.pdf

The first term on the right is asymptotically $N(0,1)$ under the null as proved in Lai and Siegmund (1983) and the third term converges in probability to zero as $c \to \infty$. Therefore, this model possesses a LAN property and the first term on the right is asymptotically $N(\delta, 1)$ under the alternative by LeCam's third lemma.

We discuss the relationship between the joint density function of $(\hat{\phi}_{\hat{\tau}_c}, \hat{\tau}_c)$ under the null and the local alternative. From Theorems 2 and 8, we can establish the following corollary, which can be regarded as LAN in terms of the joint limit.

Corollary 12. Let $f^0(z,u)$ be the joint density of the normalized estimator and stopping time under the null obtained in Theorem *2*. The joint density function of $(-\delta + B_1, \int_0^1 \frac{1}{2X_s^\delta} ds)$ under the local alternative above is given by

$$f^\delta(z,u) = \exp(-\delta z - \frac{1}{2}\delta^2) f^0(z,u).$$

This corollary implies that the log likelihood ratio turns out

$$\log \frac{f^\delta(z,u)}{f(z,u)} = -\delta z - \frac{1}{2}\delta^2,$$

which indicates that this is LAN, and it only depends on $\hat{\phi}_{\hat{\tau}_c}$.

The following theorem gives the asymptotic distribution of $s^2_{\hat{\tau}_c}$.

Theorem 13. *Suppose* ϵ_n *has a finite fourth moment. Put* $\mu_4 = E(\epsilon_n^4)$ *,* $\mu_3 = E(\epsilon_n^3)$ *and*

$$\sigma^2_{\epsilon^2} = E\left[\left(\epsilon_n^2 - \sigma^2\right)^2\right] = \mu_4 - \sigma^4. \tag{29}$$

Let W *be the Brownian motion in (34). Then as* $c \uparrow \infty$*,and*

$$\frac{c^{1/4}}{\sigma_{\epsilon^2}}\left(s^2_{\hat{\tau}_c} - \sigma^2\right) \Rightarrow \frac{1}{U_1^\delta} W'\left(U_1^\delta\right)$$

where W' *is a Brownian motion satisfying* $\langle W', W \rangle_1 = \mu_3/\sigma_{\epsilon^2}\sigma$ *with* W *being the Brownian motion in (22). Furthermore, if* $\mu_3 = 0$*, then* W *and* W' *are independent and*

$$\frac{c^{1/2}}{\sigma_{\epsilon^2}\hat{\tau}_c^{1/2}}\left(s^2_{\hat{\tau}_c} - \sigma^2\right) \Rightarrow N(0,1).$$

4.3 Power against local alternatives

From the above theorems, we easily know that the test using $\hat{\phi}_{\hat{\tau}_c}$ has a nontrivial power against the local alternatives because

$$\sqrt{c}\hat{\phi}_{\hat{\tau}_c} \Rightarrow B_1 \quad \textit{under } H_0$$
$$\sqrt{c}\hat{\phi}_{\hat{\tau}_c} \Rightarrow B_1 - \delta \textit{ under } H_1$$

asympototically. This is a standard result. We can also use the stopping time $\hat{\tau}_c$ to test the null using

$$\frac{\hat{\tau}_c}{\sqrt{c}} \Rightarrow \int_0^1 \frac{1}{2\rho_s} ds \quad under\ H_0$$
$$\frac{\hat{\tau}_c}{\sqrt{c}} \Rightarrow \int_0^1 \frac{1}{2\rho_s^\delta} ds \quad under\ H_1.$$

The distribution of the two integrals differ in such a way that the stopping time tends to be larger when DGP follows a local alternatives with $\delta > 0$. If $\delta < 0$, it becomes smaller. Therefore, if the stopping time exceeds $(1-\alpha)$ quantile of $\int_0^1 \frac{1}{2\rho_s} ds$, we may stop sampling and reject the null against the stationary alternative, and vice versa. By this procedure, the test again has a non-trivial power. Both procedures work as unit root tests, but it is not so clear which possesses more power. One advatage of the latter is that we can stop sampling earlier than the former when the alternative is statinary. In this sense, it may be favourable if sampling cost is extremely high.

5 Simulation

5.1 Simulation settings

We conduct simulation to examine the performance of the three tests T, ST and BON explained in Section 3.3. We provide the performance of test for the null versus stationary alternatives in section 5.2 and that for explosive alternatives in section 5.3. We consider local (near unity) alternatives as well as non-local alternatives to see the power of the tests.

The data generation process is

$$x_n = \beta x_{n-1} + \epsilon_i$$

where $\epsilon_n \sim i.i.d.\ N(0,1)$ and $x_0 = 0$. We compare three tests under two kinds of alternative hypothesis. The first hypotheses setting is unit root v.s. stationary alternatives,

$$H_0 : \beta = 1$$
$$H_1 : \beta < 1$$

which is considered in section 5.2, while the second is unit root v.s. explosive alternatives,

$$H_0 : \beta = 1$$
$$H_1 : \beta > 1$$

which is dealt in section 5.3. The critical region is different depending on which alternative is considered either $\beta < 1$ or $\beta > 1$. Therefore, not only the power but the size depends on the alternative setting. It is also possible to consider two-sided tests(***??***). The size must be informative, but the power properties are totally different depending on the true DGP. As the statinary alternatives, we take $\beta = 0.95$, 0.99 while we set $\beta = 1.01$, 1.05 as the explosive cases. We prove Theorems in the previous section under local alternatives, however we also report simulation results for a stationary case of $\beta = 0.80$ as a reference.

For testing against the stable alternatives with size $\alpha = 0.05$, T, ST and BON reject the null respectively when $\sqrt{c}\hat{\phi}_c < z_{0.05}$, $\hat{\tau}_c/\sqrt{c} > u_{0.95}$, and $\sqrt{c}\hat{\phi}_c < z_{0.025}$ or $\hat{\tau}_c/\sqrt{c} > u_{0.975}$. Note that if the DGP is a stationary process, the ST can possibly reject the null before the sampling hits the stopping time. Then it requires a shorter time to make a decision than the other tests, which may be favorable under a large sampling cost. For explosive alternatives, all the three testing schemes conclude at the same time after the sampling is stopped by the stopping rule. Note that ST uses the information only of the stopping time, while T and BON make use of both the estimate and the stopping time.

We take three predetermined values, $c = 600$, 2500, 10000 and the number of replication is 10000. Since we employ the asymptotic theory as $c \uparrow \infty$, we expect a larger value of c provides better asymptotic approximation.

Table 1: Size of tests when alternative hypothesis is stationary

$\beta = 1$

$c = 600$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.0518	2.0748770
ST	0.0442	2.0453525
BON	0.0297	2.0624663
$c = 2500$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.0465	2.0693040
ST	0.0438	2.0399180
BON	0.0296	2.0571760
$c = 10000$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.0488	2.0932680
ST	0.0463	2.0628650
BON	0.0300	2.0801700

5.2 Stationary alternatives

Table 1 reports the size of tests. The nominal size is 5%. The first column indeicates the tests, the second and third columns show respectively the rejection rate and the average of standardized stopping time $\hat{\tau}_c/\sqrt{c}$. The size of T and ST appears slightly conservative for all c, but mostly closer to the nominal than BON as expected. The size distortion ofBON is about 2% for all c. Average time to make a decision of ST is slightly shorter than the other tests as expected. Linetsky derive the density of Bessel processes which gives the expectation of the limit of $\hat{\tau}_c/\sqrt{c}$ under the null as $E(U_1) = 2.0921$. The mean of $\hat{\tau}_c/\sqrt{c}$ mostly close to this value, especially in the case of $c = 10000$.

Table 2 shows power of the tests and the standardized average stopping time. T has the highest power with the longest stopping time for all values of β and c. When $\beta = 0.99$ and $c = 10000$, T possesses a higher power by 3% than the ST with 8% larger stopping time. When $\beta = 0.95$ and $c = 10000$, the power of both S and ST are almost unity, but the stopping time is significantly different. T requires about 1000 observations to make a decision in average, while ST needs only around 420 observations. We also report the results when $\beta = 0.80$ which is not regarded as a local alternative. For $c = 10000$, the stopping time of ST test in this case is 424 in average. This is significantly smaller than that for T test which is more than 3600. BON has the lowest power among the three tests for all cases partly because of a conservative size.

If the sampling cost is expensive, one may be willing to employ the ST test. Otherwise s/he can choose the T test.

5.3 Explosive alternatives

In this subsection, we examine the small sample properties of the tests when the alternative hypothesis is an explosive AR(1) process or $\beta > 1$. Table 3 gives the size of the tests and the average standardized stopping time. The size of T and the ST are close to the nominal value when $c = 10000$, while it is slightly larger for $c = 600$. BON is conservative by about $0.8 - 0.9\%$ for all values of c though larger c appears to give slightly smaller size distortion. As explained above, the stopping time is exactly the same for all three tests because all tests make decisions after the sampling is stopped unlike the stationary alternative case.

Table 4 provides the power of tests for $\beta = 1.01$ and $\beta = 1.05$. As shown in the table, T has the highest power in all settings. T outperforms ST because when ST rejects the null, T also rejects the null in most cases, while T often rejects the null even if ST does not. BON has a higher power than ST except the case of $c = 600$ and $\beta = 1.01$. It might be the effect of smaller size distortion of BON than the stable alternative case (???). T always outperforms ST and BON for the explosive alternative case, thus we should employ T.

Table 2: Power of tests when alternative hypothesis is stationary

$\beta = 0.99$

$c = 600$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.0865	2.3023081
ST	0.0694	2.2514526
BON	0.0553	2.2797115

$c = 2500$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.1288	2.5729220
ST	0.1129	2.4831220
BON	0.0829	2.5293460

$c = 10000$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.2630	3.1839720
ST	0.2357	2.9486160
BON	0.1909	3.0518460

$\beta = 0.95$

$c = 600$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.3362	3.3784546
ST	0.2766	3.1016950
BON	0.2501	3.2270517

$c = 2500$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.8019	5.4098460
ST	0.7457	3.9848560
BON	0.7213	4.3702700

$c = 10000$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.9996	10.026283
ST	0.9989	4.239397
BON	0.9989	4.788537

$\beta = 0.8$

$c = 600$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.9990	9.0158491
ST	0.9980	4.2444717
BON	0.9981	4.8140883

$c = 2500$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	1.0000	18.105178
ST	1.0000	4.2400000
BON	1.0000	4.8000000

$c = 10000$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	1.0000	36.045610
ST	1.0000	4.2400000
BON	1.0000	4.7900000

Table 3: Size of tests when alternative hypothesis is explosive

$\beta = 1$

$c = 600$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.0590	2.0675163
ST	0.0504	2.0675163
Bon	0.0415	2.0675163

$c = 2500$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.0509	2.0859180
ST	0.0440	2.0859180
Bon	0.0420	2.0859180

$c = 10000$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.0506	2.0973950
ST	0.0503	2.0973950
Bon	0.0424	2.0973950

Table 4: Power of tests when alternative hypothesis is explosive

$\beta = 1.01$

$c = 600$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.0784	1.8900467
ST	0.0667	1.8900467
Bon	0.0569	1.8900467

$c = 2500$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.1196	1.7052720
ST	0.0823	1.7052720
Bon	0.0978	1.7052720

$c = 10000$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.2624	1.4189670
ST	0.1550	1.4189670
Bon	0.2060	1.4189670

$\beta = 1.05$

$c = 600$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.3299	1.3404179
ST	0.1849	1.3404179
Bon	0.2531	1.3404179

$c = 2500$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.8010	0.94602
ST	0.4100	0.94602
Bon	0.7300	0.94602

$c = 10000$	rejection rate	mean of $\hat{\tau}_c/\sqrt{c}$
T	0.9996	0.610923
ST	0.8233	0.610923
Bon	0.9990	0.610923

5.4 Comparison of the simulation and numerical computation results

****

We can evaluate the powers of our testing procedure and the expected stopping time by numerical calculations given an alternative parameter value. We can numerically compute the theoreical expected stopping time given c and a parameter value by the transition densities of Bessel diffusions with constant drift obtained by Linetsky (2004). We find that our theoretical results are consistent with the simulations.

****

6 Conclusion

Considering AR(1) process, we obtain the asymptotic distribution of the OLS estimator of the AR(1) parameter and the Fisher information based stopping time under a sequential sampling both under the unit root process and near unit root process. The t statistic is asymptotically normally distributed and the stopping time is characterized by Bessel processes. We employ diffusion approximation to prove the results which enables us to analyze the asymptotic properties in the case of local alternatives of near unit root processes. Based on the results, we propose three kinds of unit root tests using the t statistic, the stopping time and the both by Bonferroni approach. When the alternative is a stationary process, we show that the stopping time can be a useful test statistic especially when the sampling cost is large. If the alternative is an explosive process, t test is shown to perform the best. The Bonferrroni approach is one way of using both of t value and stopping time, but it is likely to be able to construct a better test exploiting both information.

Sequential probability ratio test (SPRT) based on the likelihood ratio is commonly used in sequential tests, because it is the most powerful test against a simple alternative hypothesis. It is possible to apply this approach in the present unit root testing. We comjecture that it could potentially ourperform the t and stopping time tests we propose in this paper. A practical disadvantage of SPRT is that the alternative needs to be simple. In the present context, this limitation is obviously inconvenient because we typically do not know a likely value under the alternative. One possibile approach employing SPRT is that we plug an estimator in the likelihood under the alternative to construct a test statistic, but its statistical properties are unknown. Research toward this direction has been conducted in part and is currently going on.

This paper deals with only AR(1) process with i.i.d. inovations, but we can consider a variety of extention to e.g. AR(1) processes with martingale diffrence innovations, AR(p) processes, nonparametrically correlated linear processes, GARCH processes and others, which are currently under way in part.

Appendix 1

Proof of Theorem 6

We shall use a diffusion approximation on space $D[0,\infty)$, so briefly show some characteristics of the space (see Billingsley (1999) for details). According to Billingsley (1999), $D[0,\infty)$ is a Polish space; a complete separable metric space with a suitable metric. We start with a simple lemma which enable us to deal with weak convergence in $D[0,\infty)$ easily.

We define the sup norms $\|\cdot\|_m$ for $m > 0$ and $\|\cdot\|_\infty$ for $f : [0,\infty) \to R$;

$$\|f\|_m = \sup_{t\in[0,m]} |f(t)|, \quad \|f\|_\infty = \sup_{t\in[0,\infty)} |f(t)|, \tag{30}$$

and also define the metric of $C[0,\infty)$, the set of the continuous functions on $[0,\infty)$;

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

by which $C[0,\infty)$ becomes a Polish space.

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

$$\|\lambda\|_m^\circ = \sup_{0\le s<t\le m} \left| \log \frac{\lambda t - \lambda s}{t-s} \right|.$$

Let

$$d_m^\circ(x,y) = \inf_{\lambda\in\Lambda_m} \left\{ \|\lambda\|_m^\circ \vee \|x - y\lambda\|_m \right\}.$$

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

$$f^m(t) = h_m(t) f(t), \quad t \ge 0.$$

Define the metric on $D[0,\infty)$ for $f, g \in D[0,\infty)$;

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

Let Λ_∞ be the set of strictly increasing, continuous maps of $[0,\infty)$ onto itself.

Theorem 14. *(Billingsley (1999) p.168 , Th16.1). Let $I : [0,\infty) \to [0,\infty)$ be the identity map; $I(t) = t$. Then, * in $D[0,\infty)$ as $n \to \infty$ if and only if there exist elements λ_n of Λ_∞ such that*

$$\|\lambda_n - I\|_\infty \to 0, \tag{31}$$

and for each m,

$$\|f_n \circ \lambda_n - f\|_m \to 0. \tag{32}$$

Based on this theorem, we can derive the following lemma.

Lemma 15. *Suppose $f_n \in D[0,\infty)$ and $f \in C[0,\infty)$. Then, $\lim_{n\to\infty} d_\infty^\circ(f_n, f) = 0$ if and only if $\lim_{n\to\infty} \|f_n - f\|_m = 0$ for any $m > 0$.*

Proof. Suppose (31) and (32) 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 is $0 < \delta < m$ such that $|f(s) - f(t)| < \epsilon/2$ for any $s,t \in [0,2m]$ satisfying $|s-t| < \delta$. Choose n_0 so that $\|\lambda_n - I\|_\infty < \delta$ and $\|f_n \circ \lambda_n - f\|_{2m} < \epsilon/2$ for any $n \ge n_0$. Then, we can get $\left|t - \lambda_n^{-1} t\right| < \delta$, and

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

□

Proof of Theorem 6

Let $S_n = \epsilon_1 + \epsilon_2 + \cdots + \epsilon_n$ with $S_0 = 0$, W be a Brownian motion, and

$$W_c(t) = S_{\lfloor\sqrt{c}t\rfloor}/c^{1/4}\sigma. \tag{33}$$

Then, by the functional central limit theorem for martingale differences (Theorem 18.2 in Billingsley (1999)) , as $c \uparrow \infty$,

$$W_c \Rightarrow W. \tag{34}$$

Since $D[0,\infty)$ is Polish, we can use Skorohod's representation Theorem (Theorem 6.7 in Billingsley (1999)) and create a new probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ under which $\tilde{W}$ is a Brownian motion, $\tilde{W}_c \in D[0,\infty)$ has the same distribution as W_c in $D[0,\infty)$, and

$$\left\|\tilde{W}_c - \tilde{W}\right\|_m \to 0 \tag{35}$$

as $c \uparrow \infty$ any $m > 0$ *a.s.*, where $\|\|_m$ is the sup norm defined in (30).

Let $\beta_c = 1 - \delta/\sqrt{c}$, we can write x_n in (18) as follows:

$$x_n = \sum_{k=1}^{n-1} \beta_c^k \epsilon_{n-k} + \beta_c^n x_0.$$

Then x_n can be rewritten as

$$\begin{aligned} x_n &= \beta_c^n x_0 + \sum_{i=1}^{n} \beta_c^{n-i}\epsilon_i \\ &= \beta_c^n x_0 + \sum_{i=1}^{n} \beta_c^{n-i}(S_i - S_{i-1}) \\ &= \beta_c^n x_0 - (1-\beta_c)\sum_{i=1}^{n} \beta_c^{n-i-1} S_i + \beta_c^{-1} S_n. \end{aligned}$$

Here, we define $X_c(t) = x_{\lfloor\sqrt{c}t\rfloor}/c^{1/4}\sigma$ and $X_0 = x_0/c^{1/4}\sigma$, then we have

$$X_c(t) = \frac{1}{c^{1/4}\sigma}\left(\beta_c^{\lfloor\sqrt{c}t\rfloor} x_0 - \frac{\delta}{\sqrt{c}}\sum_{i=1}^{\lfloor\sqrt{c}t\rfloor} \beta_c^{\lfloor\sqrt{c}t\rfloor - i - 1} S_i + \beta_c^{-1} S_{\lfloor\sqrt{c}t\rfloor}\right). \tag{36}$$

From (35), we can obtain

Lemma 16. *Define*

$$X(t) = -\delta e^{-\delta t}\int_0^t e^{\delta s} W(s)ds + W(t), \tag{37}$$

then $X_c \Rightarrow X$ as $c \uparrow \infty$ in the sense of $D[0,\infty)$.

Proof. Consider $\tilde{W}_c$ and $\tilde{W}$ of (35) and observe

$$\sum_{i=1}^{\lfloor\sqrt{c}t\rfloor} \beta_c^{-i-1}\frac{S_i}{c^{1/4}\sigma}\frac{1}{\sqrt{c}} = \int_0^{\lfloor\sqrt{c}t\rfloor/\sqrt{c}} \beta_c^{-\lfloor\sqrt{c}s\rfloor - 1} W_c(s)ds.$$

Then,

$$\sup_{t\le m}\left|\int_0^{\lfloor\sqrt{c}t\rfloor/\sqrt{c}} \beta_c^{-\lfloor\sqrt{c}s\rfloor-1}\tilde{W}_c(s)ds - \int_0^t \beta_c^{-\lfloor\sqrt{c}s\rfloor-1}\tilde{W}_c(s)ds\right| \le \sup_{t\le m}\left|\beta_c^{-\lfloor\sqrt{c}t\rfloor-1}\tilde{W}_c(s)\frac{1}{\sqrt{c}}\right| \to 0 \quad a.s.$$

Hence it suffices to show that

$$\begin{aligned}\tilde{X}_c(t) &= -\frac{\delta}{\sqrt{c}}\beta_c^{\lfloor\sqrt{c}t\rfloor}\int_0^{\lfloor\sqrt{c}t\rfloor/\sqrt{c}}\beta_c^{-\lfloor\sqrt{c}s\rfloor-1}\tilde{W}_c(s)ds + \beta_c^{-1}\tilde{W}_c(t)\\ &\to \tilde{X}(t) := e^{-\delta t}\tilde{X}_0 - \delta e^{-\delta t}\int_0^t e^{\delta s}\tilde{W}(s)ds + \tilde{W}(t) \qquad (38)\end{aligned}$$

uniformly in $t \in [0, m]$ for any $m > 0$. To show (38), first we will show $\beta_c^{\lfloor\sqrt{c}t\rfloor} \to e^{-\delta t}$ uniformly in $t \in [0, m]$ for any $m > 0$. Since

$$e^{-\delta t} = \sum_{k=0}^{\lfloor\sqrt{c}t\rfloor}\frac{(-\delta t)^k}{k!} + \sum_{\lfloor\sqrt{c}t\rfloor+1}^{\infty}\frac{(-\delta t)^k}{k!},$$

and

$$\begin{aligned}\beta_c^{\lfloor\sqrt{c}t\rfloor} &= \left(1-\frac{\delta}{\sqrt{c}}\right)^{\lfloor\sqrt{c}t\rfloor}\\ &= \sum_{k=0}^{\lfloor\sqrt{c}t\rfloor}\frac{(-\delta t)^k}{k!}\left(\frac{n}{\sqrt{c}t}\right)^k\frac{n(n-1)\cdots(n-k+1)}{n^k},\end{aligned}$$

where $n = \lfloor\sqrt{c}t\rfloor$. For any $m > 0$, by Dini's theorem, we have

$$\begin{aligned}&\sup_{t\le m}\left|e^{-\delta t} - \beta_c^{\lfloor\sqrt{c}t\rfloor}\right|\\ =&\sup_{t\le m}\left|\sum_{k=0}^{\lfloor\sqrt{c}t\rfloor}\frac{(-\delta t)^k}{k!}\left(1-\left(\frac{n}{\sqrt{c}t}\right)^k\prod_{i=1}^{k-1}\left(1-\frac{i}{n}\right)\right) + \sum_{\lfloor\sqrt{c}t\rfloor+1}^{\infty}\frac{(-\delta t)^k}{k!}\right|\\ \le&\sup_{t\le m}\left|\sum_{k=0}^{\lfloor\sqrt{c}t\rfloor}\frac{(\delta t)^k}{k!}\left(1-\left(1-\frac{(\sqrt{c}t-n)}{\sqrt{c}t}\right)^k\prod_{i=1}^{k-1}\left(1-\frac{i}{n}\right)\right) + \sum_{\lfloor\sqrt{c}t\rfloor+1}^{\infty}\frac{(\delta t)^k}{k!}\right|\\ \le&\sup_{t\le m}\left|\sum_{k=0}^{\infty}\frac{(\delta t)^k}{k!}\left(1-\left(1-\frac{1}{\sqrt{c}t}\right)^k\prod_{i=1}^{k-1}\left(1-\frac{i}{n}\right)\right) + \sum_{\lfloor\sqrt{c}t\rfloor+1}^{\infty}\frac{(\delta t)^k}{k!}\right|\\ \to&0.\end{aligned}$$

Hence (38) holds uniformly in $t \in [0, m]$ for any $m > 0$. □

Next, as to X_t, using d-dimentional Itô's formula (Theorem in Revus and Yor (1999) pp147), we will see (22). In fact, put $Y_t = X_t e^{\delta t}$, we can find $dY_t = e^{\delta t}dW_t$. Letting $X_t = e^{-\delta t}Y_t$, we have the differential of X_t;

$$\begin{aligned}dX_t &= d(e^{-\delta t}Y_t)\\ &= e^{-\delta t}dY_t + Y_t d(e^{-\delta t})\\ &= dW_t - \delta e^{-\delta t}Y_t dt\\ &= -\delta X_t dt + dW_t.\end{aligned}$$

Appendix 2

6.1 Proof of Theorem 8

To prove the asymptotic property of $\sqrt{c}\hat{\phi}^c_{\tau_c}$, we show the following lemmas.

Lemma 17. *Let $I : [0,\infty) \to [0,\infty)$ be the identity map; $I(t) = t$ and*

$$I_c(t) = \sum_{n=1}^{\lfloor \sqrt{c}t \rfloor} \epsilon_n^2 / \sqrt{c}\sigma^2. \tag{39}$$

Then,

$$\|I_c - I\|_m = \sup_{t \le m} |I_c(t) - t| \to 0$$

for any $m > 0$ a.s as $c \uparrow \infty$.

Proof. Fix $\omega \in \left\{ \sum_{n=1}^N \epsilon_n^2 / N \to \sigma^2 \right\}$ and for any $m > 0$ and $\varepsilon > 0$ find N_0 so that for any $N \ge N_0$

$$\left| \frac{1}{N} \sum_{n=1}^N \left(\frac{\epsilon_n^2}{\sigma^2} - 1 \right) \right| < \frac{\varepsilon}{2m}.$$

Then, for large enough $c > 0$,

$$\begin{aligned}
&\sup_{t \le m} |I_c(t) - t| \\
&= \sup_{t \le m} \left| \frac{1}{\sqrt{c}} \sum_{n=1}^{\lfloor \sqrt{c}t \rfloor} \left(\frac{\epsilon_n^2}{\sigma^2} - 1 \right) + \frac{\lfloor \sqrt{c}t \rfloor - \sqrt{c}t}{\sqrt{c}} \right| \\
&\le \max_{N \le N_0} \left| \frac{1}{\sqrt{c}} \sum_{n=1}^{N} \left(\frac{\epsilon_n^2}{\sigma^2} - 1 \right) \right| \vee \sup_{N_0 \le \sqrt{c}t \le \sqrt{c}m} \left| \frac{t}{\sqrt{c}t} \sum_{n=1}^{\lfloor \sqrt{c}t \rfloor} \left(\frac{\epsilon_n^2}{\sigma^2} - 1 \right) \right| + \frac{1}{\sqrt{c}} \\
&\le \frac{1}{4}\varepsilon + m\frac{\varepsilon}{2m} + \frac{1}{4}\varepsilon = \varepsilon.
\end{aligned} \tag{40}$$

□

Lemma 18. *Let*

$$F_c(t) = \frac{1}{\sigma^2 c} \sum_{n=1}^{\lfloor \sqrt{c}t \rfloor} x_{n-1}^2, \quad F(t) = \int_0^t X^2(u) du \tag{41}$$

and

$$J_c(t) = \frac{1}{\sigma^2 \sqrt{c}} \sum_{n=1}^{\lfloor \sqrt{c}t \rfloor} x_{n-1} \Delta x_n \quad J(t) = \int_0^t X(u) dX(u). \tag{42}$$

Under the assumption of Theorem 6,

$$(X_c, F_c, J_c, I_c) \Rightarrow (X, F, J, I), \tag{43}$$

in the sense of $D[0,\infty)^4$ as $c \uparrow \infty$, where X_c and X are defined in Theorem 6 and I_c and I in Lemma 17.

Proof. Using Skorohod's representation theorem for Theorem 6 and Lemma 17, we use

$$\left(\tilde{X}_c, \tilde{I}_c \right) \to \left(\tilde{X}, \tilde{I} \right)$$

a.s. in the sense of $D[0,\infty) \times D[0,\infty)$. Since

$$F_c(t) = \int_0^{\lfloor \sqrt{c}t \rfloor / \sqrt{c}} X_c^2(u) du,$$

we define

$$\tilde{F}_c(t) = \int_0^{\lfloor \sqrt{c}t \rfloor / \sqrt{c}} \tilde{X}_c^{\,2}(u) du, \quad \tilde{F}(t) = \int_0^t \tilde{X}^2(u) du. \tag{44}$$

For $\tilde{F}_c$, we obtain

$$\begin{aligned}\sup_{t\le M}\left|\tilde{F}_c(t)-\int_0^t \tilde{X}_c^2(u)du\right| &\le \sup_{t\le M}\left|\tilde{X}_c^2(t)\right|\frac{1}{\sqrt{c}}\\ &\to \sup_{t\le M}\left|\tilde{X}^2(t)\right|\times 0=0 \quad \text{for any } M>0\ a.s.\end{aligned}$$

Hence,

$$\tilde{F}_c \to \tilde{F}$$

in the sense of $D[0,\infty)$ a.s. For J_c, using

$$\frac{1}{\sigma^2\sqrt{c}}\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\epsilon_n = J_c(t)+\delta F_c(t), \tag{45}$$

we obtain

$$\begin{aligned}J_c(t) &= \frac{1}{\sigma^2\sqrt{c}}\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\Delta x_n\\ &= \frac{1}{\sigma^2\sqrt{c}}\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor}\left\{-\frac{1}{2}\left(\Delta x_n\right)^2+\frac{1}{2}x_n^2-\frac{1}{2}x_{n-1}^2\right\}\\ &= \frac{1}{2\sigma^2\sqrt{c}}\left\{x_{\lfloor\sqrt{c}t\rfloor}^2-x_0^2-\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor}\left(-\frac{\delta}{\sqrt{c}}x_{n-1}+\epsilon_n\right)^2\right\}\\ &= \frac{1}{2}\left\{X_c^2(t)-\frac{1}{\sigma^2\sqrt{c}}\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor}\left(-\frac{\delta}{\sqrt{c}}x_{n-1}+\epsilon_n\right)^2\right\}\\ &= \frac{1}{2}\left\{X_c^2(t)+\frac{\delta^2}{\sqrt{c}}F_c(t)+\frac{2\delta}{\sqrt{c}}J_c(t)-I_c(t)\right\}. \end{aligned}\tag{46}$$

Let

$$\tilde{J}_c(t)=\frac{1}{2}\left\{\tilde{X}_c^2(t)+\frac{\delta^2}{\sqrt{c}}\tilde{F}_c(t)-\tilde{I}_c(t)\right\}/\left(1-\frac{\delta}{\sqrt{c}}\right),\quad \tilde{J}(t)=\int_0^t \tilde{X}(u)d\tilde{X}(u). \tag{47}$$

Then

$$\begin{aligned}\tilde{J}_c(t) &= \frac{1}{2}\left\{\tilde{X}_c^2(t)+\frac{\delta^2}{\sqrt{c}}\tilde{F}_c(t)-\tilde{I}_c(t)\right\}/\left(1-\frac{\delta}{\sqrt{c}}\right)\\ &\to \left(\tilde{X}_t^2-t\right)/2=\tilde{J}(t)\end{aligned}\tag{48}$$

uniformly in $t\in[0,m]$ for any $m>0$. The last equation is obtained by the Itô's lemma;

$$X_t^2 = 2\int_0^t X_u dX_u + t.$$

Hence, (43) is obtained from

$$(X_c,F_c,J_c,I_c)\sim\left(\tilde{X}_c,\tilde{F}_c,\tilde{J}_c,\tilde{I}_c\right)\to\left(\tilde{X},\tilde{F},\tilde{J},\tilde{I}\right).$$

□

To prove the asymptotic property of τ_c, we need the following lemma.

Lemma 19. *Suppose that* $f \in C[0,\infty)$, $f_c \in D[0,\infty)$, *and*

$$\lim_{c\uparrow\infty} \|f_c - f\|_m = 0$$

for any $m > 0$. *Also suppose that* $s > 0$ *and a sequence* $\{s_{c,n}\}$ *satisfies*

$$s_{c,\lfloor\sqrt{c}t\rfloor} \to s,$$

as $c \uparrow \infty$ *for any* $t > 0$. *Then, if*

$$Leb\{s : f(s) = 0\} = 0,$$

where Leb is the Lebesgue measure, we have, as $c \uparrow \infty$,

$$\begin{aligned} h_c :=& \inf\left\{t > 0 : \int_0^t f_c^2(u)du = s_{c,\lfloor\sqrt{c}t\rfloor}\right\} \\ \to h :=& \inf\left\{t > 0 : \int_0^t f^2(u)du = s\right\}. \end{aligned}$$

Proof. Suppose $t < h$. There is a $\epsilon > 0$ such that $0 < \epsilon < s$ and $\int_0^t f^2(u)du < s - \epsilon$. Since $\int_0^t f_c^2(u)du \to \int_0^t f^2(u)du$ and $s_{c,\lfloor\sqrt{c}t\rfloor} \to s$ as $c \uparrow \infty$, $\int_0^t f_c^2(u)du < s - \epsilon < s_{c,\lfloor\sqrt{c}t\rfloor}$, so $t < h_c$, which implies $t \le \liminf_{c\uparrow\infty} h_c$. On the contrary, suppose $t' > h$. There is a $\epsilon' > 0$ such that $0 < \epsilon' < s$ and $s + \epsilon' < \int_0^{t'} f^2(u)du$. Since $\int_0^{t'} f_c^2(u)du \to \int_0^{t'} f^2(u)du$ and $s_{c,\lfloor\sqrt{c}t'\rfloor} \to s$ as $c \uparrow \infty$, $s_{c,\lfloor\sqrt{c}t'\rfloor} < s + \epsilon' < \int_0^{t'} f_c^2(u)du$, so $h_c < t'$. In the same way, we can get $\limsup_{c\uparrow\infty} h \leqq t'$. Since t and t' are arbitrary, $\lim_{c\uparrow\infty} h_c = h$. □

Lemma 20. *For* τ_c, *the convergence corresponding to(28) is true;*

$$\left(\sqrt{c}\hat{\phi}^c_{\tau_c}, \frac{\tau_c}{\sqrt{c}}\right) \Rightarrow (-\delta + B_1, U_1) \quad (c \uparrow \infty). \tag{49}$$

First, we rewrite τ_c and $\hat{\tau}_c$ as follows:

$$\frac{\tau_c}{\sqrt{c}} = \inf\left\{t : \frac{1}{\sigma^2 c}\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}^2 \ge 1\right\}, \tag{50}$$

$$\frac{\hat{\tau}_c}{\sqrt{c}} = \inf\left\{t : \frac{1}{s^2_{\lfloor\sqrt{c}t\rfloor} c}\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}^2 \ge 1\right\}. \tag{51}$$

Proof. Skorohod's representation theorem enables us to find the probability space in which the a.s.convergence

$$\left(\tilde{X}_c, \tilde{F}_c, \tilde{J}_c\right) \to \left(\tilde{X}, \tilde{F}, \tilde{J}\right)$$

holds in the sense of $D[0,\infty)^3$, which corresponds to the weak convergence $(X_c, F_c, J_c) \to (X, F, J)$ in Lemma 18 . Use $\tilde{F}_c(t)$ in (44) and define

$$\frac{\tilde{\tau}_c}{\sqrt{c}} = \inf\left\{t : \tilde{F}_c(t) \ge 1\right\} \tag{52}$$

and

$$\tilde{\tau}'_c = \inf\left\{\sqrt{c}t \ge 0 : \int_0^t \tilde{X}_c^2(u)du = 1\right\}.$$

Then

$$\left|\frac{\tilde{\tau}_c}{\sqrt{c}} - \frac{\tilde{\tau}'_c}{\sqrt{c}}\right| \le \frac{1}{\sqrt{c}}.$$

From Lemma 19 we have

$$\frac{\tilde{\tau}_c'}{\sqrt{c}} \to \inf\left\{t \geq 0 : \int_0^t \tilde{X}^2(u)du = 1\right\} := \tilde{U}_1.$$

Hence, since $\tilde{\tau}_c/\sqrt{c} \sim \tau_c/\sqrt{c}$,

$$\frac{\tau_c}{\sqrt{c}} \Rightarrow \inf\left\{t \geq 0 : \int_0^t X^2(u)du = 1\right\} = U_1.$$

On the other hand, by the fact $\int_0^{U_1} X^2(u)du = 1$ and $\int_0^{U_1} X(u)dW(u) = M_{U_1} = B_1$ from (26),

$$\begin{aligned}
\sqrt{c}\hat{\phi}^c_{\tau_c} &= \frac{\sum_{n=1}^{\tau_c} x_{n-1}\Delta x_n/\sigma^2\sqrt{c}}{\sum_{n=1}^{\tau_c} x_{n-1}^2/\sigma^2 c} \\
&= \frac{J_c(\tau_c/\sqrt{c})}{F_c(\tau_c/\sqrt{c})} && (53)\\
&\sim \frac{\tilde{J}_c(\tilde{\tau}_c/\sqrt{c})}{\tilde{F}_c(\tilde{\tau}_c/\sqrt{c})} && (54)\\
&\to \frac{\tilde{J}(\tilde{U}_1)}{\tilde{F}(\tilde{U}_1)} \quad a.s. \\
&\sim \frac{\int_0^{U_1} X(u)dX(u)}{\int_0^{U_1} X^2(u)du} && (55)\\
&= \int_0^{U_1} X(u)(-\delta X(u)du + dW(u)) \\
&= -\delta + B_1. && (56)
\end{aligned}$$

□

Lemma 21. *If I_c, F_c, and J_c are defined as in (39),(41), and (42), as for s_N^2 defined in (5) we obtain*

$$s^2_{\lfloor\sqrt{c}t\rfloor} = -\frac{\sigma^2}{\lfloor\sqrt{c}t\rfloor}\left(\delta^2 F_c(t) + \frac{J_c^2(t)}{F_c(t)} + 2\delta J_c(t)\right) + \frac{\sqrt{c}\sigma^2}{\lfloor\sqrt{c}t\rfloor} I_c(t) \tag{57}$$

and for any $t > 0$

$$s^2_{\lfloor\sqrt{c}t\rfloor} \to_p \sigma^2 \quad . \tag{58}$$

Proof. Since

$$\begin{aligned}
\sqrt{c}\hat{\phi}^c_{\lfloor\sqrt{c}t\rfloor} &= \frac{\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\Delta x_n/\sigma^2\sqrt{c}}{\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}^2/\sigma^2 c} \\
&= \frac{J_c(t)}{F_c(t)}
\end{aligned}$$

we have from (45)

$$\begin{aligned}
\frac{\lfloor\sqrt{c}t\rfloor}{\sqrt{c}} \frac{s^2_{\lfloor\sqrt{c}t\rfloor}}{\sigma^2} &= \frac{1}{\sigma^2\sqrt{c}} \sum_{n=1}^{\lfloor\sqrt{c}t\rfloor} \left(\Delta x_n - \hat{\phi}^c_{\lfloor\sqrt{c}t\rfloor} x_{n-1}\right)^2 \\
&= \frac{1}{\sigma^2\sqrt{c}} \sum_{n=1}^{\lfloor\sqrt{c}t\rfloor} \left(-\frac{1}{\sqrt{c}}\left(\delta + \frac{J_c(t)}{F_c(t)}\right) x_{n-1} + \epsilon_n\right)^2 \\
&= \frac{1}{\sqrt{c}}\left(\delta + \frac{J_c(t)}{F_c(t)}\right)^2 F_c(t) - \frac{2}{\sqrt{c}}\left(\delta + \frac{J_c(t)}{F_c(t)}\right)(J_c(t) + \delta F_c(t)) + I_c(t) \\
&= -\frac{1}{\sqrt{c}}\left(\delta^2 F_c(t) + \frac{J_c^2(t)}{F_c(t)} + 2\delta J_c(t)\right) + I_c(t)
\end{aligned}$$

Hence we have (57). Consider Skorohod Theorem as in Lemma 18. For fixed $t > 0$,

$$\frac{1}{\lfloor\sqrt{c}t\rfloor}\left(\delta^2 F_c(t) + \frac{J_c^2(t)}{F_c(t)} + 2\delta J_c(t)\right) \sim \frac{1}{\lfloor\sqrt{c}t\rfloor}\left(\delta^2 \tilde{F}_c(t) + \frac{\tilde{J}_c^2(t)}{\tilde{F}_c(t)} + 2\delta \tilde{J}_c(t)\right)$$

and the right side converges to 0 a.s. and

$$\frac{\sqrt{c}\sigma^2}{\lfloor\sqrt{c}t\rfloor} I_c(t) = \frac{1}{\lfloor\sqrt{c}t\rfloor}\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor}\epsilon_n^2 \to \sigma^2 a.s.$$

Therefore $s^2_{\lfloor\sqrt{c}t\rfloor} \to_p \sigma^2$. □

If we assume the a.s. convergence of $s^2_{\lfloor\sqrt{c}t\rfloor}$, which is the case for $\tilde{s}^2_{\lfloor\sqrt{c}t\rfloor}$ in (60) under the new probability space produced by Skorohod's representation theorem, we have

Lemma 22. *If $s^2_{\lfloor\sqrt{c}t\rfloor} \to \sigma^2 a.s.$ for any $t > 0$, then $\hat{\tau}_c \to \infty \quad a.s.$*

Proof. Fix $\epsilon \in (0, \sigma^2)$ and $t > 0$, choose c_0 so that

$$\sigma^2 - \epsilon \le s^2_{\lfloor\sqrt{c}t\rfloor} \le \sigma^2 + \epsilon,$$

for any $c \ge c_0$. Multiply $c > 0$ to each side, $c(\sigma^2 - \epsilon) \le cs^2_{\lfloor\sqrt{c}t\rfloor} \le c(\sigma^2 + \epsilon)$. Let $c' = c(1 - \epsilon/\sigma^2)$ and $c'' = c(1 + \epsilon/\sigma^2)$. Then,

$$c'\sigma^2 \le cs^2_{\lfloor\sqrt{c}t\rfloor} \le c''\sigma^2. \tag{59}$$

Now, for any $m = 1, 2, \ldots, \lfloor\sqrt{c_0}t\rfloor - 1$, let us take a large enough c to get $\sum_{n=1}^m x_{n-1}^2 < cs_m^2$. When m satisfies $\lfloor\sqrt{c_0}t\rfloor \le m < \tau_{c'}$, $\sum_{n=1}^m x_{n-1}^2 < c'\sigma^2 \le cs_m^2$. Hence, for any $m = 1, 2, \ldots, \tau_{c'} - 1$, $\sum_{n=1}^m x_{n-1}^2 < cs_m^2$, which implies $\tau_{c'} \le \hat{\tau}_c$ for large enough c.

For such c, $\lfloor\sqrt{c}_0 t\rfloor < \tau_{c''}$. Hence, $s^2_{\hat{\tau}_c} c \le c''\sigma^2 \le \sum_{n=1}^{\tau_{c''}} x_{n-1}^2$, which gives $\hat{\tau}_c \le \tau_{c''}$.

So we have $\tau_{c'} \le \hat{\tau}_c \le \tau_{c''}$ for large enough c. Since $\lim_{c\uparrow\infty} \tau_c = \infty$ *a.s.*, letting $c \uparrow \infty$, then we have$\tau_{c'} \to \infty$ and $\tau_{c''} \to \infty$ *a.s.*. Henceforce, as $c \uparrow \infty$, $\hat{\tau}_c \to \infty$ *a.s.* □

Now, we conclude the proof of the main theorem (Theorem 8) of $\hat{\tau}_c$ and $\hat{\phi}_{\hat{\tau}_c}$.

Proof. Using Skorohod's representation theorem, we can find the probability space in which the a.s.convergence

$$\left(\tilde{X}_c, \tilde{F}_c, \tilde{J}_c, \tilde{I}_c\right) \to \left(\tilde{X}, \tilde{F}, \tilde{J}, \tilde{I}\right)$$

holds in the sense of $D^4[0, \infty)$, which corresponds to the weak convergence $(X_c, F_c, J_c, I_c) \to (X, F, J, I)$ in Theorem 6 , Lemma 17 and Lemma 18 . From (57), we define

$$\tilde{s}^2_{\lfloor\sqrt{c}t\rfloor} = -\frac{\sigma^2}{\lfloor\sqrt{c}t\rfloor}\left(\delta^2 \tilde{F}_c(t) + \frac{\tilde{J}_c^2(t)}{\tilde{F}_c(t)} + 2\delta \tilde{J}_c(t)\right) + \frac{\sqrt{c}\sigma^2}{\lfloor\sqrt{c}t\rfloor}\tilde{I}_c(t) \tag{60}$$

for any $t > 0$. We certainly have $s^2_{\lfloor\sqrt{c}t\rfloor} \to \sigma^2$ a.s. for any $t > 0$. Using $\tilde{F}_c(t)$ in (44) and define

$$\frac{\tilde{\hat{\tau}}_c}{\sqrt{c}} = \inf\left\{t : \tilde{F}_c(t) \ge \frac{\tilde{s}^2_{\lfloor\sqrt{c}t\rfloor}}{\sigma^2}\right\}. \tag{61}$$

Then by Lemma 22 $\tilde{\hat{\tau}}_c \to \infty$ a.s. Let

$$\tilde{\hat{\tau}}'_c = \inf\left\{\sqrt{c}t \ge 0 : \int_0^t \tilde{X}_c^2(u)du = \frac{\tilde{s}^2_{\lfloor\sqrt{c}t\rfloor}}{\sigma^2}\right\}.$$

Then

$$\left| \frac{\tilde{\hat{\tau}}_c}{\sqrt{c}} - \frac{\tilde{\hat{\tau}}'_c}{\sqrt{c}} \right| \leq \frac{1}{\sqrt{c}}.$$

From Lemma 19 we have

$$\frac{\tilde{\hat{\tau}}'_c}{\sqrt{c}} \to \inf \left\{ t \geq 0 : \int_0^t \tilde{X}^2(u)du = 1 \right\} := \tilde{U}_1.$$

Hence, since $\tilde{\hat{\tau}}_c/\sqrt{c} \sim \hat{\tau}_c/\sqrt{c}$,

$$\frac{\tau_c}{\sqrt{c}} \Rightarrow \inf \left\{ t \geq 0 : \int_0^t X^2(u)du = 1 \right\} = U_1.$$

On the other hand, by the fact $\int_0^{U_1} X^2(u)du = 1$ and $\int_0^{U_1} X(u)dW(u) = M_{U_1} = B_1$ from (26),

$$\begin{aligned}
\sqrt{c}\hat{\phi}^c_{\hat{\tau}_c} &= \frac{\sum_{n=1}^{\hat{\tau}_c} x_{n-1}\Delta x_n/\sigma^2\sqrt{c}}{\sum_{n=1}^{\hat{\tau}_c} x_{n-1}^2/\sigma^2 c} \\
&= \frac{J_c(\hat{\tau}_c/\sqrt{c})}{F_c(\hat{\tau}_c/\sqrt{c})} && (62) \\
&\sim \frac{\tilde{J}_c(\tilde{\hat{\tau}}_c/\sqrt{c})}{\tilde{F}_c(\tilde{\hat{\tau}}_c/\sqrt{c})} && (63) \\
&\to \frac{\tilde{J}(\tilde{U}_1)}{\tilde{F}(\tilde{U}_1)} \quad a.s. \\
&\sim \frac{\int_0^{U_1} X(u)dX(u)}{\int_0^{U_1} X^2(u)du} && (64) \\
&= -\delta + B_1. && (65)
\end{aligned}$$

Finally we obtain the representation of the stopping time U_1 by using the Bessel process with a drift $-\delta$. The inverse function theorem gives $dU_s/ds = 1/X^2_{U_s}$. By Ito's formula,

$$X_u^2 = 2\int_0^u X_t dX_t + u = -2\delta \int_0^u X_t^2 dt + 2\int_0^u X_t dW_t + u. \tag{66}$$

Letting $u = U_s$, we have

$$X_{U_s}^2 = -2\delta \int_0^{U_s} X_t^2 dt + 2\int_0^{U_s} X_t dW_t + U_s = -2\delta s + 2B_s + U_s.$$

Thus

$$\frac{dU_s}{ds} = \frac{1}{-2\delta s + 2B_s + U_s}.$$

Put $\rho_s = X^2_{U_s}/2 = (-2\delta s + 2B_s + U_s)/2$, then we have

$$d\rho_s = (-\delta + \frac{1}{4\rho_s})ds + dB_s. \tag{67}$$

This indicates that ρ_s is the Bessel process of dimension $3/2$ with a drift $-\delta$ and a initial value $X_0^2/2$. Then, we have

$$U_1 = \int_0^1 dU_s = \int_0^1 \frac{1}{-2\delta s + 2B_s + U_s} ds = \int_0^1 \frac{1}{2\rho_s} ds.$$

□

6.2 Proof of Theorem 13

Proof. Put $t = \hat{\tau}_c/\sqrt{c}$ in (57) and by (39), we have

$$\begin{aligned}
c^{1/4}\left(s^2_{\hat{\tau}_c} - \sigma^2\right) &= -\frac{\sigma^2 c^{1/4}}{\hat{\tau}_c}\left(\delta^2 F_c\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right) + \frac{J_c^2\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right)}{F_c\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right)} + 2\delta J_c\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right)\right) + \frac{\sqrt{c}\sigma^2 c^{1/4}}{\hat{\tau}_c} I_c\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right) \\
&= -\frac{\sigma^2 c^{1/4}}{\hat{\tau}_c}\left(\delta^2 F_c\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right) + \frac{J_c^2\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right)}{F_c\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right)} + 2\delta J_c\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right)\right) + \frac{c^{1/4}}{\hat{\tau}_c}\sum_{n=1}^{\hat{\tau}_c}\left(\epsilon_n^2 - \sigma^2\right) \\
&= -\frac{\sigma^2\sqrt{c}}{\hat{\tau}_c}\frac{1}{c^{1/4}}\left(\delta^2 F_c\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right) + \frac{J_c^2\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right)}{F_c\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right)} + 2\delta J_c\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right)\right) + \frac{\sqrt{c}}{\hat{\tau}_c}\frac{1}{c^{1/4}}\sum_{n=1}^{\hat{\tau}_c}\left(\epsilon_n^2 - \sigma^2\right)
\end{aligned}$$

Now, we define

$$W'_c(t) = \frac{1}{c^{1/4}\sigma_{\epsilon^2}}\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor}\left(\epsilon_n^2 - \sigma^2\right) \tag{68}$$

Then, we have

$$W'_c(t) \Rightarrow W'$$

in the sense of $D[0,\infty)$, where W' is a Brownian motion with $\langle W', W\rangle_1 = \mu_3/\sigma_{\epsilon^2}\sigma$.

Using Skorohod's representation theorem, we can find the probability space in which the a.s.convergence

$$\left(\tilde{F}_c, \tilde{J}_c, \tilde{W}'_c\right) \to \left(\tilde{F}, \tilde{J}, \tilde{W}'\right)$$

holds in the sense of $D^3[0,\infty)$, which corresponds to the weak convergence $(F_c, J_c, W'_c) \to (F, J, W')$.

So we have

$$\begin{aligned}
&c^{1/4}\left(s^2_{\hat{\tau}_c} - \sigma^2\right) \\
&= -\frac{\sigma^2\sqrt{c}}{\hat{\tau}_c}\frac{1}{c^{1/4}}\left(\delta^2 F_c\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right) + \frac{J_c^2\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right)}{F_c\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right)} + 2\delta J_c\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right)\right) + \sigma_{\epsilon^2}\frac{\sqrt{c}}{\hat{\tau}_c}W'_c\left(\frac{\hat{\tau}_c}{\sqrt{c}}\right) \\
&\sim -\frac{\sigma^2\sqrt{c}}{\tilde{\hat{\tau}}_c}\frac{1}{c^{1/4}}\left(\delta^2 \tilde{F}_c\left(\frac{\tilde{\hat{\tau}}_c}{\sqrt{c}}\right) + \frac{\tilde{J}_c^2\left(\frac{\tilde{\hat{\tau}}_c}{\sqrt{c}}\right)}{\tilde{F}_c\left(\frac{\tilde{\hat{\tau}}_c}{\sqrt{c}}\right)} + 2\delta \tilde{J}_c\left(\frac{\tilde{\hat{\tau}}_c}{\sqrt{c}}\right)\right) + \sigma_{\epsilon^2}\frac{\sqrt{c}}{\tilde{\hat{\tau}}_c}\tilde{W}'_c\left(\frac{\tilde{\hat{\tau}}_c}{\sqrt{c}}\right) \\
&\to \frac{\sigma_{\epsilon^2}}{\tilde{U}_1}\tilde{W}'\left(\tilde{U}_1\right)
\end{aligned}$$

□

References

1. Andrews, D.W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," Econometrica, Vol. 59, No. 3 (May, 1991), pp. 817-858 .
2. Billingsley, P. (1999) Convergence of Probability Measures, 2nd Ed., John-Wiley and Sons.
3. BOBKOSKI, M. J. (1983). Hypothesis testing in nonstationary time series. Ph.D. dissertation, Univ. Wisconsin-Madison.
4. Borodin, A.N. and P. Salminen (2002). *Handbook of Brownian Motion - Facts and Formulae*, Second Edition, Birkhauser.
5. Cavanagh, C (1985). "Roots Local To Unity", manuscript, Department of Economics, Harvard University.
6. Chan, N.H. (1988) "On The Parameter Inference for Nearly Nonstationary Time Series", American Statistical Association, 83, 857-862
7. CHAN, N. H. and WEI, C. Z. (1987) "Asymptotic Inference For Nearly Nonstationary AR(1) Processes", Ann. Statist. 15, 1050-1063
8. CHAN, N. H. and WEI, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. Ann. Statist. 16 367-401.
9. Chang,Y. and J. Y. Park (2004). "Taking a new Contour: A Novel View on Unit Root Test", *mimeo.*
10. Dickey, D. A. and W. A. Fuller (1979). "Distribution of the estimates for autoregressive time series with a unit root," *J. Amer. Statist. Assoc.* 74, no. 366, part 1, 427–431.
11. Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1953) "Sequential Decision Problems for Processes with Continuous time Parameter. Testing Hypotheses," *Annals of Mathematical Statistics,* Vol. 24, No. 2, pp. 254-26.
12. Elliott, G., T. J. Rothenberg, and J.H. Stock (1996) "Efficient tests for an autoregressive unit root," Econometrica 64, 813–836.
13. Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York.
14. 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, p. 119-142.
15. Galtchouk, L. and Konev, V. (2006). Sequential estimation of the parameters in unstable AR(2), *Sequential Anal.* 25, p. 25-43.
16. Galtchouk, L. and Konev, V. (2010). On asymptotic normality of sequential LS-estimate for unstable autoregressive process AR(2), *J. of Multivariate Anal.* 101, p. 2616-2636.
17. Galtchouk, L. and Konev, V. (2011). On asymptotic normality of sequential LS-estimates of unstable autoregressive processes, *Sequential Anal.* 30, p.117-144 .
18. Lai, T.L. and Siegmund, D. (1977). A nonlinear renewal theory with applications to sequential analysis. I. *Ann. Statist.* 5 946-954.
19. Lai, T.L. and Siegmund, D. (1979). A nonlinear renewal theory with applications to sequential analysis. II. *Ann. Statist.* 7 60-76.
20. Lai, T.L. and D. Siegmund (1983). Fixed accuracy estimation of an autoregressive parameter, *Annals of Statistics* 11, 478-485.
21. Le Cam, L (1986) Asymptotic Methods in Statistical Decision Theory, Springer-Verlag.

22. Linetsky, V. (2004). "The spectral representation of Bessel processes with constant drift: Applications in queueing and finance," *J. Appl. Probab.*, vol. 41, no2, pp. 327-344.

23. McLeish, D. L. (1974) "Dependent Central Limit Theorems and Invariance Principles," The Annals of Probability, Vol. 2, No. 4, 620-628.

24. McLeish, D. L. (1975) "Invariance Principles for Dependent Variables," z. Wahrscheinlichkeitstheorie verw. Gebiete 32, 165-178.

25. McLeish, D. L. (1977) "On the Invariance Principle for Nonstationary Mixingales", The Annals of Probability, Vol. 5, No. 4, 616-621.

26. Nabeya, S. and Tanaka, K. (1988). Asymptotic theory of a test for the constancy of regression coef- ficients against the random walk alternative. Annals of Statistics, 16, 218–235

27. Nagai, K. and C-H. Zhang (2006), "Nonlinear renewal theorems for random walks with perturbations of intermediate order," in *IMS Lecture Notes-Monograph Series, Recent Developments in Nonparametric Inference and Probability*, Institute of Mathematical Statistics, Vol. 50, 164-175.

28. Newey, W. K., and K. D. West (1987) "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," Econometrica, 55, 703-708.

29. Novikov, A. A. (2009). "On distribution of First Passage Times and Optimal Stopping of AR(1) Sequences". Theory of Probability and Its application 53: 419-429.

30. Phillips, P.C.B. (1987a), "Time Series Regression with a Unit Root," **Econometrica, Vol. 55, No. 2 (Mar., 1987), pp. 277-301.**

31. Phillips, P.C.B. (1987b), "Towards a Unified Asymptotic Theory for Autogression", Biometrika, 71, 535-547

32. Phillips, P.C.B. and Perron, P. (1988). "Testing for a Unit Root in Time Series Regression," Biometrika, Vol. 75, No. 2 (Jun., 1988), pp. 335-346.

33. Revuz, D. and M. Yor (1999). *Continuous Martingale and Brownian Motion*, 3rd ed. Springer-Verlag: New York.

34. Shiryaev, A. N. and V.G. Spokoiny (1997). "On Sequential Estimation of an Autoregressive Parameter," *Stochastics Stochastics Rep.*, 60, 219-240.

35. Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals.* Springer-Verlag, New York.

36. Sorensen, M. (1986). "On Sequential Maximum Likelihood Estimation for Exponential Families of Stochastic Processes," International Statistical Review, Vol. 54, No. 2, pp. 191-210.

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

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

39. Sriram, T.N., I. V. Basawa, R. M. Huggins (1991). "Sequential Estimatin for Branching Processes with Immigration," *Ann. Stat.*, Vol. 19, No. 4, pp. 2232-2243.

40. Sriram, T.N., and Iaci, R. (2014). "Sequential Estimation for Time Series Models," Sequential Analysis, 33: 136–157,

41. Tanaka, K. (1996). *Time series analysis*, John Wiley and Sons.

42. Wald, A. (1947). *Sequential Analysis,* John Wiley and Sons.

43. Wei, C. Z. and J. Winnicki (1990). "Estimation of the Means in the Branching Process with Immigration," *Ann. Stat.*, Vol. 18, No. 4, pp. 1757-1773.

44. Woodroofe, M. (1976). “A renewal theorem for curved boundaries and moments of first passage times”. *Ann. Probab.* 4 67-80.

45. Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis.* CBMS-NSF Regional Conference Series in Applied Mathematics 39, SIAM, Philadelphia.

46. Yashin, A. I. (1984). "On a Problem of Sequential Hypothesis Testing," *Theory of Probability and its Applications,* Volume 28, Issue 1, Pages 157-165.

47. Zhang, C.H. (1988). A nonlinear renewal theory. *Ann. Probab.* 16 793-824.