Sequential test for a unit root in monitoring a p-th order autoregressive process

Kohtaro Hitomi[†], Keiji Nagai[*], Yoshihiko Nishiyama[‡], and Junfan Tao[‡]
[†]Kyoto Instute of Technology, *Yokohama National University, [‡]Kyoto University

October 16, 2022

Abstract

In this study, we investigate methods of sequential analysis to test prospectively for the existence of a unit root against stationary or explosive states in a p-th order autoregressive process monitored over time. Our sequential sampling schemes use stopping times based on the observed Fisher information of a local-to-unity parameter. In contrast to the Dickey-Fuller test statistic, the sequential test statistic has asymptotic normality. We derive the joint limit of the test statistic and the stopping time, which can be characterized using a 3/2-dimensional Bessel process driven by a time-changed Brownian motion. We obtain their limiting joint Laplace transform and density function under the null and local alternatives. In addition, simulations are conducted to show that the theoretical results are valid.[1]

1 Introduction

Unit roots have been an important topic of econometric research since the late 1970s. Time series processes behave quite differently depending on whether a unit root is present or not. White (1958) considered the asymptotic properties of the ordinary least squares estimator (LSE) of an autoregressive (AR) (1) unit root process. Dickey (1976) proposed using it to test the null hypothesis of the unit root against stationarity for AR(1) processes; therefore, it is called the Dickey-Fuller (DF) test. Dickey and Fuller (1979) and Said and Dickey (1984) extended the Dickey-Fuller test to AR(p) and autoregressive moving average (ARMA) models. Instead of estimating the ARMA model, Said and

[1]We are grateful for the helpful comments and suggestions from the editor and an anonymous referee, which have significantly improved the structure and readability of the paper.

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

MSC 2000 subject classifications. Primary 62M10, 62L10, Secondary 60J70, 60F17

Keywords and phrases. Sequential sampling, unit root test, observed Fisher information, DDS Brownian motion, Bessel process, functional central limit theorem in $D[0,\infty)$, local asymptotic normality

Dickey (1984) used the long autoregression model, which does not require the nonlinear optimization used for the ARMA model estimation. This feature of Said and Dickey's test makes it one of the most frequently used unit root tests, and it is often called the augmented Dickey-Fuller (ADF) test. Chang and Park (2002) relaxed the conditions on errors and lag length of AR(p) approximation. The Phillips-Perron test (Phillips (1987a), Phillips and Perron (1988)) is another widely used unit root test that assumes AR(1) processes with general stationary error terms. Chang and Park (2004) and Chang (2012) proposed asymptotically normal unit root tests using a subset of observations for time series and panel data respectively. Many authors have considered local alternatives to examine the statistical properties and performance of the test under a near-unit-root process, for example, Bobkoski (1983), Cavanagh (1985), Chan (1988), Chan and Wei (1987), and Phillips (1987b). There are a vast number of studies related to unit root tests. See, for example, Stock (1994), Choi (2015).

This study considers unit root tests for autoregressive processes of order p, AR(p), under sequential sampling schemes. We are often interested in the existence of a unit root or bubble after a particular event and seek to figure it out as quickly as possible. Asset bubbles are often described as explosive processes. The following is a motivating example. Suppose that the government announces a policy of providing a lump-sum payment to all households. Financial markets have been normal up to now, but there is a possibility of a bubble starting today because of this policy, then fund managers would like to find out whether or not it occurs as soon as possible using data from today onward. In such a monitoring situation, the method of sequential unit root tests proposed in this paper is useful. Policymakers also may want to detect asset bubbles as soon as possible to avoid macroeconomic chaos (see Blanchard (2016)).

Sequential analysis, initiated by Wald (1947), is a prospective study in which statistical inference is made by monitoring incoming observations. The stopping rule is set to collect information from the observed data until precise statistical inferences can be made, and sample size determination and statistical decision making are carried out simultaneously. Therefore, the sample size is a random variable determined by a stopping rule. Especially, sequential sampling schemes are desirable for the quick detection of anomalies.

Lai and Siegmund (1983) considered the sequential estimation of an AR(1) process. Their approach is "purely sequential." Purely sequential analysis is a prospective study that makes no assumptions about existing observations, monitors arrivals of future observations, and determine a sample size only by a stopping rule using the information obtained from the incoming observations. This study also deals with a purely sequential sampling method.

Lai and Siegmund (1983) proposed to continue sampling the AR(1) process $x_n = \beta x_{n-1} + \epsilon_n$ with $\epsilon_n \sim i.i.d.(0, \sigma^2)$ until the observed Fisher information reaches level c. That is, for any $c > 0$, Lai and Siegmund's stopping time N_c is defined as

$$N_c = \inf \{m > 1 : I_m \geq c\} \tag{1}$$

where $I_m = \sum_{n=2}^{m} x_{n-1}^2/\sigma^2$ is the observed Fisher information of β under normal disturbances. Then, at the stopping time N_c, we have a sample $(x_1, \cdots, x_{N_c})$, and I_{N_c}, computed from the sample, is close to c. They showed that if $|\beta| < 1$, $N_c/c \to 1 - \beta^2$ and if $|\beta| = 1$,

$$N_c/\sqrt{c} \Rightarrow \inf \left\{ t : \int_0^t W_s^2 ds = 1 \right\} \tag{2}$$

where W_s is a standard Brownian motion and $\Rightarrow$ indicates weak convergence. In ordinary inference with a non-random sample size, asymptotic theory is constructed by increasing the sample size. In sequential analysis, since the sample size is the stopping time, the asymptotic theory is derived by increasing c and, accordingly, increasing the random stopping time N_c. The sequential estimator $\hat{\beta}_{N_c} = \sum_{n=2}^{N_c} x_n x_{n-1} / \sum_{n=2}^{N_c} x_{n-1}^2$ possesses the uniform asymptotic normality under some assumptions:

$$\sqrt{c}(\hat{\beta}_{N_c} - \beta) \xrightarrow{d} N(0,1) \quad \text{uniformly for } |\beta| \leq 1 \text{ as } c \to \infty. \tag{3}$$

The standard error of the sequential estimator $\hat{\beta}_{N_c}$ is $1/\sqrt{c}$, which implies that c represents the accuracy of the sequential estimation. Therefore, the sequential estimation using a stopping rule based on the observed Fisher information is called a "fixed accuracy estimation."

Since σ^2 is unknown in practice, Lai and Siegmund (1983) defined a feasible stopping time in the following way. Replacing σ^2 with the estimator $s_m^2 = \sum_{n=2}^{m} (x_n - \hat{\beta}_m x_{n-1})^2/(m-2)$ in (1), one gets

$$\hat{N}_c = \inf \left\{ m > 2 : \hat{I}_m \geq c \right\}. \tag{4}$$

where $\hat{I}_m = \sum_{n=2}^{m} x_{n-1}^2 / s_m^2$ is the feasible observed Fisher Information. Here, every time a new observation x_m arrives at time m, the estimators $\hat{\beta}_m$ and s_m^2 are recomputed and the feasible observed Fisher information $\hat{I}_m$ is updated. Each time, $\hat{I}_m$ is compared with c, and $\hat{N}_c$ is the first time that $\hat{I}_m$ exceeds c.

Konev and Pergamenshchikov (1986) and Mukhopadhyay and Sriram (1992) examined the stochastic properties of sequential estimators of stationary vector AR(1) processes. Sriram (2001) proposed a risk-efficient estimator for a stationary threshold AR(1) model (see also Sriram and Iaci (2014)). Galtchouk and Konev (2003, 2004, 2005) consider the sequential estimation of stationary AR parameters. Galtchouk and Konev (2006, 2011) proved the uniform joint normality of the AR(2) and AR(p) parameter estimators, respectively, treating both stationary and nonstationary cases. Hitomi, Nagai, Nishiyama, and Tao (2020) considered sequential unit root tests for AR(1) processes and some of their results are used in this study.

The basic idea of our test is to use the ADF test for an AR(p) process under a sequential sampling scheme. For this purpose, we propose a sequential unit root test using a stopping time based on the observed Fisher information. We stop sampling if we obtain enough observed Fisher information, then perform a unit root test. We derive the joint limit of the test statistics and the stopping time, which can be characterized using a 3/2-dimensional Bessel process driven by a time-changed Brownian motion. To the best of our knowledge, previous research in this field only obtained the marginal limits. We obtain their limiting joint Laplace transform and density function under the null and local alternatives. To obtain these results, we use a diffusion approximation on $D[0,\infty)$ and a time-changed Brownian motion.

The remainder of this paper is organized as follows. Section 2 describes the models, stopping times, and test statistic. The joint asymptotic distribution of the estimator and the stopping time are included in Section 3. In section 4 we present the testing procedure and compare the theoretical values of size, power and average sample size with the values obtained from Monte Carlo simulations. We also discuss simulations using estimated models with different lag lengths than the true model. All the proofs are supplied in the online material, because of the space constraint.

2 Sequential test for near-unit-root AR(p) process

Consider a near-unit-root AR(p) process $\{x_n\}$ under a complete probability space $(\Omega, \mathcal{F}, P)$ with root α_c near unity and initial values $x_1, \ldots, x_p \in L^2$;

$$\alpha_c = 1 + \delta/\sqrt{c}, \qquad \delta \in (-\infty, \infty), \quad c > 0 \tag{5}$$

$$(1 - \alpha_c L)\, \Psi(L) x_n = \epsilon_n, \qquad n = p+1, p+2, \ldots, \tag{6}$$

where δ is a local parameter, $\epsilon_n \sim i.i.d.(0, \sigma^2)$ are independent of $x_1, \ldots, x_p$, and $\Psi(L)$ is a polynomial in the lag operator L;

$$\Psi(L) = 1 - \psi_1 L - \cdots - \psi_{p-1} L^{p-1} = (1 - \alpha_1 L) \cdots (1 - \alpha_{p-1} L). \tag{7}$$

We assume that the roots of $\Psi(z)$ lie outside the closed unit disc and

$$|\alpha_i| < |\alpha_c| \leq 1 \quad \text{or} \quad |\alpha_i| < 1 \leq |\alpha_c| \quad (i = 1, \ldots, p-1). \tag{8}$$

In this study, we investigate the asymptotic properties of the sequential procedures for testing

$$H_0 : \delta = 0 \quad \text{vs} \quad H_1 : \delta > 0 \text{ or } H_1 : \delta < 0. \tag{9}$$

Once ϵ_n are assumed to be $\epsilon_n \sim i.i.d.N(0, \sigma^2)$, the observed Fisher information of α_c is identified as $I_m^{(p)} = \sum_{n=p+1}^{m} \left(\Psi(L) x_{n-1}\right)^2 / \sigma^2$. Using this information, the stopping time is defined as

$$\tau_c = \inf \left\{ m > p : I_m^{(p)} \geq c \right\} \tag{10}$$

which is reduced to the stopping time (2) defined by Lai and Siegmund (1983) when $p = 1$. Thus the stopping time τ_c is the first time at which the observed

Fisher information reaches level c. Unlike Chang and Park (2004) and Chang (2012), our approach is a purely sequential prospective study in the sense that we do not set a maximum sample size and do not restrict the possible values of the stopping time. In ordinary inference with a non-random sample size N, asymptotic theory is constructed by increasing N. In sequential analysis, the counterpart to the sample size is τ_c, and thus the asymptotic theory is constructed by increasing c and, accordingly, increasing the random stopping time τ_c. Replacing the parameters with their estimators, a stopping time using the feasible observed Fisher information corresponding to (4) is defined as

$$\hat{\tau}_c = \inf\left\{m > p+1 : \hat{I}_m^{(p)} \geq c\right\} \tag{11}$$

where $\hat{I}_m^{(p)} = \sum_{n=p+1}^{m}\left(\hat{\Psi}_m(L)x_{n-1}\right)^2/s_m^2$ with the estimators of $\Psi(L)$ and σ^2 ;

$$\hat{\Psi}_m(L) = 1 - \hat{\psi}_{m,1}L - \cdots - \hat{\psi}_{m,p-1}L^{p-1} \tag{12}$$

using the least square estimators $\hat{\psi}_{m,1}, \ldots, \hat{\psi}_{m,p-1}$ defined in (25) and the estimator s_m^2 defined in (24). Here, when a new observation x_m arrives at time m, the estimators $\hat{\Psi}_m(L)$ and s_m^2 are recomputed and the feasible observed Fisher information $\hat{I}_m^{(p)}$ is updated. Each time, $\hat{I}_m^{(p)}$ is compared with c, and $\hat{\tau}_c$ is defined as the first time that $\hat{I}_m^{(p)}$exceeds c. Note that τ_c or $\hat{\tau}_c$ is thought to be an "augmented" version of the stopping time introduced by Lai and Siegmund (1983) for an AR(1) model. We will see that τ_c and $\hat{\tau}_c$ determine the accuracy of inference for the detection of a unit root.

Figure 1 visualizes the procedures of the sequential sampling and stopping rule by flowcharts. (A) presents the case when we know the parameter values $\psi_1, \cdots, \psi_{p-1}, \sigma^2$, whereas (B) includes estimation of the nuisance parameters.

Note that the expression of the local parameter is arbitrary in the sense that one can set $\alpha_c = 1 + \delta/c^\eta$ for any $\eta > 0$. It determines the formulation of the functional central limit theorem. For example, in the AR(1) model $x_n = \alpha_c x_{n-1} + \epsilon_n$, the functional central limit is written as follows:

$$\frac{1}{c^{\eta/2}\sigma}\sum_{n=2}^{\lfloor c^\eta t\rfloor}\epsilon_n \Rightarrow W_t \qquad \text{as} \quad c \to \infty.$$

Then, one should define the stopping time as

$$\tau_c = \inf\left\{m > 1 : I_m \geq c^{2\eta}\right\}$$

where $I_m = \sum_{n=2}^{m} x_{n-1}^2/\sigma^2$ is the observed Fisher information. As shown in the same way as Lai and Siegmund (1983) or our Theorem 2, the asymptotic property of τ_c under the null hypothesis is reduced to be

$$\tau_c/c^\eta \Rightarrow \inf\left\{t : \int_0^t W_s^2 ds = 1\right\}.$$

To align our stopping time formulation with (1) of Lai and Sigmund (1983), we have chosen $\eta = 1/2$.

2.1 A reparameterization of regression for near-unit-root AR(p) process

In this subsection we will obtain a suitable representation for the regression analysis of the near-unit-root AR(p). Defining the difference operator as $\Delta = 1 - L$, we can write (6) as

$$\Psi(L)\left(\Delta x_n - \frac{\delta}{\sqrt{c}} x_{n-1}\right) = \epsilon_n$$

and obtain the difference equation of the AR(p) process x_n in (6);

$$\Delta x_n - \sum_{i=1}^{p-1} \psi_i \Delta x_{n-i} - \frac{\delta}{\sqrt{c}}\left(x_{n-1} - \sum_{i=1}^{p-1} \psi_i x_{n-i-1}\right) = \epsilon_n. \tag{13}$$

Using a telescoping relation $x_{n-j-1} = x_{n-1} - \sum_{i=1}^{j} \Delta x_{n-i}$, we have

$$\sum_{j=1}^{p-1} \psi_j x_{n-j-1} \quad = \quad \sum_{j=1}^{p-1} \psi_j x_{n-1} - \sum_{i=1}^{p-1} \sum_{j=i}^{p-1} \psi_j \Delta x_{n-i}.$$

Then (13) can be represented as follows:

$$\Delta x_n - \frac{\delta}{\sqrt{c}} \Psi(1) x_{n-1} - \sum_{i=1}^{p-1}\left(\psi_i + \frac{\delta}{\sqrt{c}} \sum_{j=i}^{p-1} \psi_j\right) \Delta x_{n-i} = \epsilon_n, \tag{14}$$

where $\Psi(1) = 1 - \psi_1 - \cdots - \psi_{p-1}$.

Now we can reparameterize a regression model for a near-unit-root AR(p) process. We use $'$ for the matrix transpose. Put $\boldsymbol{\phi}_2^c = \left(\phi_2^c, \phi_3^c, \cdots, \phi_p^c\right)'$, $\boldsymbol{\psi} = (\psi_1, \psi_2, \cdots, \psi_{p-1})'$, and $(p-1) \times (p-1)$ matrix

$$A_c = \begin{pmatrix} \alpha_c & \alpha_c - 1 & \cdots & \alpha_c - 1 & \alpha_c - 1 \\ 0 & \alpha_c & \cdots & \alpha_c - 1 & \alpha_c - 1 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \alpha_c & \alpha_c - 1 \\ 0 & 0 & \cdots & 0 & \alpha_c \end{pmatrix}. \tag{15}$$

Let the lag polynomial of (6) be $\Phi^c(\lambda) = (1 - \alpha_c \lambda)\, \Psi(\lambda)$ and write

$$\Phi^c(\lambda) = (1 - \lambda) - \phi_1^c \lambda - \phi_2^c (1 - \lambda) \lambda - \cdots - \phi_p^c (1 - \lambda) \lambda^{p-1}, \tag{16}$$

where

$$\phi_1^c = \frac{\delta}{\sqrt{c}} \Psi(1), \quad \boldsymbol{\phi_2^c} = A_c \boldsymbol{\psi}. \tag{17}$$

The model (6) becomes

$$\Phi^c(L) x_n = \Delta x_n - \phi_1^c x_{n-1} - \phi_2^c \Delta x_{n-1} - \cdots - \phi_p^c \Delta x_{n-p+1} = \epsilon_n. \tag{18}$$

2.2 Normal equation and sequential estimator

For the sample $x_1, \ldots, x_m$, the vector representation of the AR(p) in (18) is

$$\boldsymbol{\Delta x_m} - \boldsymbol{X_m}\boldsymbol{\phi^c} = \boldsymbol{\epsilon_m}, \tag{19}$$

where $\boldsymbol{\Delta x_m} = (\Delta x_{p+1}, \Delta x_{p+2}, \cdots, \Delta x_m)'$, $\boldsymbol{\phi^c} = \left(\ \phi^c_1, \phi^c_2, \cdots, \phi^c_p \ \right)'$, $\boldsymbol{\epsilon_m} = (\epsilon_{p+1}, \epsilon_{p+2}, \cdots, \epsilon_m)'$, and the $(m-p) \times p$ matrix

$$\boldsymbol{X_m} = (x_{n-1}, \Delta x_{n-1}, \cdots, \Delta x_{n-p+1})_{n=p+1,\ldots,m}.$$

The normal equation for the least square estimator (LSE) $\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\boldsymbol{m}} = \left(\ \hat{\phi}^c_{1,m}, \hat{\phi}^c_{2,m}, \cdots, \hat{\phi}^c_{p,m} \ \right)'$ is

$$\boldsymbol{X'_m}\boldsymbol{X_m}\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\boldsymbol{m}} = \boldsymbol{X'_m}\boldsymbol{\Delta x_m}, \tag{20}$$

which has the form

$$\begin{pmatrix} \displaystyle\sum_{n=p+1}^{m} x^2_{n-1} & \left(\displaystyle\sum_{n=p+1}^{m} x_{n-1}\Delta x_{n-j}\right)_j \\ \left(\displaystyle\sum_{n=p+1}^{m} x_{n-1}\Delta x_{n-i}\right)'_i & \left(\displaystyle\sum_{n=p+1}^{m} \Delta x_{n-i}\Delta x_{n-j}\right)_{i,j} \end{pmatrix} \hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\boldsymbol{m}} = \begin{pmatrix} \displaystyle\sum_{n=p+1}^{m} x_{n-1}\Delta x_n \\ \left(\displaystyle\sum_{n=p+1}^{m} \Delta x_{n-i}\Delta x_n\right)'_i \end{pmatrix} \tag{21}$$

where i, j run over $1, 2, \ldots, p-1$. For the estimator of the lag polynomial $\Phi^c(\lambda)$ in (16), let

$$\hat{\Phi}^c_m(\lambda) = (1-\lambda) - \hat{\phi}^c_{1,m}\lambda - \hat{\phi}^c_{2,m}(1-\lambda)\lambda - \cdots - \hat{\phi}^c_{p,m}(1-\lambda)\lambda^{p-1} \tag{22}$$

and $\hat{\alpha}_{c,m}$ be the LSE of AR(1) model $x_n = \alpha_c x_{n-1} + v_n$;

$$\hat{\alpha}_{c,m} = \sum_{n=p+1}^{m} x_{n-1}x_n \Big/ \sum_{n=p+1}^{m} x^2_{n-1} \tag{23}$$

where v_n is defined in (27).

As to the stopping time $\hat{\tau}_c$ in (11), we obtain the sequential estimator $\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\hat{\boldsymbol{\tau}}_{\boldsymbol{c}}}$ and the sequential unit root test statistics $\hat{\delta}_{\hat{\tau}_c}$ in the following manner. Put

$$s^2_m = \left(\boldsymbol{\Delta x_m} - \boldsymbol{X_m}\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\boldsymbol{m}}\right)'\left(\boldsymbol{\Delta x_m} - \boldsymbol{X_m}\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\boldsymbol{m}}\right)/(m-p) \tag{24}$$

for the consistent estimator of σ^2. In view of (17), let

$$\hat{\boldsymbol{\psi}}_{\boldsymbol{m}} = \hat{A}^{-1}_{c,m}\hat{\boldsymbol{\phi}}^{c}_{\boldsymbol{2,m}}, \tag{25}$$

where $\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\boldsymbol{2,m}} = \left(\ \hat{\phi}^c_{2,m}, \hat{\phi}^c_{3,m}, \cdots, \hat{\phi}^c_{p,m} \ \right)'$ and $\hat{A}_{c,m}$ is designated by replacing α_c with $\hat{\alpha}_{c,m}$ in A_c defined in (15).

Letting $\hat{\Psi}_m(1) = 1 - \hat{\psi}_{1,m} - \cdots - \hat{\psi}_{p-1,m}$, we can set $\hat{\tau}_c$ in (11) and obtain the sequential estimators, $\hat{\Psi}_{\hat{\tau}_c}(1)$ and $\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\hat{\boldsymbol{\tau}}_{\boldsymbol{c}}}$, and the sequential unit root test statistic,

$$\hat{\delta}_{\hat{\tau}_c} = \frac{\sqrt{c}\hat{\phi}^c_{1,\hat{\tau}_c}}{\hat{\Psi}_{\hat{\tau}_c}(1)}. \tag{26}$$

3 The asymptotic properties of $\hat{\tau}_c$ and the sequential procedures for a near-unit-root AR(p)

In this section, we consider the asymptotic properties of the stopping times $\hat{\tau}_c$ in (11), the sequential unit root test statistics $\hat{\delta}_{\hat{\tau}_c}$, and the sequential estimator $\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\hat{\boldsymbol{\tau}}_{\boldsymbol{c}}}$ defined in (20) when c goes to ∞. To deal with asymptotic results relating to the stopping times, we use the weak convergence theorem in $D[0,\infty)$ and Skorohod's representation theorem (Billingsley (1999), Theorem 6.7). For brief descriptions, see Hitomi, Nagai, Nishiyama, and Tao (2020).

Let

$$z_n = \Psi(L)x_n, \quad \text{and} \quad v_n = (1-\alpha_c L)\,x_n. \tag{27}$$

Then z_n is a near-unit-root AR(1) process with initial value $z_p = x_p - \psi_1 x_{p-1} - \cdots - \psi_{p-1}x_1$ and v_n is an asymptotically stationary AR($p-1$) process with initial values $v_i = x_i - \alpha_c x_{i-1}$ ($i = 2, \ldots p$). Using them, write

$$\Delta z_n = \frac{\delta}{\sqrt{c}} z_{n-1} + \epsilon_n, \quad \text{and} \quad \Psi(L)v_n = \epsilon_n, \qquad n = p+1, p+2, \ldots. \tag{28}$$

3.1 Convergence to an Ornstein-Uhlenbeck process on $D[0,\infty)$

For z_n defined in (27), let

$$Z_c(t) = z_{\lfloor \sqrt{c}t \rfloor}/c^{1/4}\sigma. \tag{29}$$

Let $S_n = \epsilon_{p+1} + \epsilon_{p+2} + \cdots + \epsilon_n$ with $S_1 = \cdots = S_{p-1} = 0$, W be a Brownian motion, and

$$W_c(t) = S_{\lfloor \sqrt{c}t \rfloor}/c^{1/4}\sigma. \tag{30}$$

Then, by the functional central limit theorem (FCLT) (Billingsley (1999), Theorem 18.2), we have as $c \uparrow \infty$

$$W_c \Rightarrow W \quad \text{in} \quad D[0,\infty) \tag{31}$$

where $\Rightarrow$ stands for weak convergence and $D[0,\infty)$ is the space of right continuous functions with left limits on $[0,\infty)$. We then obtain

$$Z_c \Rightarrow Z \quad \text{in} \quad D[0,\infty). \tag{32}$$

Here Z is the Ornstein-Uhlenbeck (OU) process with the Brownian motion W in (31):

$$dZ_t = \delta Z_t dt + dW_t. \tag{33}$$

The test hypothesis $H_0 : \delta = 0$ vs $H_1 : \delta \neq 0$ under the OU process Z becomes

$$H_0 : dZ_t = dW_t \quad \text{vs} \quad H_1 : dZ_t = \delta Z_t dt + dW_t.$$

3.2 Asymptotic properties of the components of the normal equation

Next, we consider a functional limit theorem in $D[0,\infty)$ with respect to the normal equation (20) and (21) with proper normalization. Let $\boldsymbol{D_c}$ be the $p \times p$ matrix $diag\left(\sqrt{c}, c^{1/4}, \ldots, c^{1/4}\right)$. Multiplying the normal equation by $\boldsymbol{D_c^{-1}}$, we have

$$\boldsymbol{D_c^{-1}} X'_{\lfloor\sqrt{c}t\rfloor} X_{\lfloor\sqrt{c}t\rfloor} \boldsymbol{D_c^{-1}} \boldsymbol{D_c} \left(\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\lfloor\sqrt{\boldsymbol{c}}\boldsymbol{t}\rfloor} - \boldsymbol{\phi^c}\right) = \boldsymbol{D_c^{-1}} X'_{\lfloor\sqrt{c}t\rfloor} \boldsymbol{\epsilon}_{\lfloor\sqrt{\boldsymbol{c}}\boldsymbol{t}\rfloor} \tag{34}$$

which has the form

$$\begin{pmatrix} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}^2/c & \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\Delta x_{n-j}/c^{3/4}\right)_j \\ \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\Delta x_{n-i}/c^{3/4}\right)'_i & \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \Delta x_{n-i}\Delta x_{n-j}/\sqrt{c}\right)_{i,j} \end{pmatrix} \begin{pmatrix} \sqrt{c}(\hat{\phi}^c_{1,\lfloor\sqrt{c}t\rfloor} - \phi^c_1) \\ c^{1/4}\left(\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\boldsymbol{2},\lfloor\sqrt{\boldsymbol{c}}\boldsymbol{t}\rfloor} - \boldsymbol{\phi}^{\boldsymbol{c}}_{\boldsymbol{2}}\right) \end{pmatrix}$$
$$= \begin{pmatrix} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\epsilon_n/\sqrt{c} \\ \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \Delta x_{n-i}\epsilon_n/c^{1/4}\right)'_i \end{pmatrix} \quad (i,j = 1,\ldots,p-1)\,. \tag{35}$$

The next proposition shows the main convergence results in (35).

Proposition 1. *The asymptotic autocovariance function $\gamma_v(h)$ of v_n exists for $h = 0, 1, \ldots$;*

$$\lim_{m\to\infty} \sum_{n=p+1}^{m} v_n v_{n+h}/\left(m-p\right) = \gamma_v(h) \ a.s., \tag{36}$$

and the FCLT with respect to martingale differences $v_{n-i}\epsilon_n$ for $i = 1,\ldots,p-1$ holds; in the sense of $D[0,\infty)$,

$$\left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} v_{n-i}\epsilon_n/\left(c^{1/4}\sqrt{\gamma_v(0)}\sigma\right)\right)_i \Rightarrow \boldsymbol{W}^{(1)}_t \quad \text{as} \quad c\to\infty, \tag{37}$$

where $\boldsymbol{W}^{(1)}$ is a $(p-1)$-dimensional Brownian motion with the correlation coefficient matrix $\boldsymbol{\rho} = \left(\gamma_v(|i-j|)/\gamma_v(0)\right)_{i,j=1,\ldots,p-1}$ independent of W in (31).

In the sense of $D[0,\infty)$, as $c\to\infty$,

$$\begin{pmatrix} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}^2/\left(c\sigma^2\right) \\ \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\Delta x_{n-i}/\left(c^{3/4}\sigma^2\right)\right)_i' \\ \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \Delta x_{n-i}\Delta x_{n-j}/\left(\sqrt{c}\sigma^2\right)\right)_{i,j} \\ \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\epsilon_n/\left(\sqrt{c}\sigma^2\right) \\ \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \Delta x_{n-i}\epsilon_n/\left(c^{1/4}\sqrt{\gamma_v(0)}\sigma\right)\right)_i' \end{pmatrix} \Rightarrow \begin{pmatrix} \int_0^t Z_u^2du/\Psi(1)^2 \\ \mathbf{0} \\ \left(\gamma_v\left(|i-j|\right)t\right)_{i,j} \\ \int_0^t Z_udW_u/\Psi(1) \\ \boldsymbol{W}_t^{(1)} \end{pmatrix}, \tag{38}$$

where Z is the OU process in (33).

3.3 Main theorems

Since the stopping time τ_c in (10) has the deformation $\tau_c = \inf\left\{\sqrt{c}t > p : I^{(p)}_{\lfloor\sqrt{c}t\rfloor} \geq c\right\}$ following asymptotic property; as $c\to\infty$,

$$\tau_c/\sqrt{c} = \inf\left\{t > p/\sqrt{c} : \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}^2/c\sigma^2 \geq 1\right\} \Rightarrow \inf\left\{t \geq 0 : \int_0^t Z_u^2du = 1\right\},$$

which is a similar result as Lai and Siegmund's (1983) obtained for the stopping time N_c in (1). Therefore, we modify the expression $\hat{\tau}_c$ as follows for large c:

$$\hat{\tau}_c/\sqrt{c} = \inf\left\{t > 0 : \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(\hat{\Psi}_{\lfloor\sqrt{c}t\rfloor}(L)x_{n-1}\right)^2/cs^2_{\lfloor\sqrt{c}t\rfloor} \geq 1\right\} \tag{39}$$

where $\hat{\Psi}_m(L) = 1 - \hat{\psi}_{1,m}L - \cdots - \hat{\psi}_{p-1,m}L^{p-1}$ and $\hat{\psi}_{i,m}$ are defined in (25). Letting $t = \hat{\tau}_c/\sqrt{c}$ in Proposition 1, the main theorem is obtained. The proofs of Proposition 1 and Theorem 2 are given in the online Appendices 6.1 and 6.2.

Theorem 2. *Consider the stopping time $\hat{\tau}_c$ in (11), the sequential estimator $\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\hat{\boldsymbol{\tau}}_{\boldsymbol{c}}}$ in (20) and $\hat{\boldsymbol{\psi}}_{\hat{\boldsymbol{\tau}}_{\boldsymbol{c}}}$ in (25) by letting $m = \hat{\tau}_c$. Let Z be the OU process as defined in (33) with $Z_0 = 0$. Define the martingale M_t and its quadratic variation $\langle M\rangle_t$ as*

$$M_t = \int_0^t Z_udW_u \quad \text{and} \quad \langle M\rangle_t = \int_0^t Z_u^2du. \tag{40}$$

Then, we define $U_s = \langle M\rangle_s^{-1}$,

$$U_s = \inf\left\{t \geq 0 : \int_0^t Z_u^2du = s\right\}, \tag{41}$$

and the time-changed (Dambis-Dubins-Schwarz) Brownian motion

$$B_s = M_{U_s}. \tag{42}$$

Let $\rho_s = Z_{U_s}^2/2$, *then* ρ_t *becomes a* $3/2$*-dimensional Bessel process with drift* δ *and initial value* $\rho_0 = 0$

$$\rho_t = \rho_0 + B_s + \int_0^t \left(\frac{1}{4\rho_s} + \delta\right) ds \quad \text{and} \quad U_t = \int_0^t \frac{1}{2\rho_s} ds. \tag{43}$$

The asymptotic behavior of the stopping time $\hat{\tau}_c$ *and the sequential unit root test statistics* $\hat{\delta}_{\hat{\tau}_c}$ *are as follows. As* $c \to \infty$, $s_{\hat{\tau}_c}^2 \to_p \sigma^2$,

$$\left(\hat{\delta}_{\hat{\tau}_c}, \hat{\tau}_c/\sqrt{c}\right) \Rightarrow \left(\delta + \int_0^{U_1} Z_u dW_u, U_1\right) = \left(\delta + B_1, \int_0^1 \frac{1}{2\rho_s} ds\right), \tag{44}$$

$$\left(\ \sqrt{\hat{\tau}_c}\left(\hat{\boldsymbol{\phi}}_{\mathbf{2},\hat{\boldsymbol{\tau}}_{\mathbf{c}}}^{\mathbf{c}} - \boldsymbol{\phi}_{\mathbf{2}}^{\mathbf{c}}\right)\ \right) = \left(\ \sqrt{\hat{\tau}_c}\left(\hat{\boldsymbol{\psi}}_{\hat{\boldsymbol{\tau}}_{\mathbf{c}}} - \boldsymbol{\psi}\right)\ \right) + o_p(1) \Rightarrow \left(\ \frac{1}{\sigma\sqrt{\gamma_v(0)}}\boldsymbol{\rho}^{-1}\boldsymbol{W}_1^{(1)}\ \right) \tag{45}$$

where $\boldsymbol{W}^{(1)}$ *is a* $(p-1)$*-dimensional Brownian motion independent of* W, *with the correlation coefficient matrix* $\boldsymbol{\rho} = (\gamma_v(|i-j|)/\gamma_v(0))_{i,j=1,\ldots,p-1}$.

Note that, since $\boldsymbol{W}^{(1)}$ is independent of W, the sequential unit root test statistic $\hat{\delta}_{\hat{\tau}_c}$ and the coefficient estimator $\hat{\boldsymbol{\psi}}_{\hat{\boldsymbol{\tau}}_{\mathbf{c}}}$ of the stationary AR process v_n written in (28) are independent.

The following two theorems give the joint density of the sequential estimators and the stopping time. See the online Appendix 6.3 for the proof.

Theorem 3. *Let* B_v *be a standard Brownian motion,* ρ_v *be a Bessel process defined as (43) with* $\rho_0 = \delta = 0$, *and* $U_v = \int_0^v \frac{1}{2\rho_s} ds$. *Then, the joint density of* (B_v, U_v) *under the null hypothesis has the following form for* $u > 0$, $z > -u/2$:

$$P_{H_0}\left(B_v \in dz, U_v \in du\right) = 2\sqrt{\frac{2z+u}{2\pi}} \mathrm{es}_v\left(\frac{1}{2}, \frac{1}{2}, u, 0, \frac{2z+u}{2}\right) dz du, \tag{46}$$

where es_v *is defined by*

$$\begin{aligned} \mathrm{es}_v\left(\mu, \nu, t, x, z\right) &= \mathscr{L}_\gamma^{-1}\left(\left(\frac{(2\gamma)^{\mu/2}}{\sinh^\nu\left(t\sqrt{2\gamma}\right)}\right) \exp\left(-x\sqrt{2\gamma} - z\sqrt{2\gamma}\coth\left(t\sqrt{2\gamma}\right)\right)\right) \\ &= \sum_{k=0}^{\infty} \frac{(-z)^k}{k!} \mathrm{s}_y\left(\mu+k, \nu+k, t, x+z+kt\right) \end{aligned} \tag{47}$$

and

$$\mathrm{s}_y\left(\mu, \nu, t, z\right) = 2^\nu \sum_{l=0}^{\infty} \frac{\Gamma\left(\nu+l\right) \exp\left(-\left(\nu t + z + 2lt\right)^2/4y\right)}{\sqrt{2\pi} y^{1+\mu/2} \Gamma\left(\nu\right) l!} D_{\mu+1}\left(\frac{\nu t + z + 2lt}{\sqrt{y}}\right). \tag{48}$$

Here, $D_{\mu+1}$ *are the parabolic cylinder functions.*

See Borodin and Salminen (2002) on the definitions of es_v, s_y, and $D_{\mu+1}$ for details. Applying Girsanov's theorem to $\delta v + B_v$ in Theorem 3, we derive the explicit form of the joint density as follows. The relation in (49) indicates the LAN (local asymptotic normality) property. See Le Cam (1986).

Theorem 4. *Let B_v be a standard Brownian motion, ρ_v be a Bessel process with drift δ and $\rho_0 = 0$ defined as (43), and $U_v = \int_0^v \frac{1}{2\rho_s} ds$. The joint density of $(\delta v + B_v, U_v)$ under the alternative hypothesis in Theorem 2 becomes*

$$P_{H_1}\left(\delta v + B_v \in dz, U_v \in du\right) = \exp\left(\delta z - \delta^2 v/2\right) P_{H_0}\left(B_v \in dz, U_v \in du\right) \quad (49)$$

where $P_{H_0}\left(B_v \in dz, U_v \in du\right)$ is (46) in Theorem 3.

From the Bessel bridges in Pitman and Yor (1982), we can compute the joint moments of the stopping time and test statistic through the following theorem under the null hypothesis.

Theorem 5. *Let W_t be a standard Brownian motion and ρ_t be a Bessel process defined as (43) with $\rho_0 = \delta = 0$. Using time-change via $t = U_v = \int_0^v \frac{1}{2\rho_s} ds$, the joint Laplace transform of $\left(W_t^2, \int_0^t W_s^2 ds\right)$ can be written as a modified Laplace transform of (ρ_v, U_v);*

$$\begin{aligned} &\int_0^\infty e^{-\beta t} E_{H_0}\left[\exp\left(-\alpha W_t^2 - \gamma \int_0^t W_s^2 ds\right)\right] dt \\ =&\int_0^\infty e^{-\gamma v} E_{H_0}\left[\exp\left(-\alpha 2\rho_v - \beta U_v\right)/2\rho_v\right] dv. \end{aligned} \quad (50)$$

This relation implies

$$\begin{aligned} &E_{H_0}\left[\exp\left(-\alpha 2\rho_v - \beta U_v\right)/2\rho_v\right] \\ =&\mathscr{L}_\gamma{}^{-1}\left[\frac{2^{7/4}\gamma^{1/4}\,{}_2F_1\left(\frac{1}{2}, \frac{1}{4}\left(\frac{\beta\sqrt{2}}{\sqrt{\gamma}} + 1\right); \frac{1}{4}\left(\frac{\beta\sqrt{2}}{\sqrt{\gamma}} + 5\right); \frac{4\alpha}{2\alpha + \sqrt{2}\sqrt{\gamma}} - 1\right)}{\sqrt{2\alpha + \sqrt{2}\sqrt{\gamma}}\left(2\beta + \sqrt{2}\sqrt{\gamma}\right)}\right](v) \quad (51) \\ =&\sum_{k=0}^\infty \sum_{l=0}^\infty \frac{\alpha^k \beta^l}{k! l!} \int_0^1 J(t, v, k, l) dt, \quad (52) \end{aligned}$$

where ${}_2F_1(a, b; c; z) = \sum_{k=0}^\infty (a)_k (b)_k / (c)_k\, z^k/k!$ and

$$J(s, v, k, l) = \frac{2^{\frac{1}{2}(k-3l-2)}\left(-\frac{1}{2}\right)^{(k)} v^{\frac{1}{2}(k+l-1)}}{\Gamma\left(\frac{1}{2}(k+l+1)\right)} s^{-\frac{3}{4}}(1-s)^k(1+s)^{-k-\frac{1}{2}} \log^l(s) \quad (53)$$

with $x^{(m)}$ being the factorial power $x(x-1)\ldots(x-(m-1))$.

The joint Laplace transform under the alternative hypothesis is given as follows.

Theorem 6. *Let ρ_t be a Bessel process defined as (43) with $\rho_0 = 0$ and $\delta \neq 0$. The joint Laplace transform of (ρ_1, U_1) under the alternative hypothesis H_1 becomes*

$$\begin{aligned} &E_{H_1}\left[\exp\left(-\alpha 2\rho_1 - \beta U_1\right)/2\rho_1\right] \\ =&E_{H_0}\left[\exp\left(-\left(\alpha - \frac{\delta}{2}\right)2\rho_1 - \left(\beta + \frac{\delta}{2}\right)U_1 - \frac{\delta^2}{2}\right)/2\rho_1\right] dv \\ =&\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{\alpha^k\beta^l}{k!l!}m(\delta, v, k, l) \end{aligned} \tag{54}$$

where

$$m(\delta, v, k, l) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(-1)^i\frac{(\delta/2)^{i+j}}{i!j!}\int_0^1 J(t, v, i+k, j+l)dt. \tag{55}$$

with $J(t, v, k, l)$ defined in (53).

See the online Appendix 6.4 for the proofs of the above two theorems.

4 Testing procedure, theoretical values, and simulation results

4.1 Testing procedure

The procedures for only the one-sided test against the alternatives of $\delta < 0\,(\alpha_c < 1)$ or $\delta > 0\,(\alpha_c > 1)$ are conducted. Let α be a significance level, and select c. The sequential unit root test employs the asymptotic normality of $\hat{\delta}_{\hat{\tau}_c}$ as shown in Theorem 2 as $c \to \infty$. Since the null hypothesis is $\delta = 0$ and its asymptotic variance equals 1, the sequential unit root test simply looks at $\hat{\delta}_{\hat{\tau}_c} = \sqrt{c}\hat{\phi}^c_{1,\hat{\tau}_c}/\hat{\Psi}_{\hat{\tau}_c}(1)$. Let z_α be the α quantile of the standard normal distribution. With the stationary alternative $\delta{<}0$, we reject the null hypothesis when

$$\hat{\delta}_{\hat{\tau}_c} < z_\alpha$$

for the left tailed test. With the explosive alternative $\delta > 0$, we reject the null hypothesis when

$$\hat{\delta}_{\hat{\tau}_c} > z_{1-\alpha}$$

for the right tailed test.

In practice, $m > p + 1$ in the definition of $\hat{\tau}_c$ in (11) does not work well because the sampling may stop too early for m close to $p + 1$, which will lead to unstable parameter estimation. Replacing it by $m > m_0$ for some suitably large $m_0\,(> p + 1)$, we set the stopping time as

$$\hat{\tau}_c = \inf\left\{m > m_0 : \hat{I}^{(p)}_m \geq c\right\}. \tag{56}$$

Table 1: Theoretical expected values and standard deviations of $\hat{\tau}_c$ under null

$\sqrt{c}$	$E(\hat{\tau}_c) \approx \sqrt{c}E(U_1)$	$sd(\hat{\tau}_c) \approx \sqrt{c}sd(U_1)$
50	104.6	54.4
100	209.2	108.8
150	313.8	163.1
200	418.4	217.5

We should choose m_0 large enough to stabilize the forecast $\hat{I}_m^{(p)}$ of the observed Fisher information. Thereby, we see that the simulation results are consistent with the theory of this study, even when c is relatively small.

4.2 Theoretical values for moments of $\hat{\tau}_c$

Based on the asymptotic property of $\hat{\tau}_c$ provided in Theorem 2, Tables 1 and 2 show the theoretical values of the expectation and standard deviation of $\hat{\tau}_c$ under the null and alternative hypotheses.

The expected value and standard deviation of U_1 under the null hypothesis are calculated using the Laplace transform given in Theorem 5 as follows:

$$E(U_1) = 2.09210 \qquad sd(U_1) = 1.08758.$$

Novikov (1972) also obtained $E(U_1)$ and $E(U_1^2)$. In Table 1, for each value of $\sqrt{c}$, the theoretical values of $E(\hat{\tau}_c)$ and $sd(\hat{\tau}_c)$ are computed from the asymptotic relations $E(\hat{\tau}_c) \approx \sqrt{c} \times E(U_1)$ and $sd(\hat{\tau}_c) \approx \sqrt{c}sd(U_1)$. We immediately see that the average sample size increases as c increases.

Table 2 provides the theoretical values of the expectation and the standard deviation of U_1 under alternative hypotheses. For each value of the local parameter δ in the first column, the second and third columns provide the expected values and standard deviations of U_1 calculated from the Laplace transform given in Theorem 6. In the simulations of the two succeeding sections, we choose the numbers 0.99 and 1.01 for α_c, corresponding to the stationary and explosive alternatives, respectively. In those cases, c is determined from the relation $\alpha = 1 + \delta/\sqrt{c}$. For example, the eighth row gives the case, when $\delta = 1.5$ and $\sqrt{c} = 150$, and thus $\alpha_c = 1 + \delta/\sqrt{c} = 1.01$. Then, the expected value and standard error are $E(\hat{\tau}_c) \approx \sqrt{c}E(U_1) = 150 \times 1.19234 = 178.9$, and $sd(\hat{\tau}_c) \approx \sqrt{c}sd(U_1) = 150 \times 0.58689 = 88.0$ respectively.

Table 2: Theoretical moments of $\hat{\tau}_c$ under alternatives

δ	$E(U_1)$	$sd(U_1)$	$\sqrt{c}$	α_c	$E(\hat{\tau}_c) \approx \sqrt{c}E(U_1)$	$sd(\hat{\tau}_c) \approx \sqrt{c}sd(U_1)$
-0.5	2.59506	1.28939	50	0.99	129.8	64.5
-1	3.21232	1.47454	100		321.2	147.5
-1.5	3.93244	1.62767	150		589.9	244.2
-2	4.73487	1.74208	200		947.0	348.4
0.5	1.70132	0.89481	50	1.01	85.1	44.7
1	1.40833	0.72450	100		140.8	72.4
1.5	1.19234	0.58689	150		178.9	88.0
2	1.03252	0.48170	200		206.5	96.3

4.3 The simulation settings and results when lag length is known

We conducted simulations to examine the performance of the sequential unit root test with the correct lag length. We investigate the sizes and powers of the test and the expected sample sizes and standard deviations of the stopping times. As explained at (3) in the introduction, $1/c$ represents the accuracy of the sequential estimation. As c increases, the stopping time becomes larger, and better estimates of the parameters will be obtained.

The data generating process (DGP) is based on the following AR(3) process

$$(1-\alpha_c L)\,\Psi(L)x_n = \epsilon_n \quad n = 4, 5, \dots \tag{57}$$

where $\Psi(L) = 1 - \psi_1 L - \psi_2 L^2 = (1-\alpha_1 L)(1-\alpha_2 L)$ and $\epsilon_n \sim i.i.d.\, N(0,1)$. The initial values $x_1, x_2, x_3 \sim i.i.d.N(0,1)$ are independent of $\epsilon_4, \epsilon_5, \dots$.

There are 10,000 iterations in the simulation. We set $\alpha_1 = 0.5$, $\alpha_2 = 0.3$, then $\psi_1 = \alpha_1 + \alpha_2 = 0.8$, $\psi_2 = -\alpha_1\alpha_2 = -0.15$. For $\delta < 0\,(\alpha_c < 1)$, we consider the stationary alternative of $\alpha_c = 0.99$. We use the settings $\delta = -0.5, -1, -1.5, -2$ corresponding to $\sqrt{c} = 50, 100, 150, 200$ respectively which are the same as the upper half of Table 2. For the explosive alternative $\delta > 0\,(\alpha_c > 1)$, we set $\alpha_c = 1.01$ and have $\delta = 0.5, 1, 1.5, 2$ for $\sqrt{c} = 50, 100, 150, 200$ from the relationship $\delta = \sqrt{c}(1.01-1)$, respectively. We set $m_0 = 30$ for the definition of $\hat{\tau}_c$ in (56). We also perform simulations with $\sqrt{c} = 800$ to see at which level of $\sqrt{c}$ the size distortion disappears.

Table 3 shows the simulation results along with the theoretical values in parentheses for (i) the value of $\sqrt{c}$, (ii) the sizes of the left and right tailed tests, (iii) the expected value and standard deviation of the stopping time $\hat{\tau}_c$, and (iv) the mean and standard deviation of the sequential estimators $\left(\hat{\alpha}_{c,\hat{\tau}_c}, \hat{\psi}_{1,\hat{\tau}_c}, \hat{\psi}_{2,\hat{\tau}_c}\right)$. Tables 4 and 5 correspond to the alternative hypotheses $\alpha = 0.99$ and 1.01, respectively, and show the results for (i) the values of $\sqrt{c}$ and δ, (ii) the powers of the sequential unit root test, and (iii) and (iv), which are equivalent to those in Table 3. The theoretical values of the mean and standard deviation of $\hat{\tau}_c$ are consulted from the Tables 1 and 2.

Table 3: Sizes of left and right tailed tests and moments under the null

$\alpha_c = 1$	Size (%)		$\hat{\tau}_c$		$\hat{\alpha}_{c,\hat{\tau}_c}$		$\hat{\psi}_{1,\hat{\tau}_c}$		$\hat{\psi}_{2,\hat{\tau}_c}$	
$\sqrt{c}$	Left	Right	Mean	sd	Mean	sd	Mean	sd	Mean	sd
	(5%)	(5%)	($\sqrt{c} \times 2.0921$)	($\sqrt{c} \times 1.0876$)	(1)	($1/\sqrt{c}$)	(0.8)	(0.989)	(-0.15)	(0.989)
50	5.20	12.95	101.9(104.6)	57.4(54.4)	1.019	0.0184	0.763	1.040	-0.170	0.985
100	4.62	9.83	205.1(209.2)	110.8(108.8)	1.009	0.0086	0.782	1.016	-0.161	0.981
150	5.19	8.22	310.0(313.8)	167.4(163.1)	1.006	0.0054	0.788	1.019	-0.156	0.988
200	5.04	7.19	413.9(418.4)	218.8(217.5)	1.004	0.0039	0.791	0.995	-0.154	0.981
800	4.87	5.23	1668.8(1673.7)	871.8(870.1)	1.001	0.0009	0.797	0.986	-0.151	0.994

Table 3 shows the simulation results along with the theoretical values in parentheses for (i) the value of $\sqrt{c}$, (ii) the sizes of the left and right tailed tests, (iii) the expected value and standard deviation of the stopping time $\hat{\tau}_c$, and (iv) the mean and standard deviation of the sequential estimators $\left(\hat{\alpha}_{c,\hat{\tau}_c}, \hat{\psi}_{1,\hat{\tau}_c}, \hat{\psi}_{2,\hat{\tau}_c}\right)$.

Let us examine the performance of the sequential unit root test by looking at the columns for Size in Table 3 and the columns for Power in Tables 4 and 5. For the left tailed test, both the simulated sizes in Table 3 and powers in Table 4 perform well for all c in the sense that the theoretical values are close to the simulation results. In the case of the right tailed test, when c is relatively small, the simulated sizes in Table 3 and powers in Table 5 are quite different from the theoretical values. When the null hypothesis is true, even with $\sqrt{c} = 200$, its size (7.61% of the right size in Table 3) does not match the nominal size of 5%. However, for relatively large enough c, for example $\sqrt{c} = 800$, the simulated results of sizes approximate the theoretical values well. When the alternative hypothesis is true, the power approaches the theoretical value as c increases. The simulated powers get closer to the theoretical powers from $\sqrt{c} = 100$ under the explosive alternative. In all three tables, the results listed in the third to sixth columns clearly indicate the good performance of the stopping times and sequential estimators even when c is relatively small.

4.4 The simulation results when lag length is unknown

We use simulations to investigate the performance of the sequential unit root test when the lag length of the AR process is unknown. We find the followings. When the lag length is insufficient, there exists severe size distortion in both left tailed and right tailed tests. When sufficiently long lag lengths are used, the rejection rates are comparable to those when the correct lag lengths are used.

Table 6 gives simulation results when the lag length of the model is insufficient. The DGP is the AR(4) process with $\alpha_1 = 0.7$, $\alpha_2 = 0.7$, $\alpha_3 = 0.6$, but AR(3) is employed as the estimation model for the unit root test. The values in parentheses are the theoretical values when the AR process with true lag length, that is, AR(3), is used as the estimated model. We set $m_0 = 30$ for the definition of $\hat{\tau}_c$ in (56). The simulated sizes of left and right tailed test in Table

Table 4: Powers and moments under stationary alternatives (significance level: 5%)

$\alpha_c = 0.99$	Power (%)	$\hat{\tau}_c$		$\hat{\alpha}_{c,\hat{\tau}_c}$		$\hat{\psi}_{1,\hat{\tau}_c}$		$\hat{\psi}_{2,\hat{\tau}_c}$	
$\sqrt{c}, \delta$		Mean	sd	Mean	sd	Mean	sd	Mean	sd
50	12.50	126.0	68.4	1.012	0.0169	0.765	1.040	-0.168	1.005
$\delta = -0.5$	(12.61)	(129.8)	(64.5)	(0.99)	(0.02)	(0.8)	(0.9887)	(-0.15)	(0.989)
100	25.00	312.0	154.1	1.003	0.0069	0.782	1.011	-0.155	0.996
$\delta = -1$	(25.95)	(321.2)	(147.5)	(0.99)	(0.01)	(0.8)	(0.9887)	(-0.15)	(0.989)
150	42.64	575.8	257.6	1.000	0.0039	0.788	1.012	-0.153	0.998
$\delta = -1.5$	(44.24)	(589.9)	(244.2)	(0.99)	(0.0067)	(0.8)	(0.9887)	(-0.15)	(0.989)
200	61.88	933.5	364.2	0.999	0.0023	0.791	0.991	-0.151	0.996
$\delta = -2$	(63.88)	(947.0)	(348.4)	(0.99)	(0.005)	(0.8)	(0.9887)	(-0.15)	(0.989)

Table 4 corresponds to the alternative hypotheses $\alpha = 0.99$ and shows the simulation results along with the theoretical values in parentheses for (i) the values of $\sqrt{c}$ and δ, (ii) the powers of the left tailed test, (iii) the expected value and standard deviation of the stopping time $\hat{\tau}_c$, and (iv) the mean and standard deviation of the sequential estimators $\left(\hat{\alpha}_{c,\hat{\tau}_c}, \hat{\psi}_{1,\hat{\tau}_c}, \hat{\psi}_{2,\hat{\tau}_c}\right)$.

Table 5: Powers and moments under explosive alternatives (significance level: 5%)

$\alpha_c = 1.01$	Power (%)	$\hat{\tau}_c$		$\hat{\alpha}_{c,\hat{\tau}_c}$		$\hat{\psi}_{1,\hat{\tau}_c}$		$\hat{\psi}_{2,\hat{\tau}_c}$	
$\sqrt{c}, \delta$		Mean	sd	Mean	sd	Mean	sd	Mean	sd
50	19.71	83.4	45.3	1.026	0.0190	0.764	1.045	-0.175	0.981
$\delta = 0.5$	(12.61)	(85.1)	(44.7)	(1.01)	(0.02)	(0.8)	(0.989)	(-0.15)	(0.989)
100	28.99	140.3	72.4	1.016	0.0098	0.779	1.029	-0.163	0.981
$\delta = 1$	(25.95)	(140.8)	(72.4)	(1.01)	(0.01)	(0.8)	(0.989)	(-0.15)	(0.989)
150	46.66	179.1	88.7	1.014	0.0066	0.787	1.002	-0.162	0.980
$\delta = 1.5$	(44.24)	(178.9)	(88.0)	(1.01)	(0.0067)	(0.8)	(0.989)	(-0.15)	(0.989)
200	64.88	208.8	96.0	1.012	0.0049	0.789	1.004	-0.158	0.986
$\delta = 2$	(63.88)	(206.5)	(96.3)	(1.01)	(0.005)	(0.8)	(0.989)	(-0.15)	(0.989)

Table 5 corresponds to the alternative hypotheses $\alpha = 1.01$ and shows the simulation results along with the theoretical values in parentheses for (i) the values of $\sqrt{c}$ and δ, (ii) the powers of the right tailed test, (iii) the expected value and standard deviation of the stopping time $\hat{\tau}_c$, and (iv) the mean and standard deviation of the sequential estimators $\left(\hat{\alpha}_{c,\hat{\tau}_c}, \hat{\psi}_{1,\hat{\tau}_c}, \hat{\psi}_{2,\hat{\tau}_c}\right)$.

6 are 1.13% and 23.95% respectively though the nominal size is 5%. Because of the misspecification, it may be inevitable that the estimated parameters are biased and the sizes are distorted.

Table 6: Rejection rates (RRs) and moments under the null (DGP: AR(4), estimated model: AR(3), significance level: 5%)

	RR (%)		$\hat{\tau}_c$		$\hat{\alpha}_{c,\hat{\tau}_c}$	
$\sqrt{c} = 200$	Left	Right	Mean	sd	Mean	sd
$\alpha_c = 1$	1.13	23.95	297.1	162.1	1.008	0.0045
	(5)	(5)	(418.4)	(217.5)	(1)	$(1/\sqrt{c})$
$\alpha_c = 0.99$	16.27		521.5	242.6	1.003	0.0066
	(63.88)		(881.4)	(384.4)	(0.99)	(0.005)
$\alpha_c = 1.01$		74.23	174.1	86.5	1.015	0.0078
		(63.88)	(206.5)	(95.3)	(1.01)	(0.005)

Table 6 shows the simulation results along with the theoretical values in parentheses for (i) the sizes and powers of the left and right tailed test, (ii) the expected value and standard deviation of the stopping time $\hat{\tau}_c$, and (iii) the mean and standard deviation of the sequential estimator $\hat{\alpha}_{c,\hat{\tau}_c}$.

Tables 7-8 report the results when we apply models with sufficient lag length. From Theorem 2, we expect redundant lag lengths have no effect on the distribution of $\hat{\delta}_{\hat{\tau}_c}$ asymptotically. We confirm by simulation that the test yields similar results to those obtained when using true lag lengths. We now consider a unit root test with a lag length of 3 in the estimated model under the DGP of the AR(2) process with $\alpha_1 = 0.5$. Tables 7 and 8 respectively list the sizes and powers along with the results of the sequential estimation. For the left and right tailed tests, we see that the size properties are almost the same as when the lag length is correctly specified. In the left tailed test, the size characteristics in Table 7 are very good, as in Table 3. In the right tailed test, when $c \to \infty$, the size distortions get better as $11.72, 8.54, \ldots, 5.29$ from the top in Table 7, which is almost the same as $12.95, 9.83, \ldots, 5.23$ from the top in Table 3. The powers are also as good as those in Tables 4 and 5. The sequential estimation of the parameters of the remaining stationary process v_n with sufficiently long lag length is also well done.

We also conducted simulation studies using a lag length of 3 under the DGP of the AR(1) process. The results are similar to the the case using the DGP of the AR(2).

5 Concluding remarks

This study considers testing for the existence of a unit root for AR(p) models under sequential sampling. Using a time change via the observed Fisher information, we obtain the joint asymptotic densities and Laplace transforms of the

Table 7: Sizes of left and right tailed tests and moments under the null (DGP: AR(2), estimated model: AR(3), significance level: 5%)

$\alpha_c = 1$	Size (%)		$\hat{\tau}_c$		$\hat{\alpha}_{c,\hat{\tau}_c}$		$\hat{\psi}_{1,\hat{\tau}_c}$		$\hat{\psi}_{2,\hat{\tau}_c}$	
$\sqrt{c}$	Left	Right	Mean	sd	Mean	sd	Mean	sd	Mean	sd
	(5%)	(5%)	$\sqrt{c} \times 2.0921$	$\sqrt{c} \times 1.0876$	(1)	$(1/\sqrt{c})$	(0.5)	(1)	(0)	(1)
50	4.67	11.72	102.3 (104.6)	56.0 (54.4)	1.016	0.0181	0.468	1.053	-0.030	1.006
100	4.80	8.54	207.6 (209.2)	111.7 (108.8)	1.008	0.0085	0.483	1.026	-0.013	1.012
150	5.36	7.32	314.4 (313.8)	166.2 (163.1)	1.005	0.0055	0.490	1.026	-0.010	1.013
200	4.57	6.62	415.6 (418.4)	217.3 (217.5)	1.004	0.0039	0.493	1.017	-0.008	1.005
800	4.84	5.29	1664.4(1673.7)	867.1(870.1)	1.001	0.0009	0.499	1.003	-0.002	1.008

Table 7 shows the simulation results along with the theoretical values in parentheses for (i) the value of $\sqrt{c}$, (ii) the sizes of the left and right tailed tests, (iii) the expected value and standard deviation of the stopping time $\hat{\tau}_c$, and (iv) the mean and standard deviation of the sequential estimators $\left(\hat{\alpha}_{c,\hat{\tau}_c}, \hat{\psi}_{1,\hat{\tau}_c}, \hat{\psi}_{2,\hat{\tau}_c}\right)$.

sequential test statistic and the stopping time both under the null and local alternatives from the theory of Bessel bridges in Pitman and Yor (1982). The null distribution of the stopping time is characterized by a 3/2-dimensional Bessel process, whereas the distribution under the local alternatives is represented in terms of a 3/2-dimensional Bessel process with a drift.

In this study, the simulations have shown that the sequential unit root test performs well when the estimated model has a sufficiently long lag length. However, we also find that it does not work when the lag length of the estimated model is insufficient

There are some extensions for future research. In this study, we do not consider the upper limit of sample size, but in reality, there may be an upper limit to the acceptable sample size due to cost and budget constraints on sampling. We have developed a theory that also incorporates the upper limit of the sample size in a concurrent study. We assume that the disturbances constitute a sequence of $i.i.d.(0, \sigma^2)$ random variables, but this assumption could be relaxed to disturbances of martingale differences satisfying some additional conditions. The extension to sequential analysis of nonparametric AR processes and sequential tests for structural break or change point problems can be considered. We will consider a sequential analysis to determine the lag length consistent with the unit root test.

References

[1] Billingsley, P. (1999), *Convergence of Probability Measures*, 2nd Ed., John-Wiley and Sons.

[2] Blanchard, O. (2016), *Macroeconomics 7th Edition*, Pearson.

Table 8: Powers and moments under alternatives (DGP: AR(2), estimated model: AR(3), significance level: 5%)

$\alpha_c = 0.99$	Power (%)	$\hat{\tau}_c$		$\hat{\alpha}_{c,\hat{\tau}_c}$		$\hat{\psi}_{1,\hat{\tau}_c}$		$\hat{\psi}_{2,\hat{\tau}_c}$	
$\sqrt{c}, \delta$		Mean	sd	Mean	sd	Mean	sd	Mean	sd
50	11.90	126.5	68.3	1.009	0.0168	0.469	1.055	-0.027	1.009
$\delta = -0.5$	(12.61)	(129.8)	(70.7)	(0.99)	(0.02)	(0.5)	(1)	(0)	(1)
100	25.47	316.2	154.6	1.001	0.0069	0.486	1.025	-0.010	1.015
$\delta = -1$	(25.95)	(321.2)	(147.5)	(0.99)	(0.01)	(0.5)	(1)	(0)	(1)
150	43.01	582.4	254.8	0.999	0.0039	0.492	1.010	-0.005	1.018
$\delta = -1.5$	(44.24)	(589.8)	(241.1)	(0.99)	(0.0067)	(0.5)	(1)	(0)	(1)
200	61.45	928.4	363.8	0.998	0.0026	0.493	1.005	-0.003	1.007
$\delta = -2$	(63.88)	(947.0)	(345.0)	(0.99)	(0.005)	(0.5)	(1)	(0)	(1)
$\alpha_c = 1.01$									
50	19.34	85.4	46.5	1.024	0.0189	0.466	1.068	-0.032	1.008
$\delta = 0.5$	(12.61)	(85.1)	(44.7)	(1.01)	(0.02)	(0.5)	(1)	(0)	(1)
100	29.99	142.7	74.4	1.016	0.0098	0.483	1.037	-0.019	0.999
$\delta = 1$	(25.95)	(140.8)	(72.5)	(1.01)	(0.01)	(0.5)	(1)	(0)	(1)
150	45.99	180.9	89.2	1.013	0.0064	0.487	1.015	-0.013	1.011
$\delta = 1.5$	(44.24)	(178.9)	(88.0)	(1.01)	(0.0067)	(0.5)	(1)	(0)	(1)
200	64.13	209.1	98.1	1.012	0.0049	0.490	1.029	-0.012	1.014
$\delta = 2$	(63.88)	(206.5)	(95.3)	(1.01)	(0.005)	(0.5)	(1)	(0)	(1)

Table 8 shows the simulation results along with the theoretical values in parentheses for (i) the value of $\sqrt{c}$ and δ, (ii) the powers of the left and right tailed tests for $\alpha = 0.99$ and 1.01, (iii) the expected value and standard deviation of the stopping time $\hat{\tau}_c$, and (iv) the mean and standard deviation of the sequential estimators $\left(\hat{\alpha}_{c,\hat{\tau}_c}, \hat{\psi}_{1,\hat{\tau}_c}, \hat{\psi}_{2,\hat{\tau}_c}\right)$.

[3] Bobkoski, M. J. (1983), Hypothesis Testing in Nonstationary Time Series. Ph.D. thesis, Univ. Wisconsin-Madison.

[4] Borodin, A. N. and Salminen, P. (2002), *Handbook of Brownian Motion - Facts and Formulae*, 2nd Ed., Birkhäuser.

[5] Cavanagh, C. (1985), Roots Local To Unity. Manuscript, Department of Economics, Harvard University.

[6] Chan, N. H. and Wei, C. Z. (1987), Asymptotic Inference For Nearly Nonstationary AR(1) Processes. *Ann. Statist.* **15** 1050-1063.

[7] Chan, N. H. (1988), On The Parameter Inference for Nearly Nonstationary Time Series. *J. Amer. Math. Soc.* **83** 857-862

[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 Park, J. Y. (2002), On the Asymptotics of ADF Tests for Unit Roots. *Econometr. Rev.*, **21**, 431-447.

[10] Chang, Y. and Park, J. Y. (2004), Taking a New Contour: A Novel View on Unit Root Tests. Manuscript, Rice University.

[11] Chang, Y. (2012), Taking a New Contour: A Novel Approach to Panel Unit Root Test. *J. Econometr.*, **169** 15-28

[12] Choi, I. (2015), *Almost All about Unit Roots*, Cambridge University Press.

[13] Dickey, D. A. (1976), Estimation and Hypothesis Testing in Nonstationary Time Series. Ph.D. Thesis, Iowa State University.

[14] Dickey, D. A. and Fuller, W. A. (1979), Distribution of the Estimates for Autoregressive Time Series with a Unit Root. *J. Amer. Statist. Assoc.* **74** 427-431.

[15] 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.

[16] 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.

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

[18] Galtchouk, L. and Konev, V. (2006), Sequential Estimation of the Parameters in Unstable AR(2). *Sequential Anal.* **25** 25-43.

[19] Galtchouk, L. and Konev, V. (2011), On Asymptotic Normality of Sequential LS-Estimates of Unstable Autoregressive Processes. *Sequential Anal.* **30** 117-144.

[20] Hitomi, K., Nagai, K., Nishiyama, Y. and Tao, J. (2020), Sequential Test for Unit Root in First Order Autoregressive Model.

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

[22] Lai, T. L. and Siegmund, D. (1983), Fixed Accuracy Estimation of an Autoregressive Parameter. *Ann. Statist* **11** 478-485.

[23] Le Cam, L. (1986). Asymptotic methods in statistical decision theory. Springer, New York.

[24] Mukhopadhyay, N. and Sriram, T. N. (1992), On Sequential Comparisons of Means of First-Order Autoregressive Models. *Metrika* **39** 155-164.

[25] Novikov, A. A. (1972). Sequential estimation of the parameters of diffusion processes. Mathematical Notes of the Academy of Sciences of the USSR, 12(5), 812-818.

[26] Phillips, P. C. B. (1987a), Time Series Regression with a Unit Root. Econometrica, 1987, **55** 2 277-301.

[27] Phillips, P. C. B. (1987b), Towards a Unified Asymptotic Theory for Autoregression. *Biometrika*, **71** 535-547.

[28] Phillips, P. C. B. and Perron, P. (1988), Testing for a Unit Root in Time Series Regression, *Biometrika*, **75** 335-346.

[29] Pitman, J. and Yor, M. (1982), A Decomposition of Bessel Bridges. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, **59** 425-457.

[30] Revuz, D. and Yor, M. (1999), *Continuous Martingale and Brownian Motion*, 3rd ed. Springer-Verlag.

[31] Said, E. S. and Dickey, D. A. (1984), Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika*, **71** 599-607.

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

[33] Stock, J. H. (1994), Unit Roots, Structural Breaks and Trends, *Handbook of Econometrics Vol 4*, Elsevier.

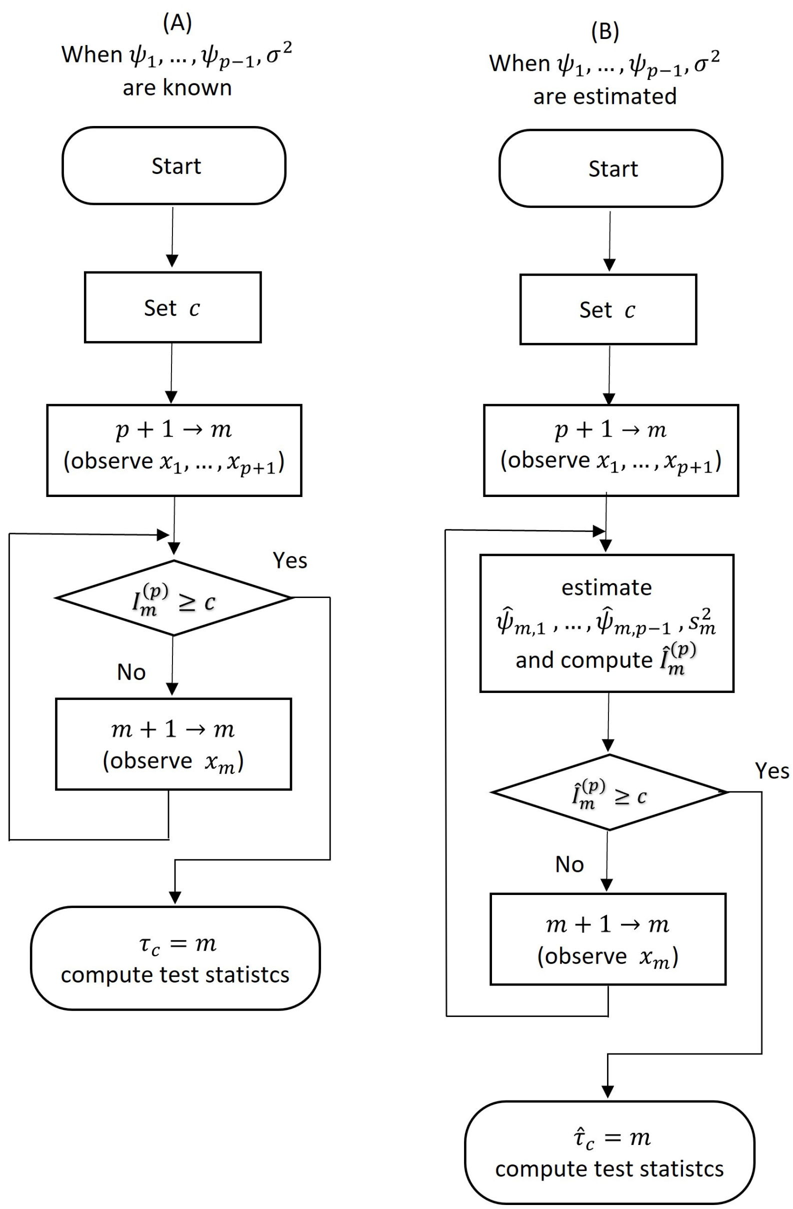
[34] Sriram, T. N. (2001), Fixed Size Confidence Regions for Parameters of Threshold AR(1) Models. *J. Statist. Plann. Inference.* **97** 293-304.

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

[36] Wald, A. (1947). Sequential Analysis. John Wiley and Sons, New York.

[37] White, J. S. (1958), The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case. *Ann. Math. Statist.* **29** 1188-1197.

Figure 1: The stopping rule of the sequential unit root tests



6 Appendix

6.1 Proof of Proposition 1

Lemma 7 proves (36) and (37) in Proposition 1. Lemmas 10 - 14 provide the proofs of the convergences of the five components of (38) in Proposition 1. Lemma 8 transforms the strong law of large numbers for discrete-time random variables into almost sure convergence of the corresponding continuous-time stochastic process in $D[0,\infty)$. Lemma 9 provides key reformulations of the processes x_n and Δx_n for the proofs. Using Lemmas 7 - 9, we prove Lemmas 10 - 14.

Lemma 7. *For α_i $(i=1,\ldots,p-1)$ in (7) with assumption (8), the asymptotic autocovariance function $\gamma_v(h)$ of v_n exists for $h=0,1,\ldots$ and satisfies (36) and (37).*

Proof. For v_n in (27), letting $\boldsymbol{v}_n=(v_n,v_{n-1},\cdots,v_{n-p+2})'$, $\boldsymbol{\epsilon_n}=(\epsilon_n,0,\cdots,0)'$, and

$$G=\begin{pmatrix} \psi_1 & \psi_2 & \cdots & \psi_{p-2} & \psi_{p-1} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

we have

$$\boldsymbol{v}_n = G\boldsymbol{v_{n-1}}+\boldsymbol{\epsilon_n}=\sum_{k=0}^{n-p-1} G^k\boldsymbol{\epsilon_{n-k}}+G^{n-p}\boldsymbol{v_p}. \tag{58}$$

We now consider another probability space $(\Omega',\mathcal{F}',P')$ on which independent variables $\epsilon_p,\epsilon_{p-1},\ldots,\epsilon_0,\epsilon_{-1},\ldots$ have the same distribution as ϵ_{p+1} on $(\Omega,\mathcal{F},P)$. Let $(\Omega^*,\mathcal{F}^*,P^*)$ be the completed probability space of the product space $(\Omega\times\Omega',\mathcal{F}\times\mathcal{F}',P\times P')$. Let $\boldsymbol{v}_n^*$ be the ergodic, stationary AR$(p-1)$ process in $(\Omega^*,\mathcal{F}^*,P^*)$;

$$=\sum_{k=0}^{\infty} G^k\boldsymbol{\epsilon}_{n-k} \quad (n=0,\pm1,\pm2,\ldots). \tag{59}$$

Then, from (58),

$$\boldsymbol{v}_n-\boldsymbol{v}_n^* = G^{n-p}\left(\boldsymbol{v_p}-\boldsymbol{v}_p^*\right) \quad (n=p+1,p+2,\ldots), \tag{60}$$

holds due to $\boldsymbol{v}_p^*=\sum_{k=0}^{\infty}G^k\boldsymbol{\epsilon}_{p-k}$. Let the autocovariance function of $\boldsymbol{v}_n^*$ be

$$\gamma_v(h)=E(v_{n+h}^*v_n^*) \quad (h=0,1,\ldots).$$

From the ergodic theorem,

$$\sum_{n=p+1}^{m} v_{n+h}^*v_n^*/\left(m-p\right)\to\gamma_v(h) \quad P^*-a.s.,$$

and

$$\sum_{n=p+1}^{m} \boldsymbol{v}^{*\prime}_{n+h} / (m-p) = \left(\sum_{n=p+1}^{m} v^*_{n+h-i} v^*_{n-j} \right)_{i,j} / (m-p) \to \boldsymbol{\Gamma}_v^h \quad P^* - a.s.,$$

where $i, j = 0, 1, \ldots, p-2$, and $\boldsymbol{\Gamma}_v^h = (\gamma_v(|h+j-i|))_{i,j}$. Using (60),

$$\begin{aligned}
&\frac{1}{m-p} \sum_{n=p+1}^{m} \boldsymbol{v}_{n+h} \boldsymbol{v}_n' \\
=&\frac{1}{m-p} \sum_{n=p+1}^{m} \boldsymbol{v}^{*\prime}_{n+h} + \frac{1}{m-p} \sum_{n=p+1}^{m} \left(G^{n-p+h} \left(\boldsymbol{v_p} - \boldsymbol{v}_p^* \right) \right) \left(G^{n-p} \left(\boldsymbol{v_p} - \boldsymbol{v}_p^* \right) \right)' \\
+&\frac{1}{m-p} \sum_{n=p+1}^{m} \boldsymbol{v}^*_{n+h} \left(G^{n-p} \left(\boldsymbol{v_p} - \boldsymbol{v}_p^* \right) \right)' + \frac{1}{m-p} \sum_{n=p+1}^{m} \left(G^{n-p+h} \left(\boldsymbol{v_p} - \boldsymbol{v}_p^* \right) \right)'.
\end{aligned} \tag{61}$$

By the Cayley-Hamilton theorem,

$$G^n = \psi_1 G^{n-1} + \psi_2 G^{n-1} + \cdots + \psi_{p-1} \boldsymbol{I}, \tag{62}$$

which implies that each element $(G^n)_{i,j}$ of G^n satisfies the difference equation with the characteristic polynomial $x^n = \psi_1 x^{n-1} + \psi_2 x^{n-1} + \cdots + \psi_{p-1}$. Thus, each element $(G^n)_{i,j}$ of G^n can be represented as a linear combination of ${}_nC_k \alpha_i^{n-k}$, $(k = 0, 1, \ldots, p-1)$ and is bounded by $K n^{p-2} \xi^n$ with some constant $K > 0$ and some $\xi \in (0,1)$ satisfying $\max_{i<p} |\alpha_i| < \xi$. Therefore, as $n \to \infty$,

$$\left(G^{n-p+h} \left(\boldsymbol{v_p} - \boldsymbol{v}_p^* \right) \right) \left(G^{n-p} \left(\boldsymbol{v_p} - \boldsymbol{v}_p^* \right) \right)' \to \boldsymbol{0} \quad P^* - a.s..$$

The Cesàro mean indicates that the second term of (61) P^*-almost surely converges to $\boldsymbol{0}$. The last two terms of the right side of (61) P^*-almost surely converge to $\boldsymbol{0}$, since there exists $n \geq \exists n_0$ such that $K n^{p-2} |\xi|^n \leq 1$.

Then, we have (36) from

$$\sum_{n=p+1}^{m} \boldsymbol{v}_{n+h} \boldsymbol{v}_n' / (m-p) \to \boldsymbol{\Gamma}_v^h \quad P - a.s..$$

From the functional central limit theorem (FCLT) with respect to ergodic, stationary martingale differences (Billingsley (1999), Theorem 18.3),

$$\sum_{n=p+1}^{\lfloor \sqrt{c} t \rfloor} \boldsymbol{v}^*_{n-1} \epsilon_n / \left(c^{1/4} \sigma \sqrt{\gamma_v(0)} \right) \Rightarrow \left(W_i(t) \right)_{i=1,\ldots p-1}.$$

Using (60), we have

$$\frac{1}{c^{1/4}\sigma\sqrt{\gamma_v(0)}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\boldsymbol{v}_{n-1}\epsilon_n = \frac{1}{c^{1/4}\sigma\sqrt{\gamma_v(0)}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\boldsymbol{v}^*_{n-1}\epsilon_n + \frac{1}{c^{1/4}\sigma\sqrt{\gamma_v(0)}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\epsilon_n G^{n-1-p}\left(\boldsymbol{v_p}-\boldsymbol{v}^*_p\right). \quad (63)$$

For $|\alpha|<1$, define $M_m=\sum_{n=p+1}^{m}{}_nC_k\alpha^{n-k}\epsilon_n$, $(k=0,1,\dots,p-1)$. Then each term of $\sum_{n=p+1}^{m}\epsilon_n G^n$ can be rewritten as a linear combination of such M_m's with $\alpha=\alpha_1,\dots,\alpha_{p-1}$. Due to $\sum_{n=p+1}^{\infty}\left({}_nC_k\alpha^{n-k}\right)^2<\infty$, M_m is a uniformly integrable L^2 martingale and $M_m\to M_\infty$ *a.s.*. There exists m_0 such that $|M_m|<|M_\infty|+1$ for any $m\geq m_0$. As $c\to\infty$

$$\max_{t\leq t_0}\left|M_{\lfloor\sqrt{c}t\rfloor}\right|/c^{1/4}\leq\max_{t_0\leq m_0}|M_m|/c^{1/4}+\left(|M_\infty|+1\right)/c^{1/4}\to 0 \quad a.s.,$$

which implies that the second term of the right side of (63) converges to $\mathbf{0}$ in $D[0,\infty)$. □

Lemma 8. *Suppose that for some random sequence w_n, $n=1,2,\dots$ and $\theta\in\mathbb{R}$ $\sum_{n=1}^{m}w_n/m\to\theta$ a.s. Let $I:[0,\infty)\to[0,\infty)$ be $I(t)=\theta t$ and*

$$I_c(t)=\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor}w_n/\sqrt{c}. \quad (64)$$

Then for any $t_0>0$ as $c\uparrow\infty$,

$$\|I_c-I\|_{t_0}=\sup_{t\leq t_0}|I_c(t)-\theta t|\to 0 \quad a.s.$$

Proof. Fix $\omega\in\{\sum_{n=1}^{m}w_n/m\to\theta\}$ and for any $t_0>0$ and $\varepsilon>0$ find m_0 so that for any $m\geq m_0$, $|\sum_{n=1}^{m}w_n/m-\theta|<\varepsilon/2t_0$. Then, for large enough $c>0$,

$$\begin{aligned}
&\sup_{t\leq t_0}|I_c(t)-\theta t|\\
&=\sup_{t\leq t_0}\left|\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor}\left(w_n-\theta\right)/\sqrt{c}+\left(\lfloor\sqrt{c}t\rfloor-\sqrt{c}t\right)\theta/\sqrt{c}\right|\\
&\leq\max_{m\leq m_0}\left|\sum_{n=1}^{m}\left(w_n-\theta\right)/\sqrt{c}\right|\vee\sup_{m_0\leq\sqrt{c}t\leq\sqrt{c}t_0}\left|t\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor}\left(w_n-\theta\right)/\left(\sqrt{c}t\right)\right|+|\theta|/\sqrt{c}\\
&\leq\varepsilon/4+\varepsilon t_0/2t_0+\varepsilon/4=\varepsilon. \quad (65)
\end{aligned}$$

□

The following lemma reforms x_n and Δx_n as linear combinations of near-unit-root AR(1) process Z_n in (27), a strongly stationary process y_n, and a nearly strongly stationary process $y_{c,n}$ defined in (66).

Lemma 9. *Let*

$$y_n = \sum_{k=1}^{p-1}\left(\sum_{i=k}^{p-1}\psi_i\right)v_{n-k+1} \quad \text{and} \quad y_{c,n} = \sum_{k=1}^{p-1}\left(\sum_{i=k}^{p-1}\frac{\psi_i}{\alpha_c^{i-k+1}}\right)v_{n-k+1}. \tag{66}$$

Then the near-unit-root AR(p) process x_n in (6) can be represented as

$$x_n = \left(z_n - y_{c,n}\right)/\Psi(\alpha_c^{-1}), \tag{67}$$

and the key relations for the proofs of asymptotic properties are obtained:

$$\Delta y_n = \epsilon_n - \Psi(1)v_n, \tag{68}$$

$$\Delta x_n = \left(\frac{\delta}{\sqrt{c}}z_{n-1} + \Psi(1)v_n + \Delta y_n - \Delta y_{c,n}\right)/\Psi(\alpha_c^{-1}), \tag{69}$$

$$y_n - y_{c,n} \to 0 \quad a.s. \quad \text{as} \quad c\to\infty, \quad \text{for} \quad \text{any} \quad n. \tag{70}$$

Proof. Since $x_{n-1} = x_n/\alpha_c - v_n/\alpha_c$ from (27), we can obtain by induction, for $i = 1,\ldots,p-1$,

$$x_{n-i} = \frac{x_n}{\alpha_c^i} - \sum_{k=1}^{i}\frac{v_{n-k+1}}{\alpha_c^{i-k+1}}.$$

Since $z_n = x_n - \sum_{i=1}^{p-1}\psi_i x_{n-i}$, we have

$$\begin{aligned} z_n &= x_n - \sum_{i=1}^{p-1}\psi_i\left(\frac{x_n}{\alpha_c^i} - \sum_{k=1}^{i}\frac{v_{n-k+1}}{\alpha_c^{i-k+1}}\right) \\ &= \left(1-\sum_{i=1}^{p-1}\frac{\psi_i}{\alpha_c^i}\right)x_n + \sum_{k=1}^{p-1}\left(\sum_{i=k}^{p-1}\frac{\psi_i}{\alpha_c^{i-k+1}}\right)v_{n-k+1} \\ &= \Psi(\alpha_c^{-1})x_n + y_{c,n}, \end{aligned} \tag{71}$$

which implies (67). As to Δy_n,

$$\begin{aligned} \Delta y_n &= \sum_{k=0}^{p-2}\sum_{i=k+1}^{p-1}\psi_i v_{n-k} - \sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\psi_i v_{n-k} = -\sum_{k=1}^{p-1}\psi_k v_{n-k} + \sum_{i=1}^{p-1}\psi_i v_n \\ &= \left(\Psi(L)-\Psi(1)\right)v_n = \epsilon_n - \Psi(1)v_n. \end{aligned}$$

Taking the difference of (67), we have

$$\begin{aligned} \Delta x_n &= \left(\Delta z_n - \Delta y_{c,n}\right)/\Psi(\alpha_c^{-1}) = \left(\frac{\delta}{\sqrt{c}}z_{n-1} + \epsilon_n - \Delta y_{c,n}\right)/\Psi(\alpha_c^{-1}) \\ &= \left(\frac{\delta}{\sqrt{c}}z_{n-1} + \Psi(1)v_n + \Delta y_n - \Delta y_{c,n}\right)/\Psi(\alpha_c^{-1}). \end{aligned} \tag{72}$$

Here we use (28) and (68) for the last two equations. Since $\alpha_c \to 1$, as $c \to \infty$,

$$y_n - y_{c,n} = \sum_{k=1}^{p-1} \left(\sum_{i=k}^{p-1} \psi_i (1 - 1/\alpha_c^{i-k+1}) \right) v_{n-k+1} \to 0 \quad a.s..$$

□

Define for $i, j = 1, \dots, p-1$,

$$F_c(t) = \sum_{n=p+1}^{\lfloor \sqrt{c} t \rfloor} z_{n-1}^2 / \left(c\sigma^2 \right), \qquad J_c(t) = \sum_{n=p+1}^{\lfloor \sqrt{c} t \rfloor} z_{n-1} \Delta z_n / \left(\sqrt{c} \sigma^2 \right), \tag{73}$$

$$I_c(t) = \sum_{n=p+1}^{\lfloor \sqrt{c} t \rfloor} \epsilon_n^2 / \sqrt{c}, \qquad I_{ij,c}(t) = \sum_{n=p+1}^{\lfloor \sqrt{c} t \rfloor} v_{n-i} v_{n-j} / \sqrt{c}, \tag{74}$$

$$H_c(t) = \sum_{n=p+1}^{\lfloor \sqrt{c} t \rfloor} v_n / \sqrt{c}, \qquad H_{i,c}(t) = \sum_{n=p+1}^{\lfloor \sqrt{c} t \rfloor} \epsilon_n v_{n-i} / \sqrt{c}, \tag{75}$$

$$W_{i,c}(t) = \sum_{n=p+1}^{\lfloor \sqrt{c} t \rfloor} v_{n-i} \epsilon_n / \left(c^{1/4} \sqrt{\gamma_v(0)} \sigma \right). \tag{76}$$

The convergence with respect to (Z_c, F_c, J_c) in Hitomi, Nagai, Nishiyama, and Tao (2020) with Lemma 8 and Lemma 7 , we obtain, in the sense of $D[0, \infty)$,

$$\begin{aligned} &\begin{pmatrix} Z_c & F_c & J_c & W_{i,c} \end{pmatrix} \Rightarrow \begin{pmatrix} Z & F & J & W_i \end{pmatrix}, \\ &\begin{pmatrix} I_c & I_{ij,c} & H_c & H_{i,c} \end{pmatrix} \to \begin{pmatrix} I & I_{ij} & 0 & 0 \end{pmatrix} \quad a.s., \end{aligned}$$

where Z is in (33), and

$$\begin{aligned} F(t) = \int_0^t Z_u^2 du, \quad J(t) = \int_0^t Z_u dZ_u, \\ I(t) = \sigma^2 t, \quad I_{ij}(t) = \gamma_v(|i-j|) t. \end{aligned}$$

Applying Skorohod's theorem and changing $(\Omega, \mathcal{F}, P)$ to a completed probability space in which, as $c \to \infty$, in the sense of space $D[0, \infty)$,

$$\begin{aligned} &\begin{pmatrix} Z_c & F_c & J_c & W_{i,c} & I_c & I_{ij,c} & H_c & H_{i,c} \end{pmatrix} \\ \to &\begin{pmatrix} Z & F & J & W_i & I & I_{ij} & 0 & 0 \end{pmatrix} \quad a.s. \end{aligned} \tag{77}$$

where

$$F_c(t) = \sum_{n=p+1}^{\lfloor \sqrt{c} t \rfloor} z_{n-1}^2 / \left(c\sigma^2 \right), \quad J_c(t) = \sum_{n=p+1}^{\lfloor \sqrt{c} t \rfloor} z_{n-1} \Delta z_n / \left(c\sigma^2 \right). \tag{78}$$

Next lemma proves the convergence of the first component in (38).

Lemma 10. *As $c \to \infty$,*

$$\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}^2/\left(c\sigma^2\right) \to F(t)/\Psi^2(1) = \int_0^t Z^2(u)du/\Psi^2(1)$$

in the sense of space $D[0,\infty)$.

Proof. We obtain the first convergence of (38). We use (67) and write

$$\begin{aligned}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}^2/c\sigma^2 &= \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(\left(z_{n-1} - y_{c,n-1}\right)/\Psi(\alpha_c^{-1})\right)^2/c\sigma^2 \\ &= \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(z_{n-1}^2 - 2z_{n-1}y_{c,n-1} + y_{c,n-1}^2\right)/\left(\Psi^2(\alpha_c^{-1})c\sigma^2\right). \quad (79)\end{aligned}$$

For the right-hand side of the last equation, we prove as $c \to \infty$,

$$\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} y_{c,n-1}^2/c \to 0, \quad \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(y_n - y_{c,n}\right)^2/\sqrt{c} \to 0,$$

$$\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(\Delta y_n - \Delta y_{c,n}\right)^2/\sqrt{c} \to 0, \quad (80)$$

uniformly in $t \in [0,m]$ for any $m > 0$ a.s.. As $c \to \infty$, we obtain

$$\begin{aligned}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} y_{c,n-1}^2/c &= \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(\sum_{k=1}^{p-1}\sum_{i=k}^{p-1} \frac{\psi_i}{\alpha_c^{i-k+1}} v_{n-k}\right)^2/c \\ &= \frac{1}{\sqrt{c}} \sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\sum_{l=1}^{p-1}\sum_{j=l}^{p-1} \frac{\psi_i}{\alpha_c^{i-k+1}} \frac{\psi_j}{\alpha_c^{j-l+1}} \left(\frac{1}{\sqrt{c}} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} v_{n-k}v_{n-l}\right) \\ &= \frac{1}{\sqrt{c}} \sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\sum_{l=1}^{p-1}\sum_{j=l}^{p-1} \frac{\psi_i}{\alpha_c^{i-k+1}} \frac{\psi_j}{\alpha_c^{j-l+1}} I_{kl,c}(t) \to 0,\end{aligned}$$

since

$$\sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\sum_{l=1}^{p-1}\sum_{j=l}^{p-1} \psi_i\psi_j/(\alpha_c^{i-k+1}\alpha_c^{j-l+1}) = O(1).$$

For $\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(y_n-y_{c,n}\right)^2/\sqrt{c}\to 0$, as $c\to\infty$, $1/\alpha_c^{i-k+1}\to 0$. We have

$$
\begin{aligned}
&\frac{1}{\sqrt{c}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(y_n-y_{c,n}\right)^2\\
=&\frac{1}{\sqrt{c}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(\sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\psi_i\left(1-\frac{1}{\alpha_c^{i-k+1}}\right)v_{n-k+1}\right)^2\\
=&\sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\sum_{l=1}^{p-1}\sum_{j=l}^{p-1}\psi_i\psi_j\left(1-\frac{1}{\alpha_c^{i-k+1}}\right)\left(1-\frac{1}{\alpha_c^{j-l+1}}\right)\left(\frac{1}{\sqrt{c}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}v_{n-k}v_{n-l}\right)\\
=&\sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\sum_{l=1}^{p-1}\sum_{j=l}^{p-1}\psi_i\psi_j\left(1-\frac{1}{\alpha_c^{i-k+1}}\right)\left(1-\frac{1}{\alpha_c^{j-l+1}}\right)I_{kl,c}(t)\to 0
\end{aligned}
$$

uniformly in $t\in[0,m]$ for any $m>0$ *a.s.*. The other convergence holds from

$$
\begin{aligned}
\left(\Delta y_n-\Delta y_{c,n}\right)^2&=\left(y_n-y_{c,n}-\left(y_{n-1}-y_{c,n-1}\right)\right)^2\\
&\le 2\left(y_n-y_{c,n}\right)^2+2\left(y_{n-1}-y_{c,n-1}\right)^2.
\end{aligned}
$$

Using the Cauchy-Schwarz inequality and (77), (80), we can obtain

$$
\left|\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}z_{n-1}y_{c,n-1}/c\right|\le\sqrt{\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}z_{n-1}^2/c}\sqrt{\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}y_{c,n-1}^2/c}\to 0.\quad a.s.,
$$

which shows Lemma 10. □

Next lemma proves the convergence of the third component in (38).

Lemma 11. $\left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\Delta x_{n-i}\Delta x_{n-j}/\sqrt{c}\right)_{i,j}\to\left(\gamma_v\left(|i-j|\right)t\right)_{i,j}\quad a.s..$

Proof. Using (72), we have

$$
\begin{aligned}
&\frac{1}{\sqrt{c}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\Delta x_{n-i}\Delta x_{n-j}\\
=&\frac{1}{\sqrt{c}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(\frac{\delta}{\sqrt{c}}z_{n-i-1}+\Psi(1)v_{n-i}+\Delta y_{n-i}-\Delta y_{c,n-i}\right)\\
&\times\left(\frac{\delta}{\sqrt{c}}z_{n-j-1}+\Psi(1)v_{n-j}+\Delta y_{n-j}-\Delta y_{c,n-j}\right).
\end{aligned}
$$

Using the ergodic theorem, we have

$$\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \Psi^2(1)v_{n-i}v_{n-j}/\left(\sqrt{c}\Psi^2(\alpha_c^{-1})\right) = \Psi^2(1)I_{kl,c}(t)/\left(\Psi^2(\alpha_c^{-1})\right) \to \gamma_v\left(|i-j|\right)t$$

uniformly in $t \in [0, m]$ for any $m > 0$ $a.s.$. The Cauchy-Schwarz inequality and (80) help for the evaluation of the other terms. □

Next lemma proves the convergence of the forth component in (38).

Lemma 12. $\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\epsilon_n/\left(\sqrt{c}\sigma^2\right) \to \int_0^t Z_s dW_s/\Psi(1)$.

Using (67), $(I_c, W_{i,c}) \to (I, W_i)$ in (77) and (80), we have

$$\begin{aligned}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\epsilon_n/\sqrt{c} = &\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(z_{n-1} - y_{n-1} + (y_{n-1} - y_{c,n-1})\right)\epsilon_n/\left(\sqrt{c}\Psi(\alpha_c^{-1})\right) \\ \sim &\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\epsilon_n/\left(\sqrt{c}\Psi(\alpha_c^{-1})\right).\end{aligned}$$

Using (77), we have

$$\begin{aligned}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\epsilon_n/\left(\sqrt{c}\sigma^2\right) &= \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\left(\Delta z_n - \delta z_{n-1}/\sqrt{c}\right)/\left(\sqrt{c}\sigma^2\right) \\ &= J_c(t) - \delta F_c(t) \to \int_0^t Z_s dZ_s - \delta\int_0^t Z_s^2 ds = \int_0^t Z_s dW_s.\end{aligned} \tag{81}$$

Next lemma proves the convergence of the fifth component in (38).

Lemma 13. *As* $c \to \infty$*,*

$$\left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \Delta x_{n-i}\epsilon_n/\left(c^{1/4}\sqrt{\gamma_v(0)}\sigma\right)\right)_i \to \left(W_i(t)\right)_i. \tag{82}$$

Proof. By (72), we have

$$\begin{aligned}&\frac{1}{c^{1/4}\sigma^2}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \Delta x_{n-i}\epsilon_n \\ =&\frac{1}{c^{1/4}\sigma^2\Psi(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(\frac{\delta}{\sqrt{c}}z_{n-i-1} + \Psi(1)v_{n-i} + \Delta y_{n-i} - \Delta y_{c,n-i}\right)\epsilon_n.\end{aligned}$$

By (28), $\Delta z_n = \delta z_{n-1}/\sqrt{c} + \epsilon_n$, $J_c \to J$ in (77) and (81) yield

$$\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} (\Delta z_n)^2/\sqrt{c} = \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(\delta^2 z_{n-1}^2/c + 2\delta z_{n-1}\epsilon_n/\sqrt{c} + \epsilon_n^2\right)/\sqrt{c} \qquad (83)$$
$$\to I(t) = \sigma^2 t.$$

(81) in Lemma 12 and (83) imply, as $c \to \infty$,

$$\frac{1}{\sqrt{c}c^{1/4}\sigma^2} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} (\delta z_{n-1-i})\,\epsilon_n = \frac{1}{\sqrt{c}c^{1/4}\sigma^2}\delta \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(z_{n-1} - \sum_{k=1}^{i} \Delta z_{n-k}\right)\epsilon_n \to 0$$

from the Cauchy-Schwarz inequality. Since $W_{c,i} \to W_i$ in (77), and $a_c \to 1$,

$$\frac{1}{c^{1/4}} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} (y_{n-i} - y_{c,n-i})\,\epsilon_n = \sum_{k=1}^{p-1}\sum_{i=k}^{p-1} \psi_i \left(1 - \frac{1}{\alpha_c^{i-k+1}}\right) \frac{1}{c^{1/4}} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} v_{n-i+k+1}\epsilon_n \to 0,$$

we have (82). □

Next lemma proves the convergence of the second component in (38).

Lemma 14. *As $c \to \infty$, $\left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\Delta x_{n-i}/c^{3/4}\sigma^2\right)_i \to \mathbf{0}$.*

Proof. Using (67) we have

$$\begin{aligned}
&\frac{1}{c^{3/4}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\Delta x_{n-i}\\
=&\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(z_{n-1}-y_{c,n-1}\right)\left(\Delta z_{n-i}-\Delta y_{c,n-i}\right)\\
=&\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(z_{n-1}-y_{n-1}+\left(y_{n-1}-y_{c,n-1}\right)\right)\\
&\times\left(\Delta z_{n-i}-\Delta y_{n-i}+\left(\Delta y_{n-i}-\Delta y_{c,n-i}\right)\right)\\
=&\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\Delta z_{n-i}-\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\Delta y_{n-i}\\
&+\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\left(\Delta y_{n-i}-\Delta y_{c,n-i}\right)+\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} y_{n-1}\Delta z_{n-i}\\
&-\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} y_{n-1}\Delta y_{n-i}+\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} y_{n-1}\left(\Delta y_{n-i}-\Delta y_{c,n-i}\right)\\
&+\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(y_{n-i}-y_{c,n-i}\right)\Delta z_{n-i}-\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(y_{n-i}-y_{c,n-i}\right)\Delta y_{n-i}\\
&+\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(y_{n-i}-y_{c,n-i}\right)\left(\Delta y_{n-i}-\Delta y_{c,n-i}\right). \qquad (84)
\end{aligned}$$

Except for the first two terms, the terms in the last equality converge to 0 by (80) and (83), and $I_{ij,c}\to I_{ij}$. As for the first two terms, since $z_{n-1}=\sum_{k=1}^{i}\Delta z_{n-k}+z_{n-i-1}$, we have

$$\begin{aligned}
\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\Delta z_{n-i}/c^{3/4}&=\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(\sum_{k=1}^{i}\Delta z_{n-k}+z_{n-i-1}\right)\Delta z_{n-i}/c^{3/4}\\
&=\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\sum_{k=1}^{i}\Delta z_{n-k}\Delta z_{n-i}/c^{3/4}+\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-i-1}\Delta z_{n-i}/c^{3/4}.
\end{aligned}$$

As $c\to\infty$, the second term in the last equation converges to 0 uniformly in $t\in[0,m]$ for any $m>0$ *a.s.*, since

$$\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor} z_{n-i-1}\Delta z_{n-i}/\left(\sqrt{c}\sigma^2\right)\sim J_c(t)\to\int_0^t Z_u dZ_u \qquad (85)$$

uniformly in $t \in [0, m]$ for any $m > 0$ $a.s.$

Next, as $c \to \infty$,

$$\begin{aligned}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\Delta y_{n-i}/c^{3/4} =& \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\left(y_{n-i} - y_{n-i-1}\right)/c^{3/4} \\ =& -\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor-1} \Delta z_n y_{n-i}/c^{3/4} - z_p y_{p-i}/c^{3/4} + z_{\lfloor\sqrt{c}t\rfloor-1} y_{\lfloor\sqrt{c}t\rfloor-i}/c^{3/4} \\ \to& 0.\end{aligned}$$

Using the Cauchy-Schwarz inequality and (83), we find that the first term converges to 0 uniformly in t. The third term in the last equation converges to 0, since $z_{\lfloor\sqrt{c}t\rfloor-1}/c^{1/4}\sigma \to Z(t)$ and $v_{\lfloor\sqrt{c}t\rfloor-i}/\sqrt{c} = H_c\left(t - i/\sqrt{c}\right) - H_c\left(t - (i-1)/\sqrt{c}\right) \to 0$. $\square$

6.2 Proof of Theorem 2

Now, we conclude the proof of the main theorem.

Proof. Let $\tilde{\boldsymbol{x}}_{\boldsymbol{n-1}} = (x_{n-1}, x_{n-2}, \cdots, x_{n-p})'$, $\boldsymbol{\psi^*} = \left(0, \boldsymbol{\psi}^t\right)'$ and $\hat{\boldsymbol{\psi}}^*_{\boldsymbol{m}} = \left(0, \hat{\boldsymbol{\psi}}^{\boldsymbol{t}}_{\boldsymbol{m}}\right)'$, according to (17), (25),

$$\boldsymbol{\psi^*} = A_I \boldsymbol{\phi^c} \quad \text{and} \quad \hat{\boldsymbol{\psi}}^*_{\boldsymbol{m}} = \hat{A}_{I,m}\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\boldsymbol{m}}, \tag{86}$$

where $A_I = \begin{pmatrix} 0 & 0 \\ 0 & A_c^{-1} \end{pmatrix}$, $\hat{A}_{I,m} = \begin{pmatrix} 0 & 0 \\ 0 & \hat{A}_{c,m}^{-1} \end{pmatrix}$ and

$$\hat{A}_{c,m}^{-1} = \begin{pmatrix} 1/\hat{\alpha}_{c,m} & \left(1-\hat{\alpha}_{c,m}\right)/\hat{\alpha}_{c,m}^2 & \cdots & \left(1-\hat{\alpha}_{c,m}\right)/\hat{\alpha}_{c,m}^{p-2} & \left(1-\hat{\alpha}_{c,m}\right)/\hat{\alpha}_{c,m}^{p-1} \\ 0 & 1/\hat{\alpha}_{c,m} & \cdots & \left(1-\hat{\alpha}_{c,m}\right)/\hat{\alpha}_{c,m}^{p-3} & \left(1-\hat{\alpha}_{c,m}\right)/\hat{\alpha}_{c,m}^{p-2} \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1/\hat{\alpha}_{c,m} & \left(1-\hat{\alpha}_{c,m}\right)/\hat{\alpha}_{c,m}^2 \\ 0 & 0 & \cdots & 0 & 1/\hat{\alpha}_{c,m} \end{pmatrix}.$$

Using (27),

$$\begin{aligned}\hat{\Psi}_m(L)x_{n-1} &= \Psi(L)x_{n-1} + \left(\hat{\Psi}_m(L) - \Psi(L)\right)x_{n-1} \\ &= z_{n-1} + \left(\hat{\boldsymbol{\psi}}^*_{\boldsymbol{m}} - \boldsymbol{\psi^*}\right)'\tilde{\boldsymbol{x}}_{\boldsymbol{n-1}} \\ &= z_{n-1} + \left(\hat{A}_{I,m}\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\boldsymbol{m}} - A_I\boldsymbol{\phi^c}\right)'\tilde{\boldsymbol{x}}_{\boldsymbol{n-1}} \\ &= z_{n-1} + \left(\hat{A}_{I,m}\left(\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\boldsymbol{m}} - \boldsymbol{\phi^c}\right) + o(1)\right)'\tilde{\boldsymbol{x}}_{\boldsymbol{n-1}}.\end{aligned}$$

The LSE

$$\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\boldsymbol{m}} - \boldsymbol{\phi^c} = \boldsymbol{D_c^{-1}\Xi_{c,m}^{-1}D_c^{-1}}X_m'\boldsymbol{\epsilon_m},$$

where $\boldsymbol{\Xi_{c,m}} = \boldsymbol{D_c^{-1}} X_m' X_m \boldsymbol{D_c^{-1}}$. Then the observed Fisher information becomes

$$\sum_{n=p+1}^{m} \left(\hat{\Psi}_m(L) x_{n-1}\right)^2 = \sum_{n=p+1}^{m} z_{n-1}^2 + 2\left(\hat{A}_{I,m}\left(\boldsymbol{\hat{\phi}_m^c} - \boldsymbol{\phi^c}\right) + o(1)\right)' \sum_{n=p+1}^{m} \boldsymbol{\tilde{x}_{n-1}} z_{n-1}$$

$$+ \left(\hat{A}_{I,m}\left(\boldsymbol{\hat{\phi}_m^c} - \boldsymbol{\phi^c}\right) + o(1)\right)' \sum_{n=p+1}^{m} \boldsymbol{\tilde{x}_{n-1}} \boldsymbol{\tilde{x}'_{n-1}} \tag{87}$$

$$\times \left(\hat{A}_{I,m}\left(\boldsymbol{\hat{\phi}_m^c} - \boldsymbol{\phi^c}\right) + o(1)\right), \tag{88}$$

and the estimator of σ^2 is

$$s_m^2 = \left(\boldsymbol{\epsilon'_m \epsilon_m} - \boldsymbol{\epsilon_m} X_m \boldsymbol{D_c^{-1}} \boldsymbol{\Xi_{c,m}^{-1}} \boldsymbol{D_c^{-1}} X_m' \boldsymbol{\epsilon_m}\right)' / (m-p).$$

As $c \to \infty$, the main terms of

$$\boldsymbol{\Xi_{c,\lfloor \sqrt{c}t \rfloor}} \to \begin{pmatrix} F(t)/\Psi^2(1) & \mathbf{0} \\ \mathbf{0} & t\boldsymbol{\Gamma_v} \end{pmatrix},$$

and

$$\boldsymbol{D_c^{-1}} X'_{\lfloor \sqrt{c}t \rfloor} \boldsymbol{\epsilon_{\lfloor \sqrt{c}t \rfloor}} \to \begin{pmatrix} \left(J(t) - \delta F(t)\right)/\Psi(1) \\ \boldsymbol{W}^{(1)}(t) \end{pmatrix}.$$

Stopping time $\hat{\tau}_c$ defined in (39) can be represented as

$$\frac{\hat{\tau}_c}{\sqrt{c}} = \inf\left\{ t : \frac{\lfloor \sqrt{c}t \rfloor - p}{\sqrt{c}} \frac{1}{c\sigma^2} \sum_{n=p+1}^{\lfloor \sqrt{c}t \rfloor} \left(\hat{\Psi}_{\lfloor \sqrt{c}t \rfloor}(L) x_{n-1}\right)^2 \geq \frac{\lfloor \sqrt{c}t \rfloor - p}{\sqrt{c}} \frac{s^2_{\lfloor \sqrt{c}t \rfloor}}{\sigma^2} \right\}.$$

As $c \to 0$,

$$\left|\boldsymbol{\Xi_{c,\lfloor \sqrt{c}t \rfloor}}\right| \to F(t) t \left|\boldsymbol{\Gamma_v}\right| / \Psi^2(1).$$

Since the right side close to 0, as $t \downarrow 0$. Multiplying both sides of the inequality in stopping time $\hat{\tau}_c$ by $\left|\boldsymbol{\Xi_{c,\lfloor \sqrt{c}t \rfloor}}\right|^2$, as $c \to \infty$, the right side

$$\frac{\lfloor \sqrt{c}t \rfloor - p}{\sqrt{c}} \left|\boldsymbol{\Xi_{c,\lfloor \sqrt{c}t \rfloor}}\right|^2 \frac{1}{c\sigma^2} \sum_{n=p+1}^{\lfloor \sqrt{c}t \rfloor} \left(\hat{\Psi}_{\lfloor \sqrt{c}t \rfloor}(L) x_{n-1}\right)^2$$
$$\to t\left(F(t) t \left|\boldsymbol{\Gamma_v}\right| / \Psi^2(1)\right)^2 F(t)$$

and the left side

$$\frac{\lfloor \sqrt{c}t \rfloor - p}{\sqrt{c}} \left|\boldsymbol{\Xi_{c,\lfloor \sqrt{c}t \rfloor}}\right|^2 \frac{s^2_{\lfloor \sqrt{c}t \rfloor}}{\sigma^2} \to t\left(F(t) t \left|\boldsymbol{\Gamma_v}\right| / \Psi^2(1)\right)^2.$$

We have

$$
\begin{aligned}
\hat{\tau}_c/\sqrt{c} &\to \inf\left\{t : t\left(F(t)t\left|\boldsymbol{\Gamma_v}\right|/\Psi^2(1)\right)^2 F(t) = t\left(F(t)t\left|\boldsymbol{\Gamma_v}\right|/\Psi^2(1)\right)^2\right\} \\
&= \inf\left\{t : F(t) = 1\right\} = U_1,
\end{aligned}
$$

where U_1 is defined in (41).

As for $\hat{\delta}_{\hat{\tau}_c} = \sqrt{c}\hat{\phi}^c_{1,\hat{\tau}_c}/\hat{\Psi}_{\hat{\tau}_c}(1) \Rightarrow \delta + B_1$, by the fact $\int_0^{U_1} Z_u^2 du = \langle M \rangle_{U_1} = 1$ and using DDS theorem $\int_0^{U_1} Z_u dW_u = M_{U_1} = B_1$, we have

$$
\begin{aligned}
\hat{\delta}_{\hat{\tau}_c} = \sqrt{c}\hat{\phi}^c_{1,\hat{\tau}_c}/\hat{\Psi}_{\hat{\tau}_c}(1) &= \frac{\sum_{n=p+1}^{\hat{\tau}_c} x_{n-1}\epsilon_n/\sqrt{c}}{\sum_{n=p+1}^{\hat{\tau}_c} x_{n-1}^2/c} \frac{1}{\hat{\Psi}_{\hat{\tau}_c}(1)} \\
&\to \frac{J(\hat{\tau}_c/\sqrt{c})}{F(\hat{\tau}_c/\sqrt{c})/\Psi^2(1)} \frac{1}{\Psi^2(1)} \\
&= \frac{\delta + \int_0^{U_1} Z_u dW_u}{\int_0^{U_1} Z_u^2 du} \quad a.s. \\
&= \delta + B_1.
\end{aligned} \tag{89}
$$

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

$$
Z_u^2 = 2\int_0^u Z_t dZ_t + u = 2\delta\int_0^u Z_t^2 dt + 2\int_0^u Z_t dW_t + u. \tag{90}
$$

Letting $u = U_s$, we have

$$
Z_{U_s}^2 = 2\delta\int_0^{U_s} Z_t^2 dt + 2\int_0^{U_s} Z_t dW_t + U_s = 2\delta s + 2B_s + U_s.
$$

Thus

$$
dU_s/ds = 1/\left(2\delta s + 2B_s + U_s\right).
$$

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

$$
d\rho_s = (\delta + \frac{1}{4\rho_s})ds + dB_s. \tag{91}
$$

This indicates that ρ_s is the Bessel process of dimension $3/2$ with a drift δ and a initial value $\rho_0 = 0$. Then, we have

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

(45) in Theorem 2 is shown as follows. Let $\boldsymbol{D_{\hat{\tau}_c}}$ be the $p \times p$ matrix $diag\left(\sqrt{c}, \sqrt{\hat{\tau}_c}, \ldots, \sqrt{\hat{\tau}_c}\right)$. Multiplying the Normal Equation by $\boldsymbol{D^{-1}_{\hat{\tau}_c}}$, we have

$$
\boldsymbol{D^{-1}_{\hat{\tau}_c}} X'_{\hat{\tau}_c} X_{\hat{\tau}_c} \boldsymbol{D^{-1}_{\hat{\tau}_c}} \boldsymbol{D_{\hat{\tau}_c}} \left(\hat{\boldsymbol{\phi}}^{\boldsymbol{c}}_{\hat{\boldsymbol{\tau}}_{\boldsymbol{c}}} - \boldsymbol{\phi^c}\right) = \boldsymbol{D^{-1}_{\hat{\tau}_c}} X'_{\hat{\tau}_c} \boldsymbol{\epsilon_{\hat{\tau}_c}}. \tag{92}
$$

Letting $t = \hat{\tau}_c/\sqrt{c}$ in Proposition 1, as $c \to \infty$, the convergence of $\boldsymbol{D}_{\hat{\boldsymbol{\tau}}_c}^{-1} X'_{\hat{\tau}_c} X_{\hat{\tau}_c} \boldsymbol{D}_{\hat{\boldsymbol{\tau}}_c}^{-1}$ is

$$\begin{pmatrix} \sum_{n=p+1}^{\hat{\tau}_c} x_{n-1}^2/c & \left(\sum_{n=p+1}^{\hat{\tau}_c} x_{n-1}\Delta x_{n-j}/\sqrt{c\hat{\tau}_c} \right)_j \\ \left(\sum_{n=p+1}^{\hat{\tau}_c} x_{n-1}\Delta x_{n-i}/\sqrt{c\hat{\tau}_c} \right)_i' & \left(\sum_{n=p+1}^{\hat{\tau}_c} \Delta x_{n-i}\Delta x_{n-j}/\hat{\tau}_c \right)_{i,j} \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \sigma^2 \int_0^{U_1} Z^2(u)du/\Psi(1)^2 & \mathbf{0}' \\ \mathbf{0} & \left(\sigma^2 \gamma_v\left(|i-j|\right)\right)_{i,j} \end{pmatrix}, \tag{93}$$

and the convergence of $\boldsymbol{D}_{\hat{\boldsymbol{\tau}}_c}^{-1} X'_{\hat{\tau}_c} \boldsymbol{\epsilon}_{\hat{\boldsymbol{\tau}}_c}$ is

$$\begin{pmatrix} \sum_{n=p+1}^{\hat{\tau}_c} x_{n-1}\epsilon_n/\sqrt{c} \\ \left(\sum_{n=p+1}^{\hat{\tau}_c} \Delta x_{n-i}\epsilon_n/\sqrt{\hat{\tau}_c} \right)_i' \end{pmatrix} \Rightarrow \begin{pmatrix} \sigma^2 \int_0^{U_1} Z(u)dW(u)/\Psi(1) \\ \sqrt{\gamma_v(0)}\sigma \boldsymbol{W}^{(1)}(1) \end{pmatrix}. \tag{94}$$

where $i, j = 1, \ldots, p-1$. (45) could be computed through (92).

Finally, we prove $s_{\hat{\tau}_c}^2 \to_p \sigma^2$. Setting $m = \hat{\tau}_c$ and substituting $\boldsymbol{\Delta x_m} = \boldsymbol{X_m \phi^c} + \boldsymbol{\epsilon_m}$ into (24), we have

$$\begin{aligned} s_{\hat{\tau}_c}^2 &= \left(\boldsymbol{\Delta x}_{\hat{\boldsymbol{\tau}}_c} - \boldsymbol{X}_{\hat{\boldsymbol{\tau}}_c}\hat{\boldsymbol{\phi}}_{\hat{\boldsymbol{\tau}}_c}^{\boldsymbol{c}}\right)' \left(\boldsymbol{\Delta x}_{\hat{\boldsymbol{\tau}}_c} - \boldsymbol{X}_{\hat{\boldsymbol{\tau}}_c}\hat{\boldsymbol{\phi}}_{\hat{\boldsymbol{\tau}}_c}^{\boldsymbol{c}}\right)/(\hat{\tau}_c - p) \\ &= \left(\boldsymbol{X}_{\hat{\boldsymbol{\tau}}_c}\left(\boldsymbol{\phi^c} - \hat{\boldsymbol{\phi}}_{\hat{\boldsymbol{\tau}}_c}^{\boldsymbol{c}}\right) + \boldsymbol{\epsilon}_{\hat{\boldsymbol{\tau}}_c}\right)' \left(\boldsymbol{X}_{\hat{\boldsymbol{\tau}}_c}\left(\boldsymbol{\phi^c} - \hat{\boldsymbol{\phi}}_{\hat{\boldsymbol{\tau}}_c}^{\boldsymbol{c}}\right) + \boldsymbol{\epsilon}_{\hat{\boldsymbol{\tau}}_c}\right)/(\hat{\tau}_c - p) \\ &= \{ \left(\boldsymbol{\phi^c} - \hat{\boldsymbol{\phi}}_{\hat{\boldsymbol{\tau}}_c}^{\boldsymbol{c}}\right)' \boldsymbol{D}_{\hat{\boldsymbol{\tau}}_c}\boldsymbol{D}_{\hat{\boldsymbol{\tau}}_c}^{-1}\boldsymbol{X}'_{\hat{\boldsymbol{\tau}}_c}\boldsymbol{X}_{\hat{\boldsymbol{\tau}}_c}\boldsymbol{D}_{\hat{\boldsymbol{\tau}}_c}^{-1}\boldsymbol{D}_{\hat{\boldsymbol{\tau}}_c}\left(\boldsymbol{\phi^c} - \hat{\boldsymbol{\phi}}_{\hat{\boldsymbol{\tau}}_c}^{\boldsymbol{c}}\right) \\ &+ 2\left(\boldsymbol{\phi^c} - \hat{\boldsymbol{\phi}}_{\hat{\boldsymbol{\tau}}_c}^{\boldsymbol{c}}\right) D_{\hat{\tau}_c} D_{\hat{\tau}_c}^{-1} \boldsymbol{X}'_{\hat{\boldsymbol{\tau}}_c}\boldsymbol{\epsilon}_{\hat{\boldsymbol{\tau}}_c} + \boldsymbol{\epsilon}'_{\hat{\boldsymbol{\tau}}_c}\boldsymbol{\epsilon}_{\hat{\boldsymbol{\tau}}_c} \}/(\hat{\tau}_c - p) \\ &\to_p \sigma^2. \end{aligned}$$

□

6.3 Proof of Theorem 3

Proof. Under null hypothesis, according to (40) $M_u = \int_0^u W_s dW_s$. Using a time-change $v = \langle M \rangle_u = \int_0^u W_s^2 ds$ and $dv = W_u^2 du$,

$$\int_0^\infty e^{-\gamma v} P\left\{U_v \le t, 2\rho_v \le y\right\} dv = E\left[\int_0^\infty e^{-\gamma \langle M \rangle_u} 1\left\{U_{\langle M \rangle_u} \le t, W_{U_{\langle M \rangle_u}}^2 \le y\right\} W_u^2 du\right]$$
$$= \int_0^t E\left[e^{-\gamma \langle M \rangle_u} 1\left\{W_u^2 \le y\right\} W_u^2\right] du.$$

By the fundamental theorem of calculus

$$\int_0^{\infty} e^{-\gamma v} \frac{\partial^2}{\partial y \partial t} P\left\{U_v \le t, 2\rho_v \le y\right\} dv = \frac{\partial}{\partial y} E\left[e^{-\gamma \langle M \rangle_t} 1\left\{W_t^2 \le y\right\} W_t^2\right]$$
$$= \int_0^{\infty} e^{-\gamma v} f_{\langle M \rangle_t, W_t^2}(v, y)\, y d\nu,$$

where $f_{\langle M \rangle_t, W_t^2}(v, y)$ is the density function of $\left(\langle M \rangle_t, W_t^2\right)$. Hence,

$$\frac{\partial^2}{\partial y \partial t} P\left\{U_v \le t, 2\rho_v \le y\right\} = f_{\langle M \rangle_t, W_t^2}(v, y)\, y.$$

According to the theory of Bessel Bridge in Pitman and Yor (1982)

$$E\left[\exp\left(-\frac{b^2}{2}\int_0^u W_s^2 ds | W_u^2 = y\right)\right]$$
$$= \lim_{x \to 0} \frac{bu}{\sinh(bu)} \exp\left(\frac{y}{2u}\left(1 - bu \coth(bu)\right)\right) \frac{I_{-1/2}\left(\frac{b\sqrt{xy}}{\sinh(bu)}\right)}{I_{-1/2}\left(\frac{\sqrt{xy}}{u}\right)} \tag{95}$$
$$= \left(\frac{bu}{\sinh(bu)}\right)^{1/2} \exp\left(\frac{y}{2u}\left(1 - bu \coth(bu)\right)\right), \tag{96}$$

where I_ν is the modified Bessel function defined as $I_\nu(x) = \sum_{k=0}^{\infty}(x/2)^{2k+\nu}/(k!\Gamma(k+\nu+1))$ for $\nu \ge -1$. Since W_u^2 has the marginal density

$$f_{W_u^2}(y) = y^{-1/2} e^{-y/(2u)} / \left(\Gamma(1/2)(2u)^{1/2}\right), \tag{97}$$

then

$$\int_0^{\infty} e^{-\gamma v} f_{\langle M \rangle_u, W_u^2}(\nu, y) y d\nu = \frac{y^{1/2}}{\sqrt{2\pi}} \left(\frac{\sqrt{2\gamma}}{\sinh\left(u\sqrt{2\gamma}\right)}\right)^{1/2} \exp\left(-\frac{y}{2}\sqrt{2\gamma}\coth\left(u\sqrt{2\gamma}\right)\right).$$

Since $2\rho_v = 2B_v + U_v$, one can obtain (46) from the expression of the es function in Borodin and Salminen (2002) . □

6.4 Proof of Theorem 5 and Theorem 6

Proof. Assuming $\gamma > 0, \alpha > 0, t > 0$, and integrating (96) with (97), we have

$$E_{H_0}\left[e^{-\gamma <M>_t - \alpha W_t^2}\right] = \int_0^{\infty} E_{H_0}\left[e^{-\gamma <M>_t} | W_t^2 = y\right] e^{-\alpha y} f_{W_t^2}(y) dy$$
$$= \frac{2^{1/4}\gamma^{1/4}}{\sqrt{\sqrt{2\gamma}\cosh\left(t\sqrt{2\gamma}\right) + 2\alpha \sinh\left(t\sqrt{2\gamma}\right)}}.$$

Using the time change $t = U_v$ and the above expression

$$\begin{aligned}
&\int_0^\infty e^{-\gamma v} E_{H_0}\left[\exp\left(-\alpha 2\rho_v - \beta U_v\right)/2\rho_v\right] dv \\
=&\int_0^\infty e^{-\beta t} E_{H_0}\left[e^{-\gamma \langle M\rangle_t - \alpha W_t^2}\right] dt \qquad (98)\\
=&\int_0^\infty e^{-\beta t} \frac{2^{1/4}\gamma^{1/4}}{\sqrt{\sqrt{2\gamma}\cosh(t\sqrt{2\gamma}) + 2\alpha\sinh(t\sqrt{2\gamma})}} dt \\
=&\int_0^1 \frac{s^{-3/4+\beta/(2\sqrt{2\gamma})}}{2^{3/4}\gamma^{1/4}\sqrt{2\alpha(1-s)+\sqrt{2\gamma}(s+1)}} ds \equiv \int_0^1 H\left(s,\alpha,\beta,\gamma\right) ds. \qquad (99)
\end{aligned}$$

Here we made a variable change $t = -\log(s)/(2\sqrt{2\gamma})$ in the third equation. Since

$$\frac{\partial^{k+l} H}{\partial\alpha^k \partial\beta^l}\left(s,0,0,\gamma\right) = \frac{2^{-\frac{k}{2}-\frac{3l}{2}-1}\left(-\frac{1}{2}\right)^{(k)}(1-s)^k(1+s)^{-k-\frac{1}{2}}\log^l(s)}{s^{3/4}} \gamma^{-\frac{1}{2}(k+l+1)}.$$

and

$$\mathscr{L}_\gamma{}^{-1}\left(\gamma^{-(j+1)/2}\right)(v) = v^{(j-1)/2}/\Gamma\left(\left(j+1\right)/2\right),$$

$\mathscr{L}_\gamma{}^{-1}\left(\partial^{k+l} H\left(s,0,0,\gamma\right)/\partial\alpha^k\partial\beta^l\right)(v)$ can be written as $J(t,v,k,l)$ in (53). Use Girsanov's theorem to obtain (54). □