SELF-WEIGHTED ESTIMATION FOR LOCAL UNIT ROOT REGRESSION WITH APPLICATIONS

By

Zhishui Hu, Nan Liu, Peter C. B. Phillips, Qiying Wang

April 2024

COWLES FOUNDATION DISCUSSION PAPER NO. 2400

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Self-weighted Estimation for Local Unit Root Regression with Applications [*]

Zhishui Hu
University of Science and Technology of China

Nan Liu
Xiamen University

Peter C. B. Phillips
Yale University, University of Auckland & Singapore Management University

Qiying Wang
University of Sydney

April 20, 2024

Abstract

A new self-weighted least squares (LS) estimation theory is developed for local unit root (LUR) autoregression with heteroskedasticity. The proposed estimator has a mixed Gaussian limit distribution and the corresponding studentized statistic converges to a standard normal distribution free of the unknown localizing coefficient which is not consistently estimable. The estimator is super consistent with a convergence rate slightly below the $O_P(n)$ rate of LS estimation. The asymptotic theory relies on a new framework of convergence to the local time of a Gaussian process, allowing for the sample moments generated from martingales and many other integrated dependent sequences. A new unit root (UR) test in augmented autoregression is developed using self-weighted estimation and the methods are employed in predictive regression, providing an alternative approach to IVX regression. Simulation results showing good finite sample performance of these methods are reported together with a small empirical application.

JEL Classification: C13, C22.

Key words and phrases: Self-weighted least squares estimation, autoregression, super consistency, limit distribution, unit root test, predictive regression.

1 Introduction

Consider the local unit root (LUR) autoregression

$$y_k = \alpha y_{k-1} + u_k, \quad \alpha = 1 + \tau/n, \quad k = 1, 2, ..., n \tag{1.1}$$

[*]Phillips acknowledges research support of a Kelly fellowship at the University of Auckland. Address correspondence to Qiying Wang, School of Mathematics and Statistics, The University of Sydney, NSW 2006, Australia; e-mail: qiying.wang@sydney.edu.au.

with initialization $y_0 = O_P(1)$, innovations $\{u_k\}_{k\geq 1}$ and constant localizing coefficient $\tau \in \mathbb{R}$. Let $\widehat{\alpha}_n = \frac{\sum_{k=1}^{n} y_{k-1}y_k}{\sum_{k=1}^{n} y_{k-1}^2}$ be the least squares (LS) estimator of α. When $u_k \sim_{iid} (0, \sigma^2 > 0)$ it is well-known (**????**) that

$$n\left(\widehat{\alpha}_n - \alpha\right) \quad \to_D \quad \Big[\int_0^1 (J_\tau(r))^2\, dr\Big]^{-1} \int_0^1 J_\tau(r) dB_r, \tag{1.2}$$

where $J_\tau = \{J_\tau(t) = \int_0^t e^{(t-s)\tau} dB_s\}_{t\geq 0}$ (here and below) is a linear diffusion (Ornstein-Uhlenbeck) process satisfying $dJ_\tau(t) = \tau\, J_\tau(t)dt + dB_t$, with $J_\tau(0) = 0$ and $B = \{B_t\}_{t\geq 0}$ is a standard Brownian motion.

In spite of the super consistency of (1.2) the limit distributions of $\widehat{\alpha}_n$ and the corresponding studentized statistic are non-pivotal and depend on the unknown localizing coefficient τ, which is identified but not consistently estimable, as is clear from the original limit theory and much later commentary and analysis (**?????**). To assist in resolving this difficulty, many different approaches have been introduced in the literature. The first-difference method in **?** has virtually no finite sample bias, is insensitive to initial conditions, and Gaussian limit theory applies continuously as α passes through unity, but with a uniform $\sqrt{n}$ convergence rate. This rate of convergence can be improved by aggregating moment conditions in differences, as in **?**, leading to an estimator that has a limiting normal distribution uniformly over stationary and unit root cases with a rate of convergence within a slowly varying factor of n when $\alpha = 1$.

In addition to this work, several instrumental variable (IV) estimators have appeared in the literature as alternatives to standard LS. In particular, **?** suggested the use of the Cauchy estimator, using the sign function as an IV. **?** generalized that approach by considering weighted estimation in model (1.1). An interesting feature of **?** is that the studentized statistic based on the weighted LS estimator in model (1.1) has a normal limit distribution free of the parameter τ, enabling feasible inference concerning α. But overweighting in that approach loses the super consistency of LS estimation in LUR regression. Using a different methodology altogether, **?** introduced the IVX approach, which was developed further in later work by **?**. The IVX method, named for its use of instruments that are constructed endogenously, utilizes linear filtering of the regressors that mildly attenuates their signal to produce new instrumental variables which enable IV estimation with a mixed Gaussian limit theory and a student t statistic with a standard normal limit distribution, thereby facilitating inference. For recent work on the IVX method, we refer to **?**, **?**, **?**, **??** and **?**.

The present paper proposes a new estimator that has a mixed Gaussian limit by using self-weighted estimation in (1.1). Noting that $y_n/\sqrt{n} = O_P(1)$, the key idea in our approach is that the function $I(|y_{k-1}| \leq b_n\sqrt{n})$, where $b_n \to 0$, can be used as a weight to reduce the signal in ordinary LS estimation theory so that Wang's extended martingale limit theorem (?; ?, Theorem 3.13) applies to establish the asymptotics. More explicitly, this paper investigates the self-weighted LS estimator $\widehat{\alpha}_{1n}$ of α defined by

$$\widehat{\alpha}_{1n} = \arg\min_{\alpha} \sum_{k=1}^{n} (y_k - \alpha y_{k-1})^2 \, I(|y_{k-1}| \le b_n \sqrt{n}), \tag{1.3}$$

for some positive numerical sequence $b_n \to 0$. We show that $\widehat{\alpha}_{1n}$ has a mixed Gaussian limit and the corresponding studentized statistic converges to a normal limit free of the parameter τ. This new estimator $\widehat{\alpha}_{1n}$ retains superconsistency although the convergence rate is slightly slower than n. The reduction in the convergence rate depends on the rate at which $b_n \to 0$. Indeed, noting that $y_{\lfloor nt \rfloor}/\sqrt{n}\sigma \Rightarrow J_\tau(t)$ on $D[0,1]$, the limit behaviour of $\widehat{\alpha}_{1n}$, when $b_n = b > 0$, is similar to that of $\widehat{\alpha}_n$, so standard normal limit theory no longer applies.

A major obstacle in deriving asymptotics for $\widehat{\alpha}_{1n}$ is the development of limit theory for sample moments of the form

$$V_n^2 := \frac{1}{n^2 b_n^3} \sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \le b_n \sqrt{n}) = \frac{c_n}{n} \sum_{k=1}^{n} f(c_n y_{nk}),$$

where $y_{nk} = y_{k-1}/\sqrt{n}$, $c_n = b_n^{-1}$ and $f(x) = x^2 I(|x| \le 1)$. When y_{nk} is a martingale array as here, new asymptotics are required for the sample moment V_n^2 since previous results, such as those in Theorem 2.1 of ? are unsuitable, as explained in Section 2. Another contribution of this paper is to establish convergence of V_n^2 under an alternative used in ?. The new condition imposed on y_{nk} is easy to verify for martingales and many other integrated dependent sequences by using well-developed strong approximation theory. There is a trade off between our new general condition on y_{nk} and the numerical sequence c_n required in V_n^2. Compared with ? the restriction on c_n is stronger, but it is sufficient for the purposes of this paper, as explained in Section 2.

The paper proceeds as follows. Section 2 explores convergence to local time of a Gaussian process, giving a significant extension of ? by allowing for martingales and many other integrated dependent sequences. Self-weighted LS estimation theory for near unit root autoregression with heteroskedasticity is developed in Section 3, providing a new unit root test. The use of self-weighted instrumentation in predictive regression is developed in Section 4. Simulations and a brief empirical application are reported in Section 5. Proofs are collected in Section 7. Throughout the paper, we use $C, C_1, C_2, ...$ for constants, which may be different at each appearance.

2 Convergence to local time by strong approximation

This section establishes convergence to the local time process of a Gaussian process by using strong approximation techniques. The result is of independent interest and this section may be read separately. The notation therefore differs slightly from other sections to allow for more general applications.

2.1 Preliminaries

Let $X = \{X_t\}_{t\geq 0}$ be a real-valued stochastic process. A measurable process $L_X = \{L_X(t,x)\}_{t\geq 0, x\in\mathbb{R}}$ is said to be a local time of X if it satisfies

$$\int_0^t h(X_s)ds = \int_{-\infty}^{\infty} h(x)L_X(t,x)dx,$$

for all positive (or bounded) Borel functions $h : \mathbb{R} \to \mathbb{R}$. For a zero mean Gaussian process X with covariance kernel $\mathrm{r}(u,v) = E\big(X_u X_v\big)$ satisfying for all $T > 0$

$$\int_0^T \int_0^T [\mathrm{r}(u,u)\mathrm{r}(v,v) - \mathrm{r}^2(u,v)]^{-1/2} du dv < \infty, \tag{2.1}$$

X has a local time L_X which can be represented as

$$L_X(t,x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux}\int_0^t e^{iuX_s} ds\, du,$$

where the right-hand side is understood in the framework of L^2-theory, i.e., as a limit of $L^N(t,x)$ in the sense that

$$\lim_{N\to\infty} \mathrm{E}\,|L^N(t,x) - L_X(t,x)|^2 = 0,$$

where

$$L^N(t,x) = \frac{1}{2\pi}\int_{-N}^{N} e^{-iux}\int_0^t e^{iuX_s} ds\, du.$$

With this convention, for any bounded Borel function $f(t)$ on $[0,1]$, $\int_0^1 f(u)L_X(du,x)$ is well defined and

$$\begin{aligned}\int_0^1 f(u)L^N(du,x) &= \frac{1}{2\pi}\int_{-N}^{N} e^{-itx}\int_0^1 f(u)e^{itX_u} du\, dt \\ &\to \int_0^1 f(u)L_X(du,x),\end{aligned} \tag{2.2}$$

in probability, as $N \to \infty$. Furthermore, if $g(t,x)$ is a Borel function on $[0,1]\times R$ satisfying $\int_{-\infty}^{\infty} \sup_{0\leq s\leq 1}|g(s,x)|dx < \infty$ and $L_X(t,x)$ is continuous, we have

$$\begin{aligned}c_n\int_0^1 g(t, c_n X_t)dt &= \int_0^1\int_{-\infty}^{\infty} g(t,x)L_X(dt,\, x/c_n)dx \\ &\to \int_0^1 G(t)L_X(dt,0), \quad a.s.,\end{aligned} \tag{2.3}$$

for any $0 < c_n \to \infty$, where $G(t) = \int_{-\infty}^{\infty} g(t, x)dx$. See, e.g. **??** and also Chapter 2 of **?**.

The linear diffusion $J_\tau = \{J_\tau(t) = \int_0^t e^{(t-s)\tau} dB_s\}_{t\geq 0}$ in (1.2) is a Gaussian process that satisfies (2.1). Another Gaussian process satisfying (2.1) is fractional Brownian motion $B_H = \{B_H(t)\}_{t\geq 0}$ defined by

$$B_H(t) = \kappa_H \int_{-\infty}^{t} \big[((t-u)_+)^{H-1/2} - ((-u)_+)^{H-1/2}\big] dB_u,$$

where $a_+ = \max\{a, 0\}$, $0 < H < 1$, and $\kappa_H > 0$ is a constant such that $\mathrm{E}B_H^2(1) = 1$. Both J_τ and B_H have local time processes satisfying (2.3).

2.2 Convergence to local time

Let $\{X_{nk}\}_{n\geq 1, k\geq 1}$ be an arbitrary random array defined on $\{\Omega, \mathcal{F}, P\}$ for which $X_{n,\lfloor nt \rfloor} \Rightarrow X_t$ on $D[0, 1]$. Note that

$$S_n \;\; := \;\; \frac{c_n}{n} \sum_{k=1}^{n} g\Big(\frac{k}{n}, c_n X_{nk}\Big) = \int_0^1 c_n g(\lfloor nt \rfloor/n, c_n X_{n,\lfloor nt \rfloor}) dt + o_P(1),$$

for any $c_n/n \to 0$ and bounded function $g(t, x)$. Similar to (2.3), in applications it is often desirable to establish that

$$S_n \;\; \to_D \;\; \int_0^1 G(t)\, L_X(dt, 0),$$

for certain $0 < c_n \to \infty$. In this regard, a framework was given in **?**, allowing for the array y_{nk} to have a certain structure depending on conditional arguments (see, also, Chapter 2.3 of **?**). Theorem 2.1 of **?** is quite general, so that the main results given by **?**, **?**, **?** and **?** are all covered, but it is difficult to verify their Assumption 2.3 in applications. To see this, let $d_{nkj}^2 = \mathrm{Var}(y_{nk} - y_{nj})$ and $\mathcal{F}_{nk} = \sigma(y_{n1}, ..., y_{nk})$. In addition to some regularity conditions on d_{nkj}, Assumption 2.3 of **?** imposed the following condition on the array y_{nk}:

> *y_{nk} is adapted to $\mathcal{F}_{nk}$ and, conditional on $\mathcal{F}_{nj}$, $(y_{nk} - y_{nj})/d_{nkj}$ has a density $h_{nkj}(x)$ which is uniformly bounded by a constant K.*

Except for explicit arrays y_{nk} of special structure such as standardized partial sums of linear processes with i.i.d. innovations, the existence of a conditional density can hardly be checked directly or is simply impossible to verify where y_{nk} is a martingale array of the type considered in the present paper.

This section provides an alternative to Assumption 2.3 of **?**. The new condition imposed on X_{nk} is easy to verify for martingales and many other dependent sequences by using well-developed strong approximation theory. We have the following explicit result.

Theorem 2.1. *Suppose that*

(a) on a richer probability space, there exists a Gaussian process X satisfying (2.1) and a random array $\{X^*_{nk}\}_{\substack{1\le k\le n\\ n\ge 1}}$ such that, for each $n \geq 1$, $\{X_{n1}, ..., X_{nn}\} =_d \{X^*_{n1}, ..., X^*_{nn}\}$ and

$$\sup_{0\le t\le 1} |X^*_{n,\lfloor nt\rfloor} - X_t| = o_P(c_n^{-2}), \tag{2.4}$$

where $0 < c_n \to \infty$ is a sequence of constants satisfying $c_n/n \to 0$;

(b) for any small $\delta > 0$, there exists two continuous functions $g_{i\delta}(s,x), i = 1,2,$ on $[0,1]\times\mathbb{R}$ satisfying $\sup_{0\le s\le 1} |g_{i\delta}(s,x)| \le C_\delta/(1+|x|^{1+b})$ for some $b > 0$ and $C_\delta > 0$ depending only on δ such that $g_{1\delta}(s,x) \le g(s,x) \le g_{2\delta}(s,x)$ and

$$\int_{-\infty}^{\infty} \sup_{0\le s\le 1} |g_{i\delta}(s,x) - g(s,x)|dx \le \delta.$$

Then, on $D_{\mathbb{R}^2}[0,1]$, we have

$$\big(X_{n,\lfloor nt\rfloor}, S_n\big) \quad\Rightarrow\quad \Big(X_t,\ \int_0^1 G(s)L_X(ds,0)\Big), \tag{2.5}$$

where $G(s) = \int_{-\infty}^{\infty} g(s,x)dx$.

Remark 2.1. The approximation condition (2.4) can be established using a Skorohod embedding (together with some minor modifications) for various dependent random arrays including martingales, partial sums of mixing sequences and nonlinear causal time series. Examples will be given in the following sections. It should be mentioned that **?** and Theorem 2.19 of **?** established versions of (2.5) suited to nonparametric cointegrating regression for any $0 < c_n \to \infty$ satisfying $c_n/n \to 0$. For the condition imposed on X_{nk} to be useful in applications there is a trade off on c_n depending on (2.4) in Theorem 2.1. To provide an illustration, we assume $X_{nk} = \frac{1}{\sqrt{n}}\sum_{j=1}^{k} u_k$ where $u_k \sim \mathcal{N}(0,1)$ are i.i.d. random variables. In this case, there exists a standard Brownian motion $B = \{B_t\}_{t\ge 0}$ so that $\{X_{n1}, ..., X_{nn}\} =_d \{\frac{1}{\sqrt{n}}B_1, , ..., \frac{1}{\sqrt{n}}B_n\}$ for each $n \geq 1$. Hence, by noting that $\sqrt{n}\sup_{0\le t\le 1} \big|B_t - B_{[nt]}/\sqrt{n}\big| \to_P \infty$, to ensure (2.4) true with X_t being replaced by $B(t)$, we have $\sqrt{n}c_n^{-2} \to \infty$, and so we require $c_n/n^{1/4} \to 0$ and $c_n \to \infty$. Such a rate requirement on c_n is insufficient for nonparametric cointegrating regression, as seen in **??** and related papers, but it is sufficient for our purpose in this paper.

Remark 2.2. Without loss of generality, we may assume in applications that X^*_{nk} given in condition (a) is the array X_{nk} itself. A simple modification to the proof shows that (2.5) still holds in case X is replaced by B if, instead of (a), the following condition (a)* is used:

(a)* on a richer probability space, there exists a standard Brownian motion $B = \{B_t\}_{t\geq 0}$ so that

$$\max_{1\leq k\leq n} \left| X_{nk} - B_k/\sqrt{n} \right| = o_P(c_n^{-2}), \tag{2.6}$$

where $0 < c_n \to \infty$ is a sequence of constants satisfying $c_n/n \to 0$.

Result (2.6) is widely studied in probability theory and the convergence in probability given in (2.6) can be strengthened to almost sure convergence in various situations when additional conditions such as finite higher moment requirements are included, as in Lemma 3.1 of **?**.

Remark 2.3. Condition (b) is usually satisfied if $g(s, x)$ is Riemann integrable on the space $[0, 1] \times \mathbb{R}$. If no additional smoothness condition on X_{nk} is imposed, the requirement that $\sup_{0\leq s\leq 1} |g(s, x)| \leq C/(1 + |x|^{1+b})$ cannot be materially improved. An example similar to Example IV. 2.3 of **?** can be constructed to show this in the present case.

2.3 Extension to more general settings

We now consider an extension of (2.5) that can be used for many different purposes. Let v_k be a sequence of arbitrary positive random variables and define the sample covariance

$$S_{1n} = \frac{c_n}{n} \sum_{k=1}^{n} g\Big(\frac{k}{n}, c_n X_{nk}\Big)\, v_k.$$

We have the following result for the asymptotics of S_{1n}.

Theorem 2.2. *In addition to the conditions of Theorem 2.1, suppose the following hold:*

(i) X is a Gaussian process satisfying

$$\sup_{\substack{|s-t|\leq \Delta \\ 0\leq t\leq 1}} |X_t - X_s| = O_P(\Delta^\eta), \quad \text{for some } 0 < \eta \leq 1;$$

(ii) $\sup_{k\geq 1} Ev_k < \infty$ and there exist $A_0 \in \mathbb{R}$ and $0 < m := m_n \to \infty$ satisfying $m = o(nc_n^{-2/\eta})$ where η is given in (i) so that

$$\max_{0\leq j\leq n-m} E\left| \frac{1}{m} \sum_{k=j+1}^{j+m} v_k - A_0 \right| = o(c_n^{-1});$$

(iii) $|g(s, x) - g(s', x)| \leq C\,|s - s'|/(1 + |x|^{1+b})$ on $[0, 1] \times \mathbb{R}$, where $b > 0$ is given as in condition (b) of Theorem 2.1.

Then, for any constant sequence $0 < c_n \to \infty$ *satisfying* $c_n/n \to 0$*, we have*

$$S_{1n} \;=\; A_0\, S_n + o_P(1). \tag{2.7}$$

Consequently, on $D_{\mathbb{R}^2}[0,1]$*, we have*

$$\big(X_{n,\lfloor nt\rfloor}, S_{1n}\big) \;\Rightarrow\; \Big(X_t,\ A_0 \int_0^1 G(s) L_X(ds, 0)\Big),$$

where $G(s) = \int_{-\infty}^{\infty} g(s,x)dx$*.*

Remark 2.4. For a version of (2.7) with $c_n = 1$, we refer to **?**, where the additional condition (i) is not required. When $c_n = 1$, we have

$$S_n = \frac{1}{n}\sum_{k=1}^{n} g\left(\frac{k}{n},\, X_{nk}\right) \to_D \int_0^1 g(t, X_t)dt,$$

providing a different limit distribution from (2.5).

Remark 2.5. Let w_k be a sequence of arbitrary random variables and write $w_k = w_k^+ - w_k^-$, where $w_k^+ = w_k I(w_k \geq 0)$ and $w_k^- = w_k I(w_k < 0)$. If the condition (ii) of Theorem 2.2 is replaced by the following (ii)$'$, the conclusion of Theorem 2.2 still holds with $A_0 = A_0^+ - A_0^-$ when v_k is replaced by w_k.

(ii)$'$ $\sup_{k\geq 1} E|w_k| < \infty$ and there exist A_0^+ and A_0^- so that

$$\max_{m\leq j\leq n-m} E\Big|\frac{1}{m}\sum_{k=j+1}^{j+m} w_k^+ - A_0^+\Big| = o(c_n^{-1}), \quad \max_{m\leq j\leq n-m} E\Big|\frac{1}{m}\sum_{k=j+1}^{j+m} w_k^- - A_0^-\Big| = o(c_n^{-1}),$$

where $0 < m := m_n \to \infty$ satisfying $m = o(nc_n^{-2/\eta})$ with η being given in (i).

In applications, we usually have $A_0^+ = \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^n Ew_j^+$ and $A_0^- = \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^n Ew_j^-$ so that $A_0 = \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^n Ew_j$.

2.4 Strong approximation for martingales

This section shows that (2.4) can be verified for a martingale sequence under certain regularity conditions. Let $\{u_k, \mathcal{F}_k\}_{k\geq 1}$ be a martingale difference sequence with $\sup_{k\geq 1} E|u_k|^p < \infty$ for some $2 < p \leq 4$. Let $S_n = \sum_{j=1}^n u_j$, $d_n^2 = \sum_{j=1}^n Eu_j^2$, $V_n^2 = \sum_{j=1}^n E\big(u_j^2|\mathcal{F}_{j-1}\big)$ and $\mathcal{L}_{np} = d_n^{-p}\sum_{j=1}^n E|u_j|^p$.

Proposition 2.1. *Suppose that, for some sequence* $0 < \delta_n \to 0$*,*

$$\max_{1\leq k\leq n} |V_k^2/d_n^2 - k/n| \;=\; O_P(\delta_n). \tag{2.8}$$

Then, on a richer probability space, there exist a standard Brownian motion $B = \{B_t\}_{t\geq 0}$ and a random arrry $\{S^*_{nk}, k = 1, 2, \cdots, n, n \geq 1\}$ such that $\{S^*_{nk}, k = 1, 2, \cdots, n\} =_d \{S_k, k = 1, 2, \cdots, n\}$ for all $n \geq 1$, and

$$\sup_{0\leq t\leq 1} \Big| \frac{1}{d_n} S^*_{n,\lfloor nt\rfloor} - B_t \Big| = O_P\big[(\delta_n^{1/2} + \mathcal{L}_{np}^{1/p}) \log^{1/2} n\big]. \tag{2.9}$$

The following result is an immediate consequence of Proposition 2.1.

Corollary 2.1. *Suppose that $Eu_j^2 = \sigma^2$ and $\sup_{k\geq 1} E|u_k|^{2+\delta} < \infty$ for some $0 < \delta \leq 2$. If $\frac{1}{n} \max_{1\leq k\leq n} \left|V_k^2 - k\sigma^2\right| = O_P(n^{-\delta/(2+\delta)})$, then, on a richer probability space, there exist a standard Brownian motion $B = \{B_t\}_{t\geq 0}$ and a random array $\{S^*_{nk}, k = 1, 2, \cdots, n, n \geq 1\}$ such that $\{S^*_{nk}, k = 1, 2, \cdots, n\} =_d \{S_k, k = 1, 2, \cdots, n\}$ for all $n \geq 1$, and*

$$\sup_{0\leq t\leq 1} \Big| \frac{1}{\sqrt{n}\sigma} S^*_{n,\lfloor nt\rfloor} - B_t \Big| = O_P\Big(n^{-\delta/(4+2\delta)} \log^{1/2} n\Big). \tag{2.10}$$

Remark 2.6. The convergence rate in (2.9) depends on the conditional variance V_k^2, matching the martingale weak convergence theory discussed in Chapter 4 of **?**. There is a simple sufficient condition to ensure (2.8). In fact, if $d_n^2/n \to \sigma^2$ for some $\sigma^2 > 0$, $n|\eta_n|$ is an eventually increasing sequence and $V_n^2/n = \sigma^2 + O(|\eta_n|), a.s.$, then (2.8) holds with $\delta_n = |\eta_n| + 1/n$, where $\eta_n = d_n^2/n - \sigma^2$. Further, if u_k is a sequence of stationary martingale differences, it automatically holds that $\eta_n = 0$ (so that $\delta_n = 1/n$) and $\mathcal{L}_{np} = O(n^{1-p/2})$, providing the optimal convergence rate. There are other strong approximation results for martingales, such as those in **?**, **?** and **?**. We may establish (2.9) by using martingale embedding methods. The proof of Proposition 2.1 is given in Section 7.1.

2.5 Other examples satisfying (2.6)

The strong approximation result (2.6) can be verified for various mixing sequences such as α-mixing sequences and causal time series. The following examples are well-studied in the literature.

Example 2.1. Let $\{u_i\}_{i\geq 0}$ be a sequence of stationary α-mixing random variables[1] with mean zero and coefficients $\alpha(n)$ satisfying the following conditions: for $r > p$ and $2 < p \leq 4$,

$$\sup_{x>0} x^r P(|u_0| \geq x) < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} n^{p-2} \alpha^{(r-p)/r}(n) < \infty.$$

[1]A sequence $\{\zeta_k, k \geq 1\}$ is said to be α-mixing if

$$\alpha(n) := \sup_{k\geq 1} \sup\{|P(AB) - P(A)P(B)| : A \in \mathcal{F}^\infty_{n+k}, B \in \mathcal{F}^k_1\}$$

converges to zero as $n \to \infty$, where $\mathcal{F}^m_l = \sigma\{\zeta_l, \zeta_{l+1}, \dots, \zeta_m\}$ denotes the σ-algebra generated by $\zeta_l, \zeta_{l+1}, \dots, \zeta_m$ with $l \leq m$.

Then, on a richer probability space, there exists a standard Brownian motion $B = \{B_t\}_{t\geq 0}$ such that

$$\max_{1\leq k\leq n} \Big| \frac{1}{\sqrt{n}\sigma} \sum_{j=1}^{k} u_j - \frac{1}{\sqrt{n}} B_k \Big| = O_P(n^{1/p-1/2} \log^{1/2-1/p} n), \tag{2.11}$$

where $\sigma^2 = Eu_0^2 + 2\sum_{k\geq 1} E(u_0 u_k)$.

For a proof of (2.11), we refer to **?** and **?**.

Example 2.2. Let $(\eta_i)_{i\in\mathbb{Z}}$ be a sequence of i.i.d. random variables. A stationary causal process $(u_k)_{k\in\mathbb{Z}}$ is defined by $u_k = F(..., \eta_{k-1}, \eta_k)$, where F is a measurable function such that u_k are well-defined stationary random variables with $Eu_0 = 0$ and $Eu_0^2 \in (0, \infty)$. Write $\mathcal{F}_k = \sigma(u_i, i \leq k)$,

$$\delta_{i,p} = \big|\big|u_i - u_i^{'}\big|\big|_p \quad \text{and} \quad \Theta_{j,p} = \sum_{i=j}^{\infty} \delta_{i,p},$$

where $u_k^{'} = F(..., \eta_{-1}, \eta_0^*, \eta_1, ..., \eta_k)$, and $\{\eta_k^*\}_{k\in\mathbb{Z}}$ is an independent copy of $\{\eta_k\}_{k\in\mathbb{Z}}$. Suppose that $E|u_0|^p < \infty$ for some $2 < p \leq 4$ and

(a) $\Theta_{j,p} = O(j^{-1} \log^{-1/p} j)$ when $2 < p < 4$;

(b) $\Theta_{j,p} = O(j^{-1} \log^{-A} j)$ with $A > 3/2$ when $p = 4$.

Then, on a richer probability space, there exists a standard Brownian motion $B = \{B_t\}_{t\geq 0}$ such that

$$\max_{1\leq k\leq n} \Big| \frac{1}{\sqrt{n}\sigma} \sum_{j=1}^{k} u_j - \frac{1}{\sqrt{n}} B_k \Big| = O_P(n^{1/p-1/2}), \tag{2.12}$$

where $\sigma^2 = Eu_0^2 + 2\sum_{k\geq 1} E(u_0 u_k)$.

For a proof of (2.12), we refer to **?** and **?**. We mention that many popular models in statistics and econometrics satisfy the two conditions (a) and (b) including the following examples:

- **TAR model**: $u_k = \phi_1 \max(u_{k-1}, 0) + \phi_2 \max(-u_{k-1}, 0) + \eta_k$, where $\max(|\phi_1|, |\phi_2|) < 1$;
- **Bilinear model**: $u_k = (\alpha_1 + \beta_1 \eta_{k-1}) u_{k-1} + \eta_k$, where α_1 and β_1 are real parameters satisfying $E|\alpha_1 + \beta_1 \eta_0|^q < 1$ for some $q > 0$;
- **GARCH model**:

$$u_k = \sigma_k \eta_k, \quad \text{with} \quad \sigma_k^2 = \gamma_0 + \sum_{i=1}^{m} \gamma_i u_{k-i}^2 + \sum_{j=1}^{l} \beta_j \sigma_{k-j}^2, \tag{2.13}$$

  where $\gamma_0 > 0$, $\gamma_i \geq 0$ for $1 \leq i \leq m$, $\beta_j \geq 0$ for $1 \leq j \leq l$ and $\sum_{i=1}^{m} \gamma_i + \sum_{j=1}^{l} \beta_j < 1$.

For more examples, we refer to **?**.

3 Self-weighted near unit root estimation

We consider model (1.1) with innovations $\{u_k\}_{k\geq 1}$ having the following form:

$$u_k = \sigma_k \, \epsilon_k, \tag{3.1}$$

where $\{\epsilon_k, \mathcal{F}_k\}_{k\geq 1}$ is a martingale difference with $E(\epsilon_k^2|\mathcal{F}_{k-1}) = 1$ ($\mathcal{F}_0 = \sigma(\emptyset, \Omega)$ is the trivial σ-field), and σ_k is a positive stationary α-mixing random variables sequence so that $\sigma_k \in \mathcal{F}_{k-1}$ and $0 < \sigma^2 = Eu_k^2 = E\sigma_1^2 < \infty$ for all $k \geq 1$. Such an innovation sequence allows for GARCH models of the form defined by (2.13) (e.g., **?**). Furthermore, σ_k allows for nonlinear processes such as TAR and Bilinear models or, more generally, stationary causal processes defined by $\sigma_k = l(\eta_k, \eta_{k-1}, \cdots)$, where $\{\eta_k\}_{k\in\mathcal{Z}}$ is a sequence of i.i.d. random variables so that $\sigma_k \in \mathcal{F}_{k-1}$ and $l(\cdot, \cdot, \cdots)$ is a nonnegative real measurable function of its components.

Define $\widehat{\alpha}_{1n}$ as in (1.3). The asymptotic theory for $\widehat{\alpha}_{1n}$ is given by the following result under mild additional moment conditions on σ_k and ϵ_k.

Theorem 3.1. *Suppose that for some $\delta > 0$,*

(a) $\sup_{k\geq 1} E(|\epsilon_k|^{2+\delta}|\mathcal{F}_{k-1}) \leq C < \infty$ *and* $E\sigma_1^{2+\delta} < \infty$*;*

(b) the mixing coefficients $\alpha(n)$ of $\{\sigma_k\}_{k\geq 1}$ satisfy $\alpha(n) \leq Cn^{-(1+\delta)}$.

Then, for any $b_n > 0$ satisfying $b_n \to 0$ and $b_n \log^K n \to \infty$ for some $K > 0$, we have

$$n\, b_n^{3/2} \left(\widehat{\alpha}_{1n} - \alpha\right) \quad \to_D \quad \sigma^{3/2} \left[\frac{2}{3}\, L_{J_\tau}(1, 0)\right]^{-1/2} N, \tag{3.2}$$

where N is a standard normal variate independent of the local time process $L_{J_\tau}(t, x)$ of the linear diffusion $J_\tau = \{J_\tau(t)\}_{t\geq 0}$ defined in (1.2), and

$$\Big[\sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \leq b_n \sqrt{n})\Big]^{1/2} \left(\widehat{\alpha}_{1n} - \alpha\right) \quad \to_D \quad \mathcal{N}(0, \sigma^2). \tag{3.3}$$

Remark 3.1. Since $\sigma^2 = Eu_k^2 = Eu_1^2$, under the given conditions a natural consistent estimator $\hat{\sigma}_{1n}^2$ of σ^2 is $\hat{\sigma}_{1n}^2 = \frac{1}{n}\sum_{k=1}^n \left(y_k - \widehat{\alpha}_{1n} y_{k-1}\right)^2$ or $\hat{\sigma}_{1n}^2 = \frac{1}{n}\sum_{k=1}^n \left(y_k - \widehat{\alpha}_n y_{k-1}\right)^2$ where $\widehat{\alpha}_n$ is the LS estimator of α. Consequently, result (3.3) can be re-written as

$$\hat{\sigma}_{1n}^{-1} \Big[\sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \leq b_n \sqrt{n})\Big]^{1/2} \left(\widehat{\alpha}_{1n} - \alpha\right) \quad \to_D \quad \mathcal{N}(0, 1), \tag{3.4}$$

with a pivotal standard normal limit which is therefore convenient for inference and an advantage in applications.

Remark 3.2. In comparison with the usual LS estimator $\widehat{\alpha}_n$, the self-weighted LS estimator $\widehat{\alpha}_{1n}$ has a slightly slower consistency rate. This reduction in the convergence rate is needed to achieve the standard normal limit theory. Indeed, since $y_{\lfloor nt\rfloor}/\sqrt{n}\sigma \Rightarrow J_\tau(t)$ on $D[0,1]$, the limit behaviour of $\widehat{\alpha}_{1n}$, when $b_n = b > 0$, is similar to that of $\widehat{\alpha}_n$. In principle, the rate $b_n \to 0$ can be chosen so that the self-weighted estimator $\widehat{\alpha}_{1n}$ has a higher consistency rate, although for finite n the quality of the approximation provided by the limit theory may suffer. The restriction $b_n \log^K n \to \infty$ for some $K > 0$ is imposed for technical reasons and in applications we may take $b_n = C_0 \log^{-1} n$ for simplicity and convenience, where C_0 is a positive constant. Both results (3.3) and (3.4) hold for all $C_0 \in \mathbb{R}^+$, but the choice of C_0 can have a significant impact in finite sample performance. Roughly speaking, in finite sample simulations, a balanced C_0 is required so that b_n is small (close to 0) while $b_n \log^K n$ is relatively large for some $K \geq 2$. This will be explored in simulations of Section 5.

Remark 3.3. Theorem 3.1 implies that, upon normalization, $\widehat{\alpha}_{1n}$ can serve as a test statistic for the unit root null hypothesis $\mathbb{H}_0: \ \alpha = 1$ against the alternative $\mathbb{H}_1 : \alpha < 1$. Further, model (1.1) with innovations (3.1) can be extended to ADF form as

$$y_k = \alpha y_{k-1} + \sum_{j=1}^{p-1} \theta_j (y_{k-j} - y_{k-j-1}) + u_k, \tag{3.5}$$

where $\alpha \in \mathbb{R}$, $\sum_{j=1}^{p-1} |\theta_j| < 1$ and $y_{-k} = 0$ for $k \geq 0$. Similarly to $\widehat{\alpha}_{1n}$, α and $\theta = (\theta_1, ..., \theta_{p-1})'$ can be estimated by using solving the weighted score equations

$$\sum_{k=1}^{n} \Big[y_k - \alpha y_{k-1} - \sum_{j=1}^{p-1} \theta_j \big(y_{k-j} - y_{k-j-1}\big)\Big] y_{k-1}\, I(|y_{k-1}| \leq b_n\sqrt{n}) = 0,$$
$$\sum_{k=1}^{n} \Big[y_k - \alpha y_{k-1} - \sum_{j=1}^{p-1} \theta_j \left(y_{k-j} - y_{k-j-1}\right)\Big] \left(y_{k-l} - y_{k-l-1}\right) = 0, \quad l = 1, ..., p-1.$$

Let $z_k = \big(y_{k-1} - y_{k-2}, ..., y_{k-p+1} - y_{k-p}\big)$ and $\widetilde{y}_k = y_k I(|y_k| \leq b_n\sqrt{n})$. Simple calculations show that the resulting estimator $(\widehat{\alpha}_{2n}, \widehat{\theta}_n)$ of (α, θ) is given by

$$\begin{pmatrix} \widehat{\alpha}_{2n} \\ \widehat{\theta}_n \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^n y_{k-1}\widetilde{y}_{k-1} & \sum_{k=1}^n z_k\, \widetilde{y}_{k-1} \\ \sum_{k=1}^n z_k'\, y_{k-1} & \sum_{k=1}^n z_k' z_k \end{pmatrix}^{-1} \begin{pmatrix} \sum_{k=1}^n y_k \widetilde{y}_{k-1} \\ \sum_{k=1}^n z_k' y_k \end{pmatrix}.$$

It is routine to show that $\widehat{\theta}_n - \theta = O_P(n^{-1/2})$ if $\alpha = 1$ and the following theorem holds for $\widehat{\alpha}_{2n}$, providing a new result in testing the unit root for the augmented Dickey-Fuller model (3.5).

Theorem 3.2. *Suppose that*

(a) model (3.5) holds with u_k satisfying (3.1);

(b) $\sup_{k\geq 1} E(|\epsilon_k|^{2+\delta}|\mathcal{F}_{k-1}) < \infty$ *and* $E\sigma_1^{2+\delta} < \infty$ *for some* $\delta > 0$*;*

(c) the mixing coefficients $\alpha(n)$ *of the sequence* $\{\sigma_k\}_{k\geq 1}$ *satisfy that* $\alpha(n) \leq Cn^{-(1+\delta)}$.

Under the null hypothesis $H_0 : \alpha = 1$*, for any* $b_n > 0$ *satisfying* $b_n \to 0$ *and* $b_n \log^K n \to \infty$ *for some* $K > 0$*, we have*

$$\Big[\sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \leq b_n \sqrt{n})\Big]^{1/2} (\widehat{\alpha}_{2n} - \alpha) \quad \to_D \quad \mathcal{N}(0, \sigma^2), \tag{3.6}$$

and

$$\hat{\sigma}_{2n}^{-1} \Big[\sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \leq b_n \sqrt{n})\Big]^{1/2} (\widehat{\alpha}_{2n} - \alpha) \quad \to_D \quad \mathcal{N}(0, 1), \tag{3.7}$$

where $\hat{\sigma}_{2n}^2 = \frac{1}{n}\sum_{k=1}^{n} \big(y_k - \widehat{\alpha}_{2n} y_{k-1} - z_k \widehat{\theta}_n\big)^2$.[2]

Remark 3.4. Since the original work of **?** which employed *iid* normal errors with no central limit theory, there has been extensive research on testing for a unit root in models with general error structures using functional central limit theory. The resulting limit theory of the tests is typically nonstandard and is expressed as functionals of Brownian motion, with nuisance parameters eliminated by both parametric and nonparametric methods. Reviews of earlier contributions appear in **?** and **?**. In recent years, the instrumental variables (IV) approach has been introduced in unit root tests. An advantage of IV unit root tests is that their asymptotics are standard normal, free of nuisance parameters – see **?**, **??**, **?**, **?**, **?** and the references therein. A feature for our approach is that it has desirable properties beyond standard normal asymptotics with a super consistency rate slightly below $O(n)$ which has some advantage in finite samples, as demonstrated in the partially aggregated differences estimation approach of **?**. Simulations show that when the truncation parameter C_0 in the truction rate $b_n = C_0 \log^{-1} n$ and lag parameter p are appropriately chosen, the empirical size can be controlled well around the nominal level across all the chosen sample sizes n. Moreover, the size performance of our new unit root test is also robust to the distribution of the innovation u_k. Furthermore, the power performance of our unit root test is comparable to that of the ADF test. In certain cases, our test has correct size control and outperforms the ADF test regarding power performance, showing its superiority in reducing the Type II error and corroborating its usefulness in specifying unit root behaviors.

[2] It is routine to see that $\hat{\sigma}_{2n}^2$ is a natural consistent estimator of $\sigma^2 = Eu_1^2$ under model (3.5). We may estimate σ^2 by using $\widetilde{\sigma}_{2n}^2 = \frac{1}{n}\sum_{k=1}^{n} \big(y_k - \widetilde{\alpha}_n y_{k-1} - z_k \widetilde{\theta}_n\big)^2$, where $(\widetilde{\alpha}, \widetilde{\theta}_n')$ is the LS estimator of (α, θ') in model (3.5).

4 Predictive regression

Self-weighted estimation can also be used in nonlinear cointegrating regression. In this regard, a general model was considered recently in **?** using strong smoothing conditions, which are difficult to verify for some dependent sequences, including martingale differences, as explained earlier in Section 2. The new local time limit theory of this paper avoids the use of strong smoothing conditions in the innovations and in doing so enables applications of self-weighted estimation to nonlinear nonparametric cointegrating regression. For brevity, that application is omitted here and the following analysis focuses on the predictive regression model

$$\begin{aligned} y_k &= \alpha\, f(x_{k-1}) + u_k, \\ x_k &= \beta x_{k-1} + \xi_k,\ \beta = 1 + \tau/n,\ k = 1, ..., n, \end{aligned} \tag{4.1}$$

where $x_0 = 0$, $f(x)$ is a given real function, $(\eta_k, u_k, \mathcal{F}_k)_{k\ge1}$ forms a martingale difference and $\xi_k = \sum_{j=0}^{\infty} \phi_j \eta_{k-j}$ with $\Phi := \sum_{j=0}^{\infty} \phi_j \neq 0$ and $\sum_{j=0}^{\infty} j|\phi_j| < \infty$. The self-weighted estimator $\widehat{\alpha}_{3n}$ of α is defined as

$$\begin{aligned} \widehat{\alpha}_{3n} &= \arg\min_{\alpha} \sum_{k=1}^{n} \big[y_k - \alpha\, f(x_{k-1})\big]^2 I(|x_{k-1}| \le b_n\sqrt{n}) \\ &= \frac{\sum_{k=1}^{n} y_k f(x_{k-1})\, I(|x_{k-1}| \le b_n\sqrt{n})}{\sum_{k=1}^{n} f^2(x_{k-1}) I(|x_{k-1}| \le b_n\sqrt{n})}, \end{aligned}$$

where $b_n \to 0$ is a seqeunce of positive constants. To explore the asymptotics of $\widehat{\alpha}_{3n}$ we use the following assumptions:

A1 $u_k = \sigma_{1k}\, \epsilon_{1k}$, $\eta_k = \sigma_{2k}\, \epsilon_{2k}$, and $\epsilon_k = (\epsilon_{1k}, \epsilon_{1k})'$, where

(a) $\{\epsilon_k, \mathcal{F}_k\}_{k\ge1}$ is a martingale difference sequence with natural filtration $\mathcal{F}_k$, $E(\epsilon_{1k}^2|\mathcal{F}_{k-1}) = E(\epsilon_{2k}^2|\mathcal{F}_{k-1}) = 1$ and, for some $\delta > 0$,

$$\sup_{k\ge1} \big[E(|\epsilon_{1k}|^{2+\delta}|\mathcal{F}_{k-1}) + E(|\epsilon_{2k}|^{2+\delta}|\mathcal{F}_{k-1})\big] < \infty;$$

(b) $\{\sigma_{1k}\}_{k\ge0}$ and $\{\sigma_{2k}\}_{k\ge0}$ are adapted to $\mathcal{F}_{k-1}$ and both are positive stationary α-mixing random processes with coefficients $\alpha(n) \le Cn^{-(1+\delta)}$ and $E\sigma_{11}^{2+\delta} + E\sigma_{21}^{2+\delta} < \infty$ for some $C > 0$ and $\delta > 0$.

A2 There exists a continuous real function $H(x)$ and a real function $\pi(\lambda) : (0, \infty) \to (0, \infty)$ such that

$$f(\lambda x) = \pi(\lambda)\, H(x) + R(\lambda, x)$$

where $|R(\lambda, x)| \le a(\lambda)\,(1 + |x|^{\delta})$ for some $\delta > 0$ and $a(\lambda)/\pi(\lambda) \to 0$, as $\lambda \to \infty$.

Condition **A1** allows for heteroskedasticity in both the regressor and the errors, which is useful in empirical work. Unlike **?**, no strong smoothness condition is required on the regressor x_k. **A2** indicates that $f(x)$ is a real homogeneous function, a form that is widely used in the literature. The following result provides the asymptotics of $\widehat{\alpha}_{3n}$.

Theorem 4.1. *Suppose **A1** and **A2** hold. Then, for any $b_n > 0$ satisfying $b_n \to 0$ and $b_n \log^K n \to \infty$ for some $K > 0$, we have*

$$\sqrt{n\, b_n}\pi(b_n\sqrt{n}\,)\left(\widehat{\alpha}_{3n} - \alpha\right) \quad \to_D \quad \sigma_1\,(\sigma_2\Phi)^{1/2}\Big(\int_{|x|\le 1} H^2(x)dx\Big)^{-1/2} L_{J_\tau}^{-1/2}(1,0)\,N, \tag{4.2}$$

where $\sigma_1^2 = E\sigma_{11}^2 = Eu_k^2$ and $\sigma_2^2 = E\sigma_{21}^2 = E\eta_k^2$ for $k \ge 1$, N is a standard normal variate independent of the local time $L_{J_\tau}(t,x)$ of the linear diffusion $J_\tau = \{J_\tau(t)\}_{t\ge 0}$, and

$$\Big[\sum_{k=1}^{n} f^2(x_{k-1})I(|x_{k-1}| \le b_n\sqrt{n})\Big]^{1/2}\left(\widehat{\alpha}_{3n} - \alpha\right) \quad \to_D \quad \mathcal{N}(0,\sigma_1^2). \tag{4.3}$$

We further have

$$\hat{\sigma}_{3n}^{-1}\Big[\sum_{k=1}^{n} f^2(x_{k-1})I(|x_{k-1}| \le b_n\sqrt{n})\Big]^{1/2}\left(\widehat{\alpha}_{3n} - \alpha\right) \quad \to_D \quad \mathcal{N}(0,1), \tag{4.4}$$

where $\hat{\sigma}_{3n}^2 = \frac{1}{n}\sum_{k=1}^n\big[y_k - \widetilde{\alpha}_n f(x_{k-1})\big]^2$ with $\widetilde{\alpha}_n = \frac{\sum_{k=1}^n y_k f(x_{k-1})}{\sum_{k=1}^n f^2(x_{k-1})}$. [3]

Remark 4.1. Theorem 4.1 provides pivotal limit theory free of β (in turn τ) and other unknown parameters associated with (ξ_k, u_k). For other IV estimation procedures used in predictive regression, we refer to **?**, **?**, **?** and **?**. In the latter paper, a chronologically trimmed LS method was developed which also has a standard normal asymptotics.

Theorem 4.1 requires a martingale structure in model (4.1). If endogenity is imposed in the model, as in other IVX estimation, Theorem 4.1 fails and a bias corrected estimator is required for a mixed normal limit. To illustrate, we conisder the following simple cointegrated predictive regression model:

$$\begin{aligned} y_k &= \alpha x_{k-1} + v_k, \\ x_k &= \beta x_{k-1} + \xi_k, \quad \beta = 1 + \tau/n, \\ \begin{pmatrix} v_k \\ \xi_k \end{pmatrix} &= \sum_{j=0}^{\infty} D_j\,\zeta_{k-j}, \end{aligned} \tag{4.5}$$

[3]We may take $\hat{\sigma}_{3n}^2 = \frac{1}{n}\sum_{k=1}^n\big[y_k - \widehat{\alpha}_{3n} f(x_{k-1})\big]^2$. Since $\widetilde{\alpha}_n$ is the LS estimator of α in model (4.1) having a faster consistency rate in comparison with $\widehat{\alpha}_{3n}$, result (4.4) usually has a better finite sample performance.

where $x_0 = 0$, $\zeta_j \sim_{iid} (0, \sigma^2)$ and the coefficients $D_j = \begin{pmatrix} \psi_j \\ \phi_j \end{pmatrix}$ satisfy that $\Psi := \sum_{j=0}^{\infty} \psi_j \neq 0$, $\Phi := \sum_{j=0}^{\infty} \phi_j \neq 0$ and $\sum_{j=0}^{\infty} j(|\psi_j|+|\phi_j|) < \infty$. For this model, the bias corrected-IVX estimator is $\hat{\alpha}_{bIVX} = \left(\sum_{k=1}^{n} y_k \tilde{z} - n\hat{\Lambda}_{\xi v}\right) / \sum_{k=1}^{n} x_{k-1}\tilde{z}$ where $\tilde{z}$ is the usual IVX instrument and

$$\hat{\Lambda}_{\xi v} = \sum_{j=1}^{M} k\big(j/M\big) \frac{1}{n} \sum_{1 \le k, k+j \le n} \hat{\xi}_k \hat{v}_{k+j}$$

is the usual lag kernel estimate of the one sided long run covariance $\Lambda_{\xi v} = \mathbb{E}(\xi_1 v_1)$ based on the residuals $\hat{v}_{k+j} = y_{k+j} - \hat{\alpha} x_{k+j-1}$ and $\hat{\xi}_k = x_k - \hat{\beta} x_{k-1}$ with OLS estimates $\hat{\alpha}$ and $\hat{\beta}$. For such OLS estimators, it is well-known that

$$|\hat{\alpha} - \alpha| + |\hat{\beta} - \beta| \quad = \quad O_P(n^{-1}). \tag{4.6}$$

In present paper, we consider the bias corrected SW estimator

$$\hat{\alpha}_{bSW} = \left(\sum_{k=1}^{n} x_{k-1}^2 I(|x_{k-1}| \le b_n \sqrt{n}) \right)^{-1} \left(\sum_{k=1}^{n} y_k x_{k-1} I(|x_{k-1}| \le b_n \sqrt{n}) - \widetilde{\Lambda}_{\xi v} \right) \tag{4.7}$$

where

$$\widetilde{\Lambda}_{\xi v} \quad = \quad \sum_{j=1}^{M} \sum_{k=1}^{n-j} \hat{v}_{k+j} \big[x_k I(|x_k| \le b_n \sqrt{n}) - x_{k-1} I(|x_{k-1}| \le b_n \sqrt{n}) \big].$$

The following theorem shows that the scaled $\hat{\alpha}_{bSW}$ has a standard normal limit when

$$L_M := M + n \,\big(\sum_{j=M}^{\infty} \psi_j^2 \big)^{1/2} = o\big[(n b_n)^{1/2} \big]. \tag{4.8}$$

If $\psi_j = 0$ for $j \ge M_0 > 0$, it is routine to see that (4.8) holds with $M = M_0$ as $n b_n \to \infty$.

Theorem 4.2. *Suppose $E|\zeta_1|^3 < \infty$ and (4.8) holds. For any $b_n > 0$ satisfying $b_n \to 0$ and $b_n \log^K n \to \infty$ for some $K > 0$, we have*

$$n b_n^{3/2} \big(\hat{\alpha}_{bSW} - \alpha \big) \quad \to_D \quad (3\,\sigma\,\Phi/2)^{1/2}\, \Psi_1\, L_{J_\tau}^{-1/2}(1,0)\, N, \tag{4.9}$$

where $\Psi_1^2 = E v_1^2 = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$, N is a standard normal variate independent of the local time $L_{J_\tau}(t,x)$ of the linear diffusion $J_\tau = \{J_\tau(t)\}_{t \ge 0}$, and

$$\Big[\sum_{k=1}^{n} x_{k-1}^2 \, I(|x_{k-1}| \le b_n \sqrt{n}) \Big]^{1/2} \big(\hat{\alpha}_{bSW} - \alpha \big) \quad \to_D \quad \mathcal{N}(0, \Psi_1^2). \tag{4.10}$$

We further have

$$\widehat{\Psi}_{1n}^{-1} \, \Big[\sum_{k=1}^{n} x_{k-1}^2 \, I(|x_{k-1}| \le b_n \sqrt{n}) \Big]^{1/2} \big(\hat{\alpha}_{bSW} - \alpha \big) \quad \to_D \quad \mathcal{N}(0,1), \tag{4.11}$$

where $\widehat{\Psi}_{1n}^2$ is a consistency estimator of Ψ_1^2. [4]

[4] In simulations, it is found that the two-sided long run variance (LRV) estimator $\widehat{\Psi}_{1n}^2$ of $\Psi_1^2 = E v_1^2$ has a better finite sample performance. In this case, $\widehat{\Psi}_{1n}^2 = \frac{1}{n} \sum_{k=1}^{n} \widetilde{v}_k^2 + \frac{2}{n} \sum_{j=1}^{M} K(\frac{j}{M}) \sum_{k=1}^{n-j} \widetilde{v}_k \widetilde{v}_{k+j}$, where $M \to \infty$, $K(x) = 1 - |x|$ for $|x| \le 1$ and $\widetilde{v}_k = y_k - \widetilde{\alpha} x_{k-1}$ with $\widetilde{\alpha}$ being the LS estimator.

Remark 4.2. The idea of self-weighted estaimation in present paper is simple and the methodology can be easily extended to general nonlinear regression function as shown in Theorem 4.1. See, also, Jin and Wang (2022). However, due to the restrication in multiple local time theory, there is a technical difficulty for this new method to be used in multi-regression with nonstationary times, as much of the research dealing with the IVX approach (e.g., **?**, **?** and **?**). Allowing for multiple regressors x_t is an important advantage of IVX compared with the self-weighted methods of predictive regression preesented in this paper. An additional line of research commenced by **?** deals with semiparametric regression in which the function $f(x_{t-1}) = g(\frac{t}{n})' x_{t-1}$ in the predictive regression equation of (4.1) has time-varying coefficients $g(\frac{t}{n})$ which are estimated by a sieve version of IVX regression, again allowing for linear process errors η_t and multiple regressors x_t.

5 Simulations

Simulations were conducted to explore the finite sample properties of the self-weighted estimator, IVX estimators and first difference estimators in relation to usual LS estimation in near unit root regression and in unit root testing. Comparisons with the IVX procedure in predictive regression are also reported.

PCB: I have done word by word revisions of the text up to here.

We may take $b_n = C_0 \log^{-1} n$, where $C_0 = 1/10, 1, 5, 10$, etc. For different C_0 values, we may further consider the impact of C_0 in finite sample simulations. It would be interested in to consider the optimal C_0, but this seems to be difficult.

This section examines the finite-sample performance of the proposed self-weighted least squares (LS) estimator and its corresponding inference procedure, which attains pivotal null limit theory for testing the unit root and predictability of persistent regressors.

5.1 Simulation for local-to-unity autoregression

We conduct numerical simulations for the general null hypothesis in the near unit root autoregression model:

$$y_k = \alpha y_{k-1} + u_k$$

with $\alpha = 1 + \tau/n$ and $k = 1, ..., n$. Values of localizing parameter τ are considered as $\tau \in \{0, -1, -5, -10\}$. The sample size $n \in \{50, 100, 200, 500\}$ and initial value $y_0 = 0$. For each $k = 1, ..., n$, the innovation u_k admits a GARCH (1,1) representation:

$$u_k = \sigma_k \epsilon_k, \quad \sigma_k^2 = \varphi_0 + \varphi_1 u_{0,k-1}^2 + \varphi_2 \sigma_{k-1}, \tag{5.1}$$

where $\epsilon_k \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $\varphi_0 = 1, \varphi_1 = 0.2$ and $\varphi_2 = 0.3$. The number of repetitions is $10,000$ for each data-generating process (DGP).

The null hypothesis is $\mathcal{H}_0$: $\alpha = \alpha_0$ where α_0 is the true value of parameter. Equivalently, we are testing $\mathcal{H}_0 : \tau = \tau_0$ where τ_0 is the true localizing parameter used in the DGPs. To calculate the self-weighted estimator, we take the truncation rate $b_n = C_0 \log^{-1}(n)$ in which $C_0 \in \{1, 2, 3, 4, 5, 10, 20\}$. The nominal size is set as 5%. The size performance of the self-normalized statistic defined in (3.4) of the main paper is provided in Table 1.

[Insert Table 1 here]

The simulation results of Table 1 show that the test that relies on the self-weighted estimator can control size performance well around the nominal level free of the localizing parameter τ. It is well evident that when the truncation parameter C_0 is as small as 3, oversized phenomena can seldom be observed around all the chosen sample sizes n and localizing parameter τ. Even when C_0 is chosen as 20, the size performance can be substantially improved when the sample size is as large as 200. These observations corroborate the usefulness of our testing procedure in obtaining the unified Gaussian null limit theory by trimming extremely large observations. With the numerical evidence provided in Table 1, we suggest the choice of C_0 as any value between 1 and 3 in testing the autoregressive model with local-to-unity regressors.

5.2 Simulation for new unit root test

We examine the empirical size and power performance of the new unit root test statistic $\widehat{\alpha}_{2n}$ defined in the main paper. As in Section 3, we consider the unit root model:

$$y_k = \alpha y_{k-1} + \theta(y_{k-1} - y_{k-2}) + u_k \tag{5.2}$$

where $\alpha = 1$, $\theta \in \{0, -0.5\}$, and u_k is an GARCH (1,1) process given in (5.1) with ϵ_k being specified in the following two cases:

Case I: $\epsilon_k \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$; **Case II:** $\epsilon_k \overset{i.i.d.}{\sim} U\left(-\sqrt{3}, \sqrt{3}\right)$.

We set the sample size $n \in \{50, 100, 200, 500\}$ and the number of repetitions $10,000$. The considered null hypothesis is $\mathcal{H}_0$: $\alpha = 1$ against the alternative $\mathcal{H}_1$: $\alpha < 1$. To compute our new unit root test statistic for the augmented autoregression model, we take the truncation rate $b_n = C_0 \log^{-1} n$ in which $C_0 \in \{1, 2, 3, 4, 5, 10\}$ and the number of lag terms in augmented autoregression $p \in \{0, 1, 2\}$. We evaluate the size performance of our new unit root testing statistic $\widehat{\alpha}_{2n}$ and the classical ADF unit root test with the nominal size set as 5% in the simulation study. The present paper also examines the power performance of the unit root test under the following sequence of local alternative hypotheses $\mathcal{H}_{1n}$: $\alpha = 1 - \delta/n$ where $\delta \in [0, 30]$. The empirical size and power performance of the self-normalized statistic defined in (3.7) of the main paper and the classical ADF test are provided in Tables 2–3 and Figure 1.

[Insert Tables 2–3 and Figure 1 here]

Table 2 and 3 examine the cases that y_k follows a AR(1) process by letting $\theta = 0$ and a AR(2) process by letting $\theta = -0.5$ in equation (5.2). Both tables discuss the cases in which the underlying error process ϵ_k follows a Gaussian distribution $\mathcal{N}(0, 1)$ and uniform distribution $U(-\sqrt{3}, \sqrt{3})$. The results show that when the truncation parameter C_0 and lag parameter p are appropriately chosen, the empirical size can be controlled well around the nominal level across all the chosen sample sizes n. Moreover, the size performance of our new unit root test is also robust to the distribution of the innovation ϵ_k considered.

Figure 1 further plots the local power of the proposed unit root test and ADF test with sample size $n = 200$. We set the lag parameter $p = 0$ for the cases of the y_k is an AR(1) process and $p = 1$ for the cases of the y_k is an AR(2) process. It is well evident that $\delta = 0$ implies $\alpha = 1$, corresponding to the model under the null hypothesis, while $\delta \neq 0$ represents the model under various local alternatives. Generally speaking, the power performance of our unit root test is comparable to that of the ADF test. In several cases, especially when $y_k \sim$AR(2), our test with $C_0 = 5$ has correct size control and outperforms the ADF test regarding power performance, showing its superiority in reducing the Type II error and corroborating its usefulness in specifying unit root behaviors.

5.3 Simulations for predictive regression

We consider the univariate regression model:

$$
\begin{aligned}
y_k &= \alpha x_{k-1} + v_k, \\
x_k &= \beta x_{k-1} + \xi_k, \quad \beta = 1 + \tau/n, \\
v_k &= \phi_v v_{k-1} + \zeta_k, \ \xi_k = \phi_\xi \xi_{k-1} + \zeta_k,
\end{aligned} \tag{5.3}
$$

where both y_k and x_k are scalars, the initial value $x_0 = 0$ and the noise component $\zeta_k \sim$ *i.i.d.* $\mathcal{N}(0, 1)$. The numerical simulations are constructed by using $10,000$ repetitions for parameter settings $\tau \in \{0, -1, -5, -10\}$, $\phi_v \in \{0, 0.5\}$, $\phi_\xi \in \{0, 0.5\}$, and sample size $n \in \{50, 100, 200, 500\}$.

To test null hypothesis $\mathcal{H}_0$: $\alpha = \alpha_0$ against the alternative $\mathcal{H}_1$: $\alpha \neq \alpha_0$, we utilize the self-weighted (SW) estimator $\widehat{\alpha}_{3n}$ with $f(x) = x$ in Theorem 4.1 in the case of predictive regression (i.e., $\phi_v = 0$), and the bias-corrected self-weighted (BC-SW) estimator $\widehat{\alpha}_{bSW}$ in Theorem 4.2 with serially correlated error v_k (i.e., $\phi_v \neq 0$). In both statistics, we take the truncation rate $b_n = C_0 \log^{-1} n$ in which $C_0 \in \{1, 2, 3, 4, 5, 10\}$ and the Bartlett kernel with the bandwidth $M := \lfloor n^{0.33} \rfloor$ is used in constructing the LRV estimator $\hat{\Psi}_{1n}$ given in (4.11).

For comparisons, we also apply the test statistic that depends on the IVX-based test for the slope coefficient α. In particular, the IVX instrument is given by

$$\widetilde{z}_k = R_{nz}\widetilde{z}_{k-1} + \Delta x_k \text{ with } R_{nz} = I_n + C_z/n^{\gamma}, \tag{5.4}$$

in which we set the distancing parameter $C_z = -1$ and rate parameter $\gamma \in \{0.6, 0.7, 0.8, 0.9\}$. Under the null hypothesis $\mathcal{H}_0$: $\alpha = \alpha_0$, we consider two IVX-based test statistics, namely the IVX and BC-IVX tests (**??**) as

$$t_{IVX} = \frac{\widehat{\alpha}_{IVX} - \alpha_0}{\sqrt{\mathrm{Var}(\widehat{\alpha}_{IVX})}}, \quad \text{(IVX test)} \tag{5.5}$$

$$t_{bIVX} = \frac{\widetilde{\alpha}_{IVX} - \alpha_0}{\sqrt{\mathrm{Var}(\widetilde{\alpha}_{IVX})}}, \quad \text{(BC-IVX test)} \tag{5.6}$$

in which the IVX estimator $\widehat{\alpha}_{IVX} = (\sum_{k=1}^{n} \widetilde{z}_{k-1}x_{k-1})^{-1}(\sum_{k=1}^{n} \widetilde{z}_{k-1}y_k)$ and the bias-corrected IVX estimator $\widetilde{\alpha}_{IVX} = (\sum_{k=1}^{n} \widetilde{z}_{k-1}x_{k-1})^{-1}(\sum_{k=1}^{n} \widetilde{z}_{k-1}y_k - n\Lambda_{v\xi})$. The bias correction term $\hat{\Lambda}_{v\xi} = \sum_{j=1}^{M} k\left(\frac{j}{M}\right)\frac{1}{n}\sum_{1\le k, k+j\le n} \widetilde{\xi}_k \widetilde{v}_{k+j}$ is based on the OLS residuals $\widetilde{v}_k$ and $\widetilde{\xi}_k$, the Bartlett kernel function, and the choice of bandwidth $M = \lfloor n^{0.33} \rfloor$.

[Insert Tables 4–6 here]

We report the size performance of the SW, BC-SW, IVX, and BC-IVX tests in Tables 4–6. To be specific, Tables 4–5 show the empirical size performance of the SW and IVX tests when the prediction error component v_k has no serial correlations. As the errors of predictive regression models accommodate no serial correlations, no bias correction terms are requested. By mechanisms of SW and IVX tests, both tests reduce the degrees of persistence in regressors, generating limiting Gaussian distributions that facilitate pivotal test procedures. It is well evident that in Tables 4–5, both tests control size performances reasonably well when the tuning parameters C_0 and γ are appropriately chosen. Also, similar patterns can be observed in both SW and IVX tests: when the tuning parameters C_0 and γ get larger, the degrees of persistence in regressors can be less reduced. Accordingly, we find that the oversizing phenomenon becomes more severe in such cases. Therefore, these simulated results show that a suitable choice of C_0 is 3 to 5, with which value the SW estimator performs reasonably well in finite samples.

Table 6 examines the empirical size performance of the SW, BC-SW, IVX, and BC-IVX tests when the prediction error v_k allows serial correlations. As v_k follows an AR(1) process, an additional bias-correction term is requested; otherwise, terribly behaved size performances can be observed. Table 6 clearly shows that the self-weighted test without a bias-correction term can be highly conservative, while the IVX test can also be severely oversized without bias-correction. This fact calls upon bias-corrected inference procedures, including those relying on the BC-SW and BC-IVX test statistics. Generally, both bias-corrected inference procedures can control the finite-sample size performance reasonably well around the nominal level when the prediction error v_k is serially correlated. Their performances are comparable, though moderate under-sizing phenomena can be observed for the self-weighted types of test statistics. Similar to serially uncorrelated cases of Tables 4–5, tuning parameters C_0 and γ determine the magnitude of persistence reduction in regressors. Typically, with a larger value of C_0 and γ, the regressors become more persistent and present challenges to control sizes. All these simulated results show that a suitable choice of C_0 is 3 to 5, with which value the BC-SW performs reasonably well in finite samples.

[Insert Figures 2–4 here]

This paper also examines the local power performance of SW, BC-SW tests under the following sequence of local alternative hypotheses $\mathcal{H}_{2n}$: $\alpha = \alpha_0 + \delta/n$ where $\delta \in [0, 50]$. Figures 2–4 present the local power performance of our SW related test statistics. For comparison, we also plot the local power function for the IVX and BC-IVX tests. In particular, Figure 2 and 3 are concerned with the self-weighted estimator with the iid error term, while Figures 4 focuses on the BC-SW estimator with a serially correlated error term. It is evident that when the tuning parameter C_0 is set as 3 or 5, the SW, BC-SW test statistics are consistent with power functions approaching unity. Another pattern that can be observed is that as C_0 increases, the local power performances for all the self-weighted tests are improved. When C_0 is selected as 5, the self-weighted test has power performances comparable to those of the IVX-based test.

6 Empirical Illustration

This section applies the self-weighted test statistic to examine the predictability of market fundamentals on the S&P 500 excess returns. Specifically, we apply the widely used financial dataset of **?**, ranging from January 1952 to December 2022. In the discussion of predictability testing, we mainly consider monthly and quarterly data.

Based on **?**, the present paper computes the excess return as

$$\text{Excess Return}_t = \log(1 + P_t) - \log(1 + \text{Rfree}_{t-1}),$$

where P_t is the S&P 500 value-weighted return and Rfree_{t-1} denotes the risk-free interest rate. Based on **?**, we employ the following persistent predictors: book-to-market value ratio (**b/m**), dividend payout ratio (**d/e**), default yield spread (**dfy**), dividend-price ratio (**d/p**), dividend yield (**d/y**), earnings-price ratio (**e/p**), inflation rate (**infl**), long-term yield (**lty**), net equity expansion (**nits**), T-bill rate (**tbl**) and term spread (**tms**). **?** also showed that nearly all these regressors have roots that are close to unity, indicating the necessity of introducing testing procedures that stay robust to the nonstandard null distribution generated by persistent regressors. In terms of this concern, they proposed the IVX instrumentation to reduce the degree of persistence in regressors and further obtained empirical findings in which the chance of spurious statistical significance is greatly reduced.

Our self-weighted test statistic can also obtain a robust null limit theory in the presence of unit root regressors. Though our mechanism attains the Gaussian asymptotic theory by trimming extreme values, which is different from the idea of the IVX method, the Gaussian critical values can still be applied to our analysis. In the following discussions, we intend to apply our self-weighted predictability test to the data of **?** and investigate the predictability of the aforementioned economic fundamentals in both monthly and quarterly frequencies.

For the self-weighted estimators, we choose the indicator weight function with the truncation rate $b_n = C_0 \log^{-1}(n)$ for the standardized variables. We set $C_0 = 1/2$ in our empirical study. For the IVX estimator, we set the parameter $\gamma = 0.9$. The empirical results are presented in Tables 7 and 8.

[Insert Tables 7–8 here]

Table 7 examines the predictability of regressors using monthly data. In this case, we rely on the univariate predictive regression model along with testing procedures that use OLS, SW, and IVX estimators. It is evident that both SW and IVX tests can help reduce the fake significance of regressors. For instance, the least squares test statistic detects the predictive power of the long-term yield (**lty**) on the 10% significance level, while both our SW test and IVX statistic cannot reject the null hypothesis of no predictability. Similar observations can also be found for the predictor, T-bill rate (**tbl**), whose predictability can be detected by both the OLS and IVX tests on the 5% percent level but not by the SW test statistic. All these findings corroborate the validity of applying the robust inference procedure to eliminate the fake predictability phenomenon. Moreover, the results of the T-bill rate (**tbl**) show that our newly proposed testing method performs better than the IVX test in controlling over-rejection. Table 7 shows that our SW testing procedure is able to reduce over-rejections in finding predictive phenomena in the empirical analysis of monthly data.

Table 8 discusses the case of quarterly data and presents the results of all three tests. Similar to the main results of monthly data, the SW test statistic manages to reduce the statistical significance of inflation rate (**infl**) and T-bill rate (**tbl**). In particular, both the OLS and IVX tests find inflation rate (**infl**) and T-bill rate (**tbl**) regressors can forecast the excess return on the 10% level while our SW test fails to the evidence of predictability. Basically, the proposed test statistic of the present paper detects no predictive ability in terms of quarterly data. Similar observations of diminishing predictability are also provided in **?**. One possible economic explanation for the diminishing predictability in the quarterly data is the loss of information in the low-frequency sampling, which further generates the low noise-signal ratio for long-term forecasting of stock returns. Our discussion of quarterly data also differs significantly from the case of long-run predictability in the literature, in which the sampling data of higher frequency is used for long-term prediction testing (**?**). This difference further explains why the long-run forecasting relationship is always statistically significant while our low-frequency testing is insignificant.

7 Proofs

7.1 Proof of Proposition 2.1

As in **?**, page 269, on a richer probability space, there exists a Brownian motion B and non-negative random variables $\tau_1, \tau_2, ...$ such that $\{S_k, k \geq 1\} =_d \{S_k^* = B_{T_k}, k \geq 1\}$, and for $k \geq 1$, $E(\tau_k | \mathcal{F}_{k-1}^*) = E(u_k^2 | \mathcal{F}_{k-1})$ and

$$E(\tau_k^{p/2} | \mathcal{F}_{k-1}^*) \leq C_p E(|u_k|^p | \mathcal{F}_{k-1}),$$

where $\mathcal{F}_k^*$ is generated by $S_1^*, ..., S_k^*$, $T_k = \sum_{j=1}^k \tau_j$ is $\mathcal{F}_k^*$-measurable and C_p is a constant depending only on p. Result (2.9) will follow if we prove:

$$\Delta_n \;:=\; \sup_{0 \leq t \leq 1} \big| B_{T_{\lfloor nt \rfloor}} - B_{d_n^2 t} \big| = O_P\big[d_n(\delta_n^{1/2} + \mathcal{L}_{np}^{1/p}) \log^{1/2} n\big]. \tag{7.1}$$

In fact, $\{B_t^n := B_{d_n^2 t}/d_n\}_{t \geq 0}$ is a Brownian motion and there exists a random array $\{S_{nk}^*, k = 1, 2, \cdots, n, n \geq 1\}$ such that (see e.g. Theorem 6.10 in **?**)

$$(\{B_t\}_{t \geq 0}, S_{n1}^*, \cdots, S_{nn}^*) =_d (\{B_t^n\}_{t \geq 0}, S_1^*, \cdots, S_n^*).$$

Hence

$$\sup_{0\le t\le 1}\Big|\frac{1}{d_n}\,S^*_{n,\lfloor nt\rfloor}-B_t\Big| =_d \frac{\Delta_n}{d_n}=O_P\big[(\delta_n^{1/2}+\mathcal{L}_{np}^{1/p})\log^{1/2}n\big].$$

This proves (2.9).

We now prove (7.1). By using Kolmogorov's inequality for martingales, we have

$$\begin{aligned}
P\big(\max_{1\le k\le n}|T_k-V_k^2|\ge\Delta\big) &= P\Big(\max_{1\le k\le n}\Big|\sum_{j=1}^k\big[\tau_j-E(\tau_j|\mathcal{F}^*_{j-1})\big]\Big|\ge\Delta\Big)\\
&\le \Delta^{-p/2}\,E\Big|\sum_{j=1}^n\big[\tau_j-E(\tau_j|\mathcal{F}^*_{j-1})\big]\Big|^{p/2}\\
&\le C_p\,\Delta^{-p/2}\sum_{j=1}^n E|u_j|^p
\end{aligned}$$

for any $\Delta>0$ and $2<p\le 4$, where C_p depends only on p. This, together with (2.8), yields that, for any $\eta>0$, there exists an $A>0$ so that, whenever $\Delta\ge 4A\,(\delta_n+\mathcal{L}_{np}^{2/p})$ and $n\ge A$,

$$\begin{aligned}
&P\big(\sup_{0\le t\le 1}|T_{\lfloor nt\rfloor}-d_n^2t|\ge d_n^2\Delta\big)\\
\le\;& P\big(\max_{1\le k\le n}|T_k-V_k^2|\ge d_n^2\Delta/2\big)+P\big(\max_{1\le k\le n}|V_k^2/d_n^2-k/n|\ge\Delta/2-1/n\big)\\
\le\;& C_p\,(\Delta/2)^{-p/2}\,\mathcal{L}_{np}+P\big(\max_{1\le k\le n}|V_k^2/d_n^2-k/n|\ge A\delta_n\big)\\
\le\;& (2A)^{-p/2}\,C_p+\eta\le 2\eta,
\end{aligned}\tag{7.2}$$

where we have used the fact that $\mathcal{L}_{np}^{2/p}\ge n^{-1+2/p}$ since by Hölder's inequality,

$$d_n^2=\sum_{j=1}^n Eu_j^2\le n^{1-2/p}\Big(\sum_{j=1}^n E|u_j|^p\Big)^{2/p}=n^{1-2/p}d_n^2\mathcal{L}_{np}^{2/p}.$$

On the other hand, it is well-known (e.g., **?**, (4.32)) that, for any $\epsilon>0$ and $\Delta>0$,

$$\begin{aligned}
P\Big(\sup_{\substack{|s-t|\le d_n^2\Delta\\ 0\le t\le d_n^2}}|B_t-B_s|\ge d_n\epsilon\Big) &= P\Big(\sup_{\substack{|s-t|\le\Delta\\ 0\le t\le 1}}|B_t-B_s|\ge\epsilon\Big)\\
&\le 14\,\epsilon^{-1}\Delta^{-1/2}\exp(-\epsilon^2/(18\Delta)).
\end{aligned}\tag{7.3}$$

By using (7.2) and (7.3) and taking $\Delta=4A\,(\delta_n+\mathcal{L}_{np}^{2/p})$ and $\epsilon=6(\Delta\log n)^{1/2}$, simple calculations yield, for any $\eta>0$, there exists an $A>0$ so that, whenever $n\ge A$,

$$\begin{aligned}
&P\big(\Delta_n\ge 12A^{1/2}\,d_n\,(\delta_n^{1/2}+\mathcal{L}_{np}^{1/p})\log^{1/2}n\big)\le P\big(\Delta_n\ge d_n\epsilon\big)\\
\le\;& P\Big(\sup_{\substack{|s-t|\le d_n^2\Delta\\ 0\le t\le d_n^2}}|B_t-B_s|\ge d_n\epsilon\Big)+P\Big(\sup_{0\le t\le 1}|T_{\lfloor nt\rfloor}-d_n^2t|\ge d_n^2\Delta\Big)\\
\le\;& C\,n^{-2}\,\Delta^{-1}\log^{-1/2}n+2\eta\le 3\eta,
\end{aligned}$$

yielding (7.1).

7.2 Proof of Theorem 2.1

Without loss of generality, assume that $0 < b < 1$ and $X_{nk} = X^*_{nk}$ for each $n \ge 1$ and $1 \le k \le n$. In step 1, we show (2.5) when $g(s,x)$ is continuous on $[0,1] \times \mathbb{R}$ satisfying $|g(s,x)| \le C_0/(1+|x|^{1+b})$. We start with an additional condition:

Con: there exists an $a > 0$ such that $\widehat{g}(s,x) = 0$ for all $0 \le s \le 1$ and $|x| \ge a$, where $\widehat{g}(s,x) = \int_{-\infty}^{\infty} e^{ixt} g(s,t)dt$.

This additional **Con** will be removed later. Let $\widetilde{g}(t) = \sup_{0\le s\le 1} |g(s,t)|$. Since the **Con** implies that $g(s,x) = \frac{1}{2\pi}\int_{-a}^{a} e^{itx}\widehat{g}(s,-t)dt$ and $|\widehat{g}(s,x)| \le \int_{-\infty}^{\infty} \widetilde{g}(t)dt < \infty$ uniformly on $[0,1]\times\mathbb{R}$, it is readily seen from (2.4) that, for any $A > 0$,

$$\begin{aligned}
\frac{c_n}{n}\sum_{k=1}^{n} g\Big(\frac{k}{n}, c_n X_{nk}\Big) &= \frac{1}{2\pi n}\sum_{k=1}^{n}\int_{-ac_n}^{ac_n}\widehat{g}\Big(\frac{k}{n}, -\frac{t}{c_n}\Big)\, e^{it X_{nk}} dt \\
&= \frac{1}{2\pi}\int_{-ac_n}^{ac_n}\int_0^1 \widehat{g}\Big(\frac{[nu]}{n}, -\frac{t}{c_n}\Big)\, e^{it X_{n,[nu]}} du\, dt + o_{\mathrm{P}}(1) \\
&= \frac{1}{2\pi}\int_{-ac_n}^{ac_n}\int_0^1 \widehat{g}\Big(\frac{[nu]}{n}, -\frac{t}{c_n}\Big)\, e^{it X_u} du\, dt + o_{\mathrm{P}}(1) \\
&= R_{1n}(A) + R_{1n}(A), \qquad (7.4)
\end{aligned}$$

where

$$\begin{aligned}
R_{1n}(A) &= \frac{1}{2\pi}\int_{|t|\le A}\int_0^1 \widehat{g}\Big(\frac{[nu]}{n}, -\frac{t}{c_n}\Big)\, e^{it X_u} du\, dt, \\
R_{2n}(A) &= \frac{1}{2\pi}\int_{A<|t|\le ac_n}\int_0^1 \widehat{g}\Big(\frac{[nu]}{n}, -\frac{t}{c_n}\Big)\, e^{it X_u} du\, dt.
\end{aligned}$$

Note that, for any $A > 0$, as $n \to \infty$,

$$\sup_{0\le u\le 1}\sup_{|t|\le A} |\widehat{g}([nu]/n, -t/c_n) - \widehat{g}(u,0)| \to 0.$$

It follows that, for any $A > 0$, as $n \to \infty$,

$$\Big(X_{n,\lfloor nt\rfloor},\ R_{1n}(A)\Big) \Rightarrow \Big(X_t,\ \frac{1}{2\pi}\int_{|t|\le A}\int_0^1 \widehat{g}(u,0)e^{isX_u} du\, ds\Big), \qquad (7.5)$$

on $D_{\mathbb{R}^2}[0,1]$. As $\widehat{g}(u,0) = G(u)$ and by recalling (2.2), the result (2.5) under the **Con** will follow if we prove

$$\mathrm{E}|R_{2n}(A)|^2 \ \to\ 0, \qquad (7.6)$$

as $n \to \infty$ first and then $A \to \infty$. In fact, for a Gaussian process X satisfying (2.1), we have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^T\int_0^T \big|\mathrm{E}e^{itX_u - isX_v)}\big| du\, dv\, dt\, ds < \infty, \quad \text{for any } T > 0.$$

Thus, by recalling $|\widehat{g}(s,x)| \leq \int_{-\infty}^{\infty} \widetilde{g}(t)dt < \infty$ uniformly on $[0,1]\times\mathbb{R}$, it follows that

$$\begin{aligned} \mathrm{E}|R_{2n}(A)|^2 &\leq C\int_{|t|>A}\int_{|t|>A}\int_0^1\int_0^1 \widehat{g}\Big(\frac{[nu]}{n},\, -\frac{t}{c_n}\Big)\,\widehat{g}\Big(\frac{[nv]}{n},\, \frac{s}{c_n}\Big)\mathrm{E}e^{it\,X_u - isX_v}du\,dv\,dt\,ds \\ &\leq C\int_{|t|>A}\int_{|t|>A}\int_0^1\int_0^1 \big|\,\mathrm{E}e^{it\,X_u - isX_v}\big|\,du\,dv\,dt\,ds \\ &\to 0, \end{aligned}$$

as $n\to\infty$ first and then $A\to\infty$. This proves (7.6) and also completes the proof of (2.5) when $g(s,x)$ is continuous by imposing the additional **Con**.

We next remove the additional **Con** when $g(s,x)$ is continuous on $[0,1]\times\mathbb{R}$ satisfying $|g(s,x)| \leq C/(1+|x|^{1+b})$. This is essentially the same as in the proof of Theorem IV 2.1 in **?** (see also **?**), and so we only provide an outline. Write, for $0<b<1$,

$$f(x) = \sum_{n=1}^{\infty} n^{-1-b/2}\frac{\sin^2(x-n)}{(x-n)^2}$$

and, for $\delta>0$ ($\sin y/y \equiv 1$ if $y=0$),

$$g_\delta(s,x) \;=\; \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin^2 y}{y^2}g(s,x+\delta y)dy.$$

Simple calculation shows that, for some constants $c_1>0$ and $c_2>0$,

$$c_1\,/(1+|x|^{1+b/2}) \leq f(x) \leq c_2\,/(1+|x|^{1+b/2}). \tag{7.7}$$

Furthermore, for any $\epsilon>0$, there exists $\delta_0>0$ such that, whenever $0<\delta\leq\delta_0$,

$$\begin{aligned} &\sup_{\substack{0\leq s\leq 1\\ x\in R}} |g_\delta(s,x)-g(s,x)| \\ \leq\;& \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin^2 y}{y^2}\sup_{\substack{0\leq s\leq 1\\ x\in R}}\big|g(s,x+\delta y)-g(s,x)\big|dy \;\leq\; \epsilon, \end{aligned} \tag{7.8}$$

$$\begin{aligned} &\int_{-\infty}^{\infty}\sup_{0\leq s\leq 1}|g_\delta(s,x)-g(s,x)|dx \\ \leq\;& \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin^2 y}{y^2}\int_{-\infty}^{\infty}\sup_{0\leq s\leq 1}\big|g(s,x+\delta y)-g(s,x)\big|dxdy \;\leq\; \epsilon, \end{aligned} \tag{7.9}$$

where we have used the facts that $\int_{-\infty}^{\infty}\frac{\sin^2 x}{x^2}dx = \pi$ and $g(s,x)$ is uniformly continuous on $[0,1]\times[-A,A]$ for any $A\in R$ satisfying $\int_{-\infty}^{\infty}\sup_{0\leq s\leq 1}\big|g(s,x)\big|dx<\infty$.

It follows from (7.7)-(7.9) that, for small $\epsilon>0$, there exists an $\delta>0$ such that

$$\begin{aligned} g(s,x) &\leq g^+(s,x) := g_\delta(s,x) + c_1^{-1}\,\epsilon f(x) \\ &\leq g(s,x) + 2c_1^{-1}\,\epsilon f(x) \leq g(s,x) + 2c_1^{-1}c_2\,\epsilon\,(1+|x|^{-1-b}), \\ g(s,x) &\geq g^-(s,x) := g_\delta(s,x) - c_2^{-1}\,\epsilon f(x) \\ &\geq g(s,x) - 2c_2^{-1}\,\epsilon f(x) \geq g(s,x) - 2c_2^{-1}c_1\,\epsilon\,(1+|x|^{-1-b}). \end{aligned}$$

Now, to show (2.5), it suffices to verify that

$$\text{both } g^+(s,x) \text{ and } g^-(s,x) \text{ satisfy the } \mathbf{Con}. \tag{7.10}$$

Indeed, in terms of (7.10), it follows from the result proved above that

$$\begin{aligned}
\frac{c_n}{n}\sum_{k=1}^n g\Big(\frac{k}{n},\, c_n\, X_{nk}\Big) &\le \frac{c_n}{n}\sum_{k=1}^n g^+\Big(\frac{k}{n},\, c_n\, X_{nk}\Big) \to_D \int_0^1 G^+(t)L_X(dt,0),\\
\frac{c_n}{n}\sum_{k=1}^n g\Big(\frac{k}{n}, c_n\, X_{nk}\Big) &\ge \frac{c_n}{n}\sum_{k=1}^n g^-\Big(\frac{k}{n},\, c_n\, X_{nk}\Big) \to_D \int_0^1 G^-(t)L_X(dt,0)
\end{aligned}$$

where $G^{\pm}(t)=\int_{-\infty}^{\infty} g^{\pm}(t,x)dx$. Since $G^+(t)-G^-(t)=\epsilon\,(c^{-1}+c_2^{-1})\int_{-\infty}^{\infty} f(x)dx$, the result (2.5) is established due to the arbitrary of ϵ.

The verification of (7.10) is simple. In fact, it is trivial to see that $g^+(s,x)$ is continuous satisfying $\int_{-\infty}^{\infty}\sup_{0\le s\le 1}|g^+(s,x)|dx<\infty$. On the other hand, by noting that

$$\int_{-\infty}^{\infty}\frac{\sin^2 t}{t^2}e^{itx}dt=\begin{cases}\pi(1-|x|/2), & \text{if } |x|<2,\\ 0, & \text{otherise,}\end{cases}$$

we have

$$\begin{aligned}
\widehat{g}^+(s,x) &= \int_{-\infty}^{\infty} e^{itx} g^+(s,t)dt,\\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} g(s,t)e^{itx}dt\int_{-\infty}^{\infty}\frac{\sin^2 y}{y^2}e^{i\delta x\, y}dy\\
&\quad +c_1^{-1}\epsilon\sum_{n=1}^{\infty} n^{-1-b/2}e^{ixn}\int_{-\infty}^{\infty}\frac{\sin^2 y}{y^2}e^{iy\,x}dy\\
&= 0,
\end{aligned}$$

for all $0\le s\le 1$ and $\min\{|x|,|x|\delta\}\ge 2$. Hence $g^+(s,x)$ satisfies the **Con**. The verification for $g^-(s,x)$ is similar and the details are omitted.

We finally show that (2.5) still holds when $g(s,x)$ satisfies (b) without the assumption of continuity. In fact, since $g_{2\delta}(s,x)$ is continuous on $[0,1]\times\mathbb{R}$ satisfying $|g_{2\delta}(s,x)|\le C/(1+|x|^{1+b})$, it follows from the result in step 1 that, for any $\delta>0$,

$$\frac{c_n}{n}\sum_{k=1}^n g\Big(\frac{k}{n},\, c_n\, X_{nk}\Big) \le \frac{c_n}{n}\sum_{k=1}^n g_{2\delta}\Big(\frac{k}{n},\, c_n\, X_{nk}\Big) \to_D \int_0^1 G_{2\delta}(t)L_X(dt,0),$$

where $G_{2\delta}(t)=\int_{-\infty}^{\infty} g_{2\delta}(t,x)dx$. Similarly, for any $\delta>0$, we have

$$\frac{c_n}{n}\sum_{k=1}^n g\Big(\frac{k}{n},\, c_n\, X_{nk}\Big) \ge \frac{c_n}{n}\sum_{k=1}^n g_{1\delta}\Big(\frac{k}{n},\, c_n\, X_{nk}\Big) \to_D \int_0^1 G_{1\delta}(t)L_X(dt,0),$$

where $G_{1\delta}(t)=\int_{-\infty}^{\infty} g_{1\delta}(t,x)dx$. Since, for $i=1$ and 2,

$$\sup_{0\le t\le 1}\Big|G_{i\delta}(t)-\int_{-\infty}^{\infty} g(t,x)dx\Big| \le \int_{-\infty}^{\infty}\sup_{0\le s\le 1}|g_{i\delta}(s,x)-g(s,x)|dx\le\delta,$$

the result (2.5) follows by taking $\delta\to 0$. The proof of Theorem 2.1 is now complete. □

7.3 Proof of Theorem 2.2

We only prove (2.7) when $g(s,x)$ is continuous on $[0,1]\times\mathbb{R}$ satisfying the additional **Con**. Since v_k is positive, the remaining proof is similar to that of Theorem 2.1 with minor modification. We omit the details.

As in the proof of Theorem 2.1, similarly to (7.4), we may write

$$\frac{c_n}{n}\sum_{k=1}^{n} g\Big(\frac{k}{n}, c_n X_{nk}\Big)v_k = A_0\,\frac{c_n}{n}\sum_{k=1}^{n} g\Big(\frac{k}{n}, c_n X_{nk}\Big) + \frac{1}{2\pi}\int_{-ac_n}^{ac_n} A_n(t)dt, \qquad (7.11)$$

where $A_n(t)=\frac{1}{n}\sum_{k=1}^{n}\widehat{g}\big(\frac{k}{n}, -\frac{t}{c_n}\big)\,e^{it\,X_{nk}}\widetilde{v}_k$ and $\widetilde{v}_k = v_k - A_0$. Recall that $|\widehat{g}(s,x)|\le\int_{-\infty}^{\infty}\widetilde{g}(t)dt := C_1<\infty$ and note that, under condition (iii),

$$|\widehat{g}(s,x)-\widehat{g}(s',x)|\le \widehat{C}\,|s-s'|,$$

uniformly on $[0,1]\times\mathbb{R}$, where $\widehat{C} = C\int_{-\infty}^{\infty}\frac{1}{1+|x|^{1+b}}dx$. By letting $T_n=[n/m]$ and using $\sup_{k\ge1}Ev_k<\infty$, we have

$$\begin{aligned}
|A_n(t)| &\le \frac{1}{n}\sum_{j=0}^{T_n}\Big|\sum_{k=jm+1}^{(j+1)m}\widehat{g}\Big(\frac{k}{n}, -\frac{t}{c_n}\Big)\,e^{it\,X_{nk}}\,\widetilde{v}_k\Big| + \frac{C_1}{n}\sum_{k=mT_n+1}^{n}|\,\widetilde{v}_k|\\
&\le \frac{1}{n}\sum_{j=0}^{T_n}\Big|\sum_{k=jm+1}^{(j+1)m}\widehat{g}\Big(\frac{k}{n}, -\frac{t}{c_n}\Big)\,\widetilde{v}_k\Big| + \frac{C_1}{n}\sum_{k=mT_n+1}^{n}|\widetilde{v}_k|\\
&\quad + C|t|\max_{0\le j\le T_n}\max_{jm<k\le(j+1)m}|X_{nk}-X_{n,jm}|\,\frac{1}{n}\sum_{k=1}^{n}|\widetilde{v}_k|\\
&\le \frac{C_1}{n}\sum_{j=0}^{T_n}\Big|\sum_{k=jm+1}^{(j+1)m}\widetilde{v}_k\Big| + O_P(|t|)\sup_{0\le u\le1}\sup_{|u-v|\le m/n}|X_{n,[nu]}-X_{n,[nv]}|\\
&\quad + O_P\Big(\frac{|n-mT_n|+m}{n}\Big). \qquad (7.12)
\end{aligned}$$

Since $m/n = o(c_n^{-2/\eta})$, under (2.4) and condition (i), we have

$$\begin{aligned}
&\sup_{0\le u\le1}\sup_{|u-v|\le m/n}|X_{n,[nu]}-X_{n,[nv]}|\\
\le\ &\sup_{0\le u\le1}\sup_{|u-v|\le m/n}|X_u-X_v| + o_P(c_n^{-2})\\
=\ &O_P(1)(m/n)^{\eta} + o_P(c_n^{-2}) = o_P(c_n^{-2}).
\end{aligned}$$

This, together with condition (ii) and (7.12), yields $\sup_{|t|\le ac_n}|A_n(t)| = o_P(c_n^{-1})$. Taking this estimate into (7.11), the required (2.7) is proved. □

7.4 Proof of Theorem 3.1

Write

$$S_n = \frac{1}{\sqrt{n^2 b_n^3}} \sum_{k=1}^n u_k\, y_{k-1} I(|y_{k-1}| \le b_n \sqrt{n}),$$
$$V_n^2 = \frac{1}{n^2 b_n^3} \sum_{k=1}^n y_{k-1}^2 I(|y_{k-1}| \le b_n \sqrt{n}).$$

It suffices to show that, as $n \to \infty$,

$$\big(S_n,\ V_n^2\big) \to_D \Big(\Big[\sigma \int_{|x|\le 1} x^2 dx\, L_{J_\tau}(1,0)\Big]^{1/2} N,\ \ \sigma^{-1} \int_{|x|\le 1} x^2 dx\, L_{J_\tau}(1,0)\Big). \tag{7.13}$$

Indeed, by noting that

$$\widehat{\alpha}_{1n} - \alpha = \frac{\sum_{k=1}^n u_k\, y_{k-1} I(|y_{k-1}| \le b_n\sqrt{n})}{\sum_{k=1}^n y_{k-1}^2 I(|y_{k-1}| \le b_n \sqrt{n})}, \tag{7.14}$$

the required results follow from a simple algebra by using (7.13), $\int_{|x|\le 1} x^2 dx = 2/3$ and the continuous mapping theorem.

To prove (7.13), we start with some preliminaries. Let $z_k = \sum_{i=1}^k u_i$ and $z_0 = 0$. Recall that $\{u_k, \mathcal{F}_k\}_{k\ge 1}$ forms a martingale difference with

$$\sigma^2 = E u_k^2 \quad \text{and} \quad \sup_{k\ge 1} E|u_k|^{2+\delta} < \infty, \tag{7.15}$$

for some $\delta > 0$. Since $\sigma_{k,A} := \sigma_k^2 I(\sigma_k^2 \ge A)$, for each $A \ge 0$, keeps a stationary α-mixing random sequence with $E\sigma_{1,A}^{1+\delta/2} \le E\sigma_1^{2+\delta} < \infty$ and coefficients $\alpha(n) \le C n^{-(1+\delta)}$, it follows from Lemma 2 of **?** with a simple algebra that there exists a positive constant K depending only on δ such that

$$P\big(\max_{1\le m\le n} \big|\sum_{k=1}^m \big(\sigma_{k,A} - E\sigma_{k,A}\big)\big| \ge x\big) \le C\, K\, n x^{-1-\delta/(4+3\delta)}, \tag{7.16}$$

for every $x \ge K n^{1/2} \log n$. Let $Y_n = \max_{1\le m\le n} \big|\sum_{k=1}^m \big(\sigma_{k,A} - E\sigma_{k,A}\big)\big|$. It follows from (7.16) that, when n is sufficiently large,

$$\begin{aligned} EY_n &= \int_0^\infty P(Y_n \ge y) dy \le n^{1-\delta/4(1+\delta)} + C\, K n \int_{n^{1-\delta/4(1+\delta)}}^\infty x^{-1-\delta/(4+3\delta)} dx \\ &\le C_\delta\, n^{1-\delta/4(1+\delta)} \end{aligned}$$

for some constant C_δ depending only on δ. This, together with the stationarity of $\sigma_{k,A}$ and the fact that $\sigma_{k,0} = \sigma_k^2 = E(u_k^2|\mathcal{F}_{k-1})$, implies that

$$\max_{1\le m\le n} \Big|\sum_{k=1}^m E(u_k^2|\mathcal{F}_{k-1}) - m\,\sigma^2\Big| = O_P(n^{1-\delta/4(1+\delta)}) \tag{7.17}$$

and, for each $A \geq 0$ and $m = m_n \to \infty$ satisfying $m/n \to 0$,

$$\max_{0\leq j\leq n-m} E\Big|\frac{1}{m}\sum_{k=j+1}^{j+m} \sigma_{k,A} - E\sigma_{1,A}\Big| \;=\; O(m^{-\delta/4(1+\delta)}). \tag{7.18}$$

Due to (7.15) and (7.17), the strong approximation result from Proposition 2.1 shows that, on a richer probability space, there exists a Brownian motion $B = \{B_t\}_{t\geq 0}$ and a random array $\{z^*_{nk}, n \geq 1, k = 1, 2, \cdots, n\}$ such that $\{z^*_{nk}, k = 1, 2, \cdots, n\} =_d \{z_k, k = 1, 2, \cdots, n\}$ for all $n \geq 1$ and

$$\sup_{0\leq t\leq 1} \big|\frac{1}{\sqrt{n}\sigma}\, z^*_{n,\lfloor nt\rfloor} - B_t\big| = O_P(n^{-\eta}), \tag{7.19}$$

where $\eta > 0$ is a constant depending only on δ. On the other hand, by noting that

$$\begin{aligned} y_k &= \sum_{i=1}^{k}\Big(1+\frac{\tau}{n}\Big)^{k-i} u_i + \Big(1+\frac{\tau}{n}\Big)^{k} y_0 \\ &= \sum_{i=1}^{k}\Big(1+\frac{\tau}{n}\Big)^{k-i}(z_i - z_{i-1}) + \Big(1+\frac{\tau}{n}\Big)^{k} y_0 \\ &= z_k + \frac{\tau}{n}\sum_{i=1}^{k-1}\Big(1+\frac{\tau}{n}\Big)^{k-i-1} z_i + \Big(1+\frac{\tau}{n}\Big)^{k} y_0, \end{aligned}$$

we have

$$\frac{y_{\lfloor nt\rfloor}}{\sqrt{n}} \;=\; \frac{z_{\lfloor nt\rfloor}}{\sqrt{n}} + \tau\int_0^t e^{\tau(t-s)}\frac{z_{[ns]}}{\sqrt{n}}ds + R_n(t), \tag{7.20}$$

where $\sup_{0\leq t\leq 1}|R_n(t)| = O_P(n^{-1/2})$. This, together with (7.19), indicates that there exists a random array $\{y^*_{nk}, n \geq 1, k = 1, 2, \cdots, n\}$ such that

$$\{(y^*_{nk}, z^*_{nk}), k = 1, 2, \cdots, n\} =_d \{(y_k, z_k), k = 1, 2, \cdots, n\}$$

for all $n \geq 1$, and

$$\sup_{0\leq t\leq 1} \big|\frac{1}{\sqrt{n}\sigma} y^*_{n,\lfloor nt\rfloor} - J_\tau(t)\big| = O_P(n^{-\eta}) \tag{7.21}$$

for some $\eta > 0$, where $J_\tau(t) = B_t + \tau\int_0^t e^{\tau(t-s)}B_s ds$. Since $J_\tau(t)$ is a Gaussian process satisfying

$$\sup_{\substack{|s-t|\leq\Delta\\ 0\leq t\leq 1}} |J_\tau(t) - J_\tau(s)| \leq C_\tau \sup_{\substack{|s-t|\leq\Delta\\ 0\leq t\leq 1}} |B_t - B_s| \;=\; O_P(\Delta^{1/2}),$$

it follows easily from Theorem 2.2 with $X_{nk} = y_{n,k-1} = y_{k-1}/\sqrt{n}\sigma$ and $v_k = \sigma_{k,A}$ [recall (7.18)] that, for any $b_n > 0$ satisfying $b_n \to 0$ and $b_n \log^K n \to \infty$ for some $K > 0$,

$$\begin{aligned} &\frac{1}{nb_n}\sum_{k=2}^{n}\sigma_{k,A}\, h(y_{k-1}/b_n\sqrt{n}) I(|y_{k-1}| \leq b_n\sqrt{n}) \\ &= \frac{1}{nb_n}\sum_{k=2}^{n}\sigma_{k,A}\, H(y_{n,k-1}/b_n) = \frac{E\sigma_{1,A}}{nb_n}\sum_{k=1}^{n} H(y_{nk}/b_n) + o_P(1), \end{aligned} \tag{7.22}$$

for each $A \ge 0$, where $H(x) = h(\sigma x) I(\sigma |x| \le 1)$ and $h(x)$ is a continuous function. By taking $A = 0$, as a consequence of (7.22) and Theorem 2.1, we have

$$\begin{aligned} & \Big(\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^{\lfloor nt \rfloor} u_i, \ \frac{1}{nb_n} \sum_{k=2}^{n} \sigma_k^2 H(y_{n,k-1}/b_n) \Big) \\ = & \Big(\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^{\lfloor nt \rfloor} u_i, \ \frac{\sigma^2}{nb_n} \sum_{k=1}^{n} H(y_{nk}/b_n) \Big) + o_P(1) \\ \Rightarrow & \big(B_t, \ \sigma \int_{|x| \le 1} h(x) dx \, L_{J_\tau}(1, 0) \big), \end{aligned} \tag{7.23}$$

on $D_{R^2}[0,1]$, where $L_{J_\tau}(t,s)$ is the local time process of the linear diffusion $J_\tau = \{J_\tau(t)\}_{t \ge 0}$ and we have used the fact: $\int_{-\infty}^{\infty} H(x) dx = \sigma^{-1} \int_{|x| \le 1} h(x) dx$.

We are now ready to prove (7.13). Let $S_n = \sum_{k=1}^n x_{nk} u_k$, where

$$x_{nk} = \frac{1}{n b_n^{3/2}} y_{k-1} I(|y_{k-1}| \le b_n \sqrt{n}).$$

First note that $\{u_k, \mathcal{F}_k\}_{k \ge 1}$ is a martingale difference and x_{nk} is a function of $u_{k-1}, u_{k-2}, ..., u_1$ so that $\{S_n\}_{n \ge 1}$ is a martingale having the conditional variance:

$$\widetilde{V}_n^2 = \frac{1}{n^2 b_n^3} \sum_{k=1}^n \sigma_k^2 y_{k-1}^2 I(|y_{k-1}| \le b_n \sqrt{n}).$$

Since result (7.23) with $h(x) = x^2$ yields that, on $D_{R^2}[0,1]$,

$$\begin{aligned} \Big(\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^{\lfloor nt \rfloor} u_i, \ \ \widetilde{V}_n^2 \Big) &= \Big(\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^{\lfloor nt \rfloor} u_i, \ \ \sigma^2 V_n^2 \Big) + o_P(1) \\ &\Rightarrow \Big(B_t, \ \sigma \int_{|x| \le 1} x^2 dx \, L_{J_\tau}(1, 0) \Big), \end{aligned}$$

by using Wang's extended martingale limit theorem(c.g. Theorem 3.13, **?** with $X_{nk} = x_{nk} u_k$ and $Z_{nk} = u_k/\sqrt{n}$), (7.13) will follow if we prove:

$$\frac{1}{n} \sum_{k=1}^n E\big[u_k^2 I(|u_k| \ge \epsilon \sqrt{n}) | \mathcal{F}_{k-1} \big] \to_P 0, \quad \text{for any } \epsilon > 0, \tag{7.24}$$

$$\sum_{k=1}^n x_{nk}^2 E\big[u_k^2 I(|x_{nk} u_k| \ge \epsilon) | \mathcal{F}_{k-1} \big] \to_P 0, \quad \text{for any } \epsilon > 0, \tag{7.25}$$

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n |x_{nk}| E(u_k^2 | \mathcal{F}_{k-1}) \to_P 0. \tag{7.26}$$

The proof of (7.26) is simple. In fact, by noting $nb_n \to \infty$, it follows from (7.23) with $h(x) = |x|$ $(H(x) = h(\sigma x) I(\sigma |x| \le 1))$ that

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n |x_{nk}| E(u_k^2 | \mathcal{F}_{k-1}) = \frac{1}{\sqrt{n b_n}} \frac{1}{n b_n} \sum_{k=1}^n \sigma_k^2 H(y_{n,k-1}/b_n) \to_P 0,$$

as required. As for (7.25), for any $\epsilon > 0$ and $A > 0$, it follows from $u_k = \sigma_k \epsilon_k$, σ_k is adapted to $\mathcal{F}_{k-1}$ and $E(\epsilon_k^2 | \mathcal{F}_{k-1}) = 1$ that

$$\begin{aligned}
& \sum_{k=1}^n x_{nk}^2 E\big[u_k^2 I(|x_{nk} u_k| \geq \epsilon) | \mathcal{F}_{k-1}\big] \\
\leq & \sum_{k=1}^n x_{nk}^2 E\big[u_k^2 I(|u_k| \geq A) | \mathcal{F}_{k-1}\big] + \sum_{k=1}^n x_{nk}^2 I(|x_{nk}| \geq \epsilon / A) E\big[u_k^2 | \mathcal{F}_{k-1}\big] \\
\leq & \sum_{k=1}^n x_{nk}^2 \sigma_k^2 I(\sigma_k \geq \sqrt{A}) + \sup_{k \geq 1} E\big[\epsilon_k^2 I(|\epsilon_k| \geq \sqrt{A}) | \mathcal{F}_{k-1}\big] \sum_{k=1}^n x_{nk}^2 \sigma_k^2 \\
& \quad + A^2 \epsilon^{-2} \sum_{k=1}^n x_{nk}^4 \, \sigma_k^2 \\
= & \frac{1}{n b_n} \sum_{k=1}^n H_1(y_{n,k-1}/b_n) \sigma_k^2 I(\sigma_k \geq \sqrt{A}) + \sup_{k \geq 1} E\big[\epsilon_k^2 I(|\epsilon_k| \geq \sqrt{A}) | \mathcal{F}_{k-1}\big] \frac{1}{n b_n} \sum_{k=1}^n H_1(y_{n,k-1}/b_n) \sigma_k^2 \\
& \quad + A^2 \epsilon^{-2} \frac{1}{(n b_n)^2} \sum_{k=1}^n H_2(y_{n,k-1}/b_n) \sigma_k^2, \qquad (7.27)
\end{aligned}$$

where $H_1(x) = (\sigma\, x)^2 I(\sigma\, |x| \leq 1)$ and $H_2(x) = (\sigma\, x)^4 I(\sigma\, |x| \leq 1)$. Since $n b_n \to \infty$ and $E \sigma_1^2 I(\sigma_1 \geq \sqrt{A}) \to 0$ and $\sup_{k \geq 1} E\big[\epsilon_k^2 I(|\epsilon_k| \geq \sqrt{A}) | \mathcal{F}_{k-1}\big] \to_P 0$ as $A \to \infty$, it follows easily from (7.22) and (7.27) that

$$\sum_{k=1}^n x_{nk}^2 E\big[u_k^2 I(|x_{nk} u_k| \geq \epsilon) | \mathcal{F}_{k-1}\big] \to_P 0,$$

as $n \to \infty$ first and then $A \to \infty$. This proves (7.25). Similarly, for $\epsilon > 0$, we have

$$\begin{aligned}
& \frac{1}{n} \sum_{k=1}^n E\big[u_k^2 I(|u_k| \geq \epsilon \sqrt{n}) | \mathcal{F}_{k-1}\big] \\
\leq & \frac{1}{n} \sum_{k=1}^n \sigma_k^2 I(\sigma_k \geq \log n) + \sup_{k \geq 1} E\big[\epsilon_k^2 I(|\epsilon_k| \geq \epsilon \sqrt{n} / \log n) | \mathcal{F}_{k-1}\big] \frac{1}{n} \sum_{k=1}^n \sigma_k^2 \\
= & o_P(1),
\end{aligned}$$

implying (7.24). (7.13) is now proved and we also complete the proof of Theorem 3.1. □

7.5 Proof of Theorem 3.2

We start with some preliminaries. Let $\Theta(z) = 1 - \sum_{j=1}^{p-1} \theta_j z^j$, L be the usual lag operator and $\Theta(L) \widetilde{u}_k = u_k$. Since $\sum_{j=1}^{p-1} |\theta_j| < 1$, $\Theta(z)$ is invertible when $|z| \leq 1$ so that

$$\widetilde{u}_k = \Theta^{-1}(L)\, u_k = \sum_{j=0}^{\infty} \pi_j u_{k-j}, \quad \text{where} \quad \Theta^{-1}(z) = \sum_{j=0}^{\infty} \pi_j z^j \text{ and } u_{-i} = 0 \text{ for } i \geq 0.$$

It follows from **?** that

$$\widetilde{u}_k = \Theta^{-1}(1)\, u_k + \hat{u}_{k-1} - \hat{u}_k, \qquad (7.28)$$

where $\hat{u}_k = \sum_{i=0}^{\infty} \hat{\pi}_i u_{k-i}$ with $\hat{\pi}_i = \sum_{j=i+1}^{\infty} \pi_j$, satisfying that $E\hat{u}_k^2 = \sigma^2 \sum_{i=0}^{\infty} \hat{\pi}_i^2 < \infty$ and (recall $u_{-i} \geq 0$ if $i \geq 0$)

$$\max_{1\leq k\leq n} |\hat{u}_k| \quad \leq \quad \sum_{i=0}^{\infty} |\hat{\pi}_i| \max_{1\leq k\leq n} |u_k| = O_P(n^{1/(2+\delta)}), \tag{7.29}$$

where we have used the facts that $\sum_{i=0}^{\infty} |\hat{\pi}_i| \leq \sum_{i=0}^{\infty} i|\pi_i| < \infty$ and, due to (7.15),

$$\max_{1\leq k\leq n} |u_k| \leq \big(\sum_{k=1}^{n} |u_k|^{2+\delta}\big)^{1/(2+\delta)} = O_P(n^{1/(2+\delta)}).$$

Consequently, we have $\sup_{k\geq 1} E\widetilde{u}_k^2 < \infty$ and, for any $m \geq 1$ and $j \geq 0$,

$$\max_{0\leq j\leq n-m} E\Big|\sum_{k=j+1}^{j+m} \widetilde{u}_k\Big| \quad \leq \quad C\, m^{1/2}, \tag{7.30}$$

for some constant $C > 0$. Furthermore, when $\alpha = 1$, we may rewrite (3.5) as

$$y_k = y_{k-1} + \widetilde{u}_k, \tag{7.31}$$

so that $z_k = (\widetilde{u}_{k-1}, ..., \widetilde{u}_{k-p})$ (recall $\widetilde{u}_j = 0$ if $j < 0$). See **?**, for instance.

The proof of (3.6) now follows from the following lemmas. Define $||z|| = (z_1^2 + ... + z_p^2)^{1/2}$ for a vector $z = (z_1, \cdots, z_p)$ and $||A|| = \max_z ||Az||/||z||$ for a $p \times p$ matrix A.

Lemma 7.1. *When $\alpha = 1$, we have*

(a) $||\big(\sum_{k=1}^{n} z_k' z_k\big)^{-1}|| = O_P(n^{-1})$;

(b) $||\sum_{k=1}^{n} z_k y_{k-1}|| = O_P(n)$;

(c) $||\sum_{k=1}^{n} z_k u_k|| = O_P(\sqrt{n})$;

(d) $\Big|\Big|\sum_{k=1}^{n} z_k y_{k-1} I(|y_{k-1}| \leq b_n\sqrt{n})\Big|\Big| = O_P(n^{3/2} b_n^2)$.

Proof. (a)-(b) follows from Lemma 3.2 of **?**. See, also, **?**. (c) is obvious by using the martingale properties.

We next prove (d), starting with some preliminaries. It follows from (7.28) and (7.31) that

$$y_n = \Theta^{-1}(1) \sum_{k=1}^{n} u_k + \hat{u}_0 - \hat{u}_n.$$

Since $|\hat{u}_0| + \max_{1\leq k\leq n} |\hat{u}_k| = O_P(n^{1/(2+\delta)})$ for some $\delta > 0$ by using (7.29), on a richer probability space, there exists a Brownian motion $B = \{B(t)\}_{t\geq 0}$ so that, for some $\eta > 0$,

$$\sup_{0\leq t\leq 1} \Big| y_{n,\lfloor nt\rfloor} - \frac{\Theta^{-1}(1)B(nt)}{\sqrt{n}} \Big| = O_P(n^{-\eta}), \tag{7.32}$$

where $y_{nk} = y_k/(\sqrt{n}\,\sigma)$. As a consenqence, for any $m \to \infty$ and $m/n \to 0$, we have

$$\begin{aligned}
&\sup_{0\le u\le 1}\sup_{|u-v|\le m/n}|y_{n,[nu]}-y_{n,[nv]}| \\
\le\ & \Theta^{-1}(1)\sup_{0\le u\le 1}\sup_{|u-v|\le m/n}\Big|\frac{B(nu)}{\sqrt{n}}-\frac{B(nv)}{\sqrt{n}}\Big| + o_P(n^{-\eta}) \\
=\ & O_P(1)(m/n)^{1/2}\,\log(m/n)+o_P(n^{-\eta}). \qquad (7.33)
\end{aligned}$$

Let $H_n(x)$ be a real function satisfying the following condition: for any $x, y \in R$,

$$|H_n(x)-H_n(y)| \ \le\ C\,b_n^{-1}\,|x-y| \quad \text{and} \quad \sup_{x,n}|H_n(x)|\le C.$$

Recall that $z_k = (\widetilde{u}_{k-1}, ..., \widetilde{u}_{k-p})$. In terms of (7.30) and (7.33), by letting $T_n = [n/m]$ and choosing m such that $(m/n)^{1/2}\log(m/n) = O(b_n^3)$ and $mb_n^2 \to \infty$, we have

$$\begin{aligned}
\Big|\Big|\sum_{k=1}^n z_k H_n(b_n^{-1}\,y_{nk})\Big|\Big| \ \le\ & \sum_{j=0}^{T_n}\Big|\Big|\sum_{k=jm+1}^{(j+1)m} z_k H_n(b_n^{-1}\,y_{nk})\Big|\Big| + C\sum_{k=mT_n+1}^{n}||z_k|| \\
\le\ & C\sum_{j=0}^{T_n}\Big|\Big|\sum_{k=jm+1}^{(j+1)m} z_k\Big|\Big| + C\sum_{k=mT_n+1}^{n}||z_k|| \\
& + \max_{0\le j\le T_n}\max_{jm<k\le(j+1)m}|H_n(b_n^{-1}\,y_{nk})-H_n(b_n^{-1}\,y_{n,(j+1)m})|\sum_{k=1}^n||z_k|| \\
\le\ & O_P\big[n\,m^{-1/2}+(n-mT_n)\big]+O_P(n\,b_n^{-2})\sup_{0\le u\le 1}\sup_{|u-v|\le m/n}|y_{n,[nu]}-y_{n,[nv]}| \\
=\ & O_P(nb_n). \qquad (7.34)
\end{aligned}$$

We are now ready to prove (d). We may write

$$\Big|\Big|\sum_{k=1}^n z_k y_{k-1} I(|y_{k-1}|\le b_n\sqrt{n})\Big|\Big| = \sigma\,\sqrt{n}b_n\,\Big|\Big|\sum_{k=1}^n z_k l(\sigma\,b_n^{-1}y_{n,k-1})\Big|\Big|$$

where $l(x) = xI(|x| \le 1)$. Let $l_{1n}(x)$ be a continuous function so that

$$l_{1n}(x) = \begin{cases} x, & |x|\le 1-b_n, \\ 0, & |x|\ge 1+b_n, \\ \text{linear}, & 1-b_n<|x|<1+b_n. \end{cases}$$

and $l_{2n}(x) = l(x) - l_{1n}(x)$. Since $|l_{1n}(x)| \le 1$ and

$$|l_{1n}(x)-l_{1n}(y)|\le\max\{(2b_n)^{-1},1\}\,|x-y|,$$

it follows from (7.34) that

$$\Big|\Big|\sum_{k=1}^n z_k l_{1n}(b_n^{-1}\,y_{nk})\Big|\Big| = O_P(nb_n).$$

Hence result (d) will follow if we prove

$$R_n \ := \Big|\Big|\sum_{k=1}^n z_k l_{2n}(b_n^{-1}\, y_{nk})\Big|\Big| = O_P(nb_n). \tag{7.35}$$

Note that $|l_{2n}(x)| \le I(1-b_n \le |x| \le 1+b_n)$. Applying the Cauchy-Schwarz inequality and recalling $\sup_{k\ge 1} E||z_k||^2 < \infty$ gives

$$\begin{aligned} R_n &\le \sum_{k=1}^n I(1-b_n \le |b_n^{-1} y_{n,k}| \le 1+b_n)||z_k|| \\ &\le \Big(\sum_{k=1}^n I(1-b_n \le |b_n^{-1} y_{nk}| \le 1+b_n)\sum_{k=1}^n ||z_k||^2\Big)^{1/2} \\ &= O_P(n^{1/2})\Big(\sum_{k=1}^n I(1-b_n \le |b_n^{-1} y_{nk}| \le 1+b_n)\Big)^{1/2}. \end{aligned} \tag{7.36}$$

Write $B_n(t) = \Theta^{-1}(1)B(nt)/\sqrt{n}$ for $0 \le t \le 1$, and let

$$\Omega_n = \Big\{\sup_{0\le t\le 1} |y_{n,\lfloor nt\rfloor} - B_n(t)| \le b_n^2\Big\}.$$

Then on Ω_n, we have

$$\begin{aligned} &\sum_{k=1}^n I(1-b_n \le |b_n^{-1} y_{nk}| \le 1+b_n) \\ = \;& n\int_0^1 I(1-b_n \le |b_n^{-1} y_{n,\lfloor nt\rfloor}| \le 1+b_n)dt + O_P(1) \\ \le \;& n\int_0^1 I(1-2b_n \le |b_n^{-1} B_n(t)| \le 1+2b_n)dt + O_P(1) = O_P(nb_n^2), \end{aligned} \tag{7.37}$$

where we have used the fact that, by writing $a_0 = \sqrt{|\Theta^{-1}(1)|}$,

$$\begin{aligned} &\int_0^1 I(1-2b_n \le |b_n^{-1} B_n(t)| \le 1+2b_n)dt \\ \stackrel{d}{=} \;& \int_0^1 I(1-2b_n \le |b_n^{-1} B(a_0\, t)| \le 1+2b_n)dt \\ = \;& a_0^{-1}\int_0^{a_0} I(1-2b_n \le |b_n^{-1} B(t)| \le 1+2b_n)dt \\ = \;& a_0^{-1}\int_{-\infty}^{\infty} I(1-2b_n \le |b_n^{-1} x| \le 1+2b_n)L_B(a_0, x)dx \\ = \;& a_0^{-1} b_n\int_{-\infty}^{\infty} I(1-2b_n \le |x| \le 1+2b_n)L_B(a_0, xb_n)dx \\ = \;& 8a_0^{-1} b_n^2 L_B(a_0, 0)(1+o_P(1)). \end{aligned}$$

Note that (7.32) implies $\lim_{n\to\infty} \mathbb{P}(\Omega_n) = 1$ for any $b_n > 0$ satisfying $b_n \to 0$ and $b_n \log^K n \to \infty$ for some $K > 0$. Now (7.35) follows from (7.36)−(7.37) and hence the proof of (d) is complete. □

Lemma 7.2. *Recall $\widetilde{y}_k = y_k I(|y_k| \le b_n\sqrt{n})$. When $\alpha = 1$, we have*

$$\Big(\frac{1}{\sqrt{n^2 b_n^3}}\sum_{k=1}^n \widetilde{y}_{k-1}\, u_k,\ \frac{1}{n^2 b_n^3}\sum_{k=1}^n \widetilde{y}_{k-1}^2\Big) \quad \to_D \quad \big(\sigma\, Z^{1/2}\, N,\ Z\big), \tag{7.38}$$

where $Z = \sigma^{-1}\,\Theta^{-2}(1)\int_{|x|\le 1} x^2 dx\, L_B(1,0)$ and N is a standard normal variate independent of the local time $L_B(t,x)$ of Brownian motion $B = \{B_t\}_{t\ge 0}$.

Proof. The proof is similar to that of (7.13) by using (7.32) and Theorem 2.2. The details are omitted. □

We now come back to the proofs of (3.6) and (3.7). Note that, when $\alpha = 1$,

$$\widehat{\alpha}_{2n} - 1 \quad = \quad A_n/B_n, \tag{7.39}$$

where $\widetilde{y}_{k-1} = y_{k-1} I(|y_{k-1}| \le b_n\sqrt{n})$ and

$$\begin{aligned} A_n &= \sum_{k=1}^n \widetilde{y}_{k-1}\, u_k - \sum_{k=1}^n z_k\, \widetilde{y}_{k-1}\Big(\sum_{k=1}^n z_k' z_k\Big)^{-1}\sum_{k=1}^n z_k'\, u_k, \\ B_n &= \sum_{k=1}^n y_{k-1}\widetilde{y}_{k-1} - \sum_{k=1}^n z_k\, \widetilde{y}_{k-1}\Big(\sum_{k=1}^n z_k' z_k\Big)^{-1}\sum_{k=1}^n z_k'\, y_{k-1}. \end{aligned}$$

Using Lemmas 7.1 and 7.2, simple calculations show that

$$\begin{aligned} &\Big[\sum_{k=1}^n y_{k-1}^2 I(|y_{k-1}| \le b_n\sqrt{n})\Big]^{1/2}\big(\widehat{\alpha}_{2n} - \alpha\big) \\ =\ & \Big(\frac{1}{n^2 b_n^3}\sum_{k=1}^n \widetilde{y}_{k-1}^2\Big)^{1/2} \frac{\frac{1}{\sqrt{n^2 b_n^3}}\sum_{k=1}^n \widetilde{y}_{k-1}\, u_k - \frac{1}{\sqrt{n^2 b_n^3}}\sum_{k=1}^n z_k\, \widetilde{y}_{k-1}\big(\sum_{k=1}^n z_k' z_k\big)^{-1}\sum_{k=1}^n z_k'\, u_k}{\frac{1}{n^2 b_n^3}\sum_{k=1}^n \widetilde{y}_{k-1}^2 - \frac{1}{n^2 b_n^3}\sum_{k=1}^n z_k\, \widetilde{y}_{k-1}\big(\sum_{k=1}^n z_k' z_k\big)^{-1}\sum_{k=1}^n z_k'\, y_{k-1}} \\ =\ & \frac{\frac{1}{\sqrt{n^2 b_n^3}}\sum_{k=1}^n \widetilde{y}_{k-1}\, u_k + O_P(b_n^{1/2})}{\big(\frac{1}{n^2 b_n^3}\sum_{k=1}^n \widetilde{y}_{k-1}^2\big)^{1/2} + O_P(n^{-1/2} b_n^{-1})} \\ \to_D\ & \mathcal{N}(0, \sigma^2), \end{aligned}$$

due to $b_n \to 0$ and $b_n \log^K n \to \infty$ for some $K > 0$. This proves (3.6). The proof of (3.7) is simple and hence the details are omitted. □

7.6 Proof of Theorem 4.1

We start with some preliminaries. First note that, as in (7.28), it follows from **?** that

$$\xi_j = \Phi\, \eta_j + \hat{\xi}_{j-1} - \hat{\xi}_j, \tag{7.40}$$

where $\hat{\xi}_j = \sum_{i=0}^{\infty} \hat{\phi}_i \eta_{j-i}$ with $\hat{\phi}_i = \sum_{k=i+1}^{\infty} \phi_k$, satisfying that $E\hat{\xi}_j^2 = \sigma_2^2 \sum_{i=0}^{\infty} \hat{\phi}_i^2 < \infty$. As a consequence, we have

$$\begin{aligned} x_k &= \sum_{j=1}^{k} \beta^{k-j} \xi_j = \Phi \sum_{j=1}^{k} \beta^{k-j} \eta_j + \sum_{j=1}^{k} \beta^{k-j} \big[\hat{\xi}_{j-1} - \hat{\xi}_j\big] \\ &= \Phi \sum_{j=1}^{k} \beta^{k-j} \eta_j + \Delta_k, \end{aligned}$$

where $\Delta_k = \beta^{k-1}\hat{\xi}_0 - \hat{\xi}_k + (1-\beta)\sum_{j=1}^{k-1} \beta^{k-j-1}\hat{\xi}_j$. Since, as in the proof of (7.29),

$$\max_{1\le k\le n} |\Delta_k| \le C_\tau \big(|\hat{\xi}_0| + \max_{1\le k\le n} |\hat{\xi}_k|\big) = O_P(n^{1/(2+\delta)}),$$

the same arguments used in the proof of (7.21) yield that, on a richer probability space, there exists a Brownian motion $B = \{B_t\}_{t\ge 0}$ and a random array $\{x_{nk}^*, n \ge 1, k = 1, 2, \cdots, n\}$ such that $\{x_{nk}^*, k = 1, 2, \cdots, n\} =_d \{x_k, k = 1, 2, \cdots, n\}$ for all $n \ge 1$ and

$$\sup_{0\le t\le 1} \Big| \frac{\Phi^{-1}}{\sqrt{n}\sigma_2} x_{n,\lfloor nt\rfloor}^* - J_\tau(t) \Big| = O_P(n^{-\eta}) \tag{7.41}$$

for some $\eta > 0$, where $J_\tau(t) = B_t + \tau \int_0^t e^{\tau(t-s)} B_s ds$ is a Gaussian process satisfying

$$\sup_{\substack{|s-t|\le\Delta \\ 0\le t\le 1}} |J_\tau(t) - J_\tau(s)| \le C_\tau \sup_{\substack{|s-t|\le\Delta \\ 0\le t\le 1}} |B_t - B_s| \;=\; O_P(\Delta^{1/2}).$$

We are now ready to prove Theorem 4.1. Note that

$$\widehat{\alpha}_{3n} - \alpha \;=\; \frac{\sum_{k=1}^{n} u_k f(x_{k-1})\, I(|x_{k-1}| \le b_n \sqrt{n})}{\sum_{k=0}^{n-1} f^2(x_k) I(|x_{k-1}| \le b_n \sqrt{n})}. \tag{7.42}$$

Recalling condition **A2**, we have

$$\frac{1}{\sqrt{n b_n \pi^2(b_n\sqrt{n})}} \sum_{k=1}^{n} u_k f(x_{k-1})\, I(|x_{k-1}| \le b_n \sqrt{n}) \;=\; \frac{1}{\sqrt{n b_n}} \sum_{k=1}^{n} u_k \widetilde{H}(\sigma_2 \Phi\, x_{nk}/b_n) + R_n,$$

where $x_{nk} = \Phi^{-1} x_{k-1}/\sqrt{n}\sigma_2$, $\widetilde{H}(x) = H(x) I(|x| \le 1)$ and

$$R_n \;=\; \frac{1}{\sqrt{n b_n \pi^2(b_n\sqrt{n})}} \sum_{k=1}^{n} u_k R(b_n\sqrt{n}, \sigma_2\Phi\, x_{nk}/b_n)\, I(\sigma_2\Phi\, |x_{nk}|/b_n \le 1).$$

Under given conditions **A1** and **A2**, R_n is a martingale with the conditional variance:

$$\begin{aligned} &\frac{1}{n b_n \pi^2(b_n\sqrt{n})} \sum_{k=1}^{n} \sigma_{1k}^2\, R^2(b_n\sqrt{n}, \sigma_2\Phi\, x_{nk}/b_n)\, I(\sigma_2\Phi\, |x_{nk}|/b_n \le 1) \\ \le\; &\frac{a^2(b_n\sqrt{n})}{\pi^2(b_n\sqrt{n})}\, \frac{1}{n b_n} \sum_{k=1}^{n} \sigma_{1k}^2\, A(\sigma_2\Phi\, x_{nk}/b_n) \;= o_P(1), \end{aligned}$$

where $A(x) = \big(1+|x|^{\delta}\big) I(|x|\le 1)$ and we have used the fact:

$$\frac{1}{nb_n}\sum_{k=1}^{n}\sigma_{1k}^2\, A(\sigma_2\Phi\, x_{nk}/b_n) \;= O_P(1),$$

as in the proof of (7.23). As a consequence, the classical martingale limit theorem (see, e.g., **?**) implies that $R_n = o_P(1)$, i.e.,

$$\frac{1}{\sqrt{nb_n\pi^2(b_n\sqrt{n})}}\sum_{k=1}^{n} u_k f(x_{k-1})\, I(|x_{k-1}|\le b_n\sqrt{n}) \;\;=\;\; \frac{1}{\sqrt{nb_n}}\sum_{k=1}^{n} u_k \widetilde{H}(\sigma_2\Phi\; x_{nk}/b_n) + o_P(1).$$

Similarly, we have

$$\frac{1}{nb_n\pi^2(b_n\sqrt{n})}\sum_{k=0}^{n-1} f^2(x_k) I(|x_{k-1}|\le b_n\sqrt{n}) \;\;=\;\; \frac{1}{nb_n}\sum_{k=0}^{n-1}\widetilde{H}^2(\sigma_2\Phi\, x_{nk}/b_n) + o_P(1).$$

Since the same arguments used in the proof of (7.13) implies

$$\begin{aligned}
&\Big(\frac{1}{\sqrt{nb_n}}\sum_{k=1}^{n} u_k\widetilde{H}(\sigma_2\Phi\; x_{nk}/b_n), \frac{1}{nb_n}\sum_{k=0}^{n-1}\widetilde{H}^2(\sigma_2\Phi\, x_{nk}/b_n)\Big)\\
\to_D\;\; &\Big[\sigma_1\Big(\int_{-\infty}^{\infty}\widetilde{H}^2(\sigma_2\Phi\; x)dx\Big)^{1/2} L_{J_\tau}^{1/2}(1,0)\, N,\; \int_{-\infty}^{\infty}\widetilde{H}^2(\sigma_2\Phi\; x)dx\, L_{J_\tau}(1,0)\Big], \qquad (7.43)
\end{aligned}$$

where N is a standard normal variate independent of $L_{J_\tau}(1,0)$, taking all estimates above into (7.42), we obtain

$$\begin{aligned}
\sqrt{n\, b_n}\pi(b_n\sqrt{n}\,)\,(\widehat{\alpha}_{3n}-\alpha) \;\;&=\;\; \frac{\frac{1}{\sqrt{nb_n}}\sum_{k=1}^{n} u_k\widetilde{H}(\sigma_2\Phi\; x_{nk}/b_n) + o_P(1)}{\frac{1}{nb_n}\sum_{k=0}^{n-1}\widetilde{H}^2(\sigma_2\Phi\, x_{nk}/b_n) + o_P(1)}\\
&\to_D\;\; \sigma_1\,(\sigma_2\Phi)^{\,1/2}\Big(\int_{|x|\le 1} H^2(x)dx\Big)^{-1/2} L_{J_\tau}^{-1/2}(1,0)\, N,
\end{aligned}$$

where we have used the fact that $\int_{-\infty}^{\infty}\widetilde{H}^2(\sigma_2\Phi\; x)dx \;=\; (\sigma_2\Phi)^{\,-1}\int_{|x|\le 1} H^2(x)dx$. This proves (4.2). Similarly, we have

$$\begin{aligned}
&\Big[\sum_{k=1}^{n} f^2(x_{k-1}) I(|x_{k-1}|\le b_n\sqrt{n})\Big]^{1/2}(\widehat{\alpha}_{3n}-\alpha)\\
=\;\; &\frac{\frac{1}{\sqrt{nb_n}}\sum_{k=1}^{n} u_k\widetilde{H}(\sigma_2\Phi\; x_{nk}/b_n) + o_P(1)}{\Big(\frac{1}{nb_n}\sum_{k=0}^{n-1}\widetilde{H}^2(\sigma_2\Phi\, x_{nk}/b_n) + o_P(1)\Big)^{1/2}} \;\to_D\; \mathcal{N}(0,\sigma_1^2),
\end{aligned}$$

i.e., (4.3) holds. Finally, the proof of (4.4) is simple and hence the details are omitted. □

7.7 Proof of Theorem 4.2

Let $\widetilde{v}_k = v_k - \hat{v}_k$ and note that

$$\begin{aligned}
&\sum_{k=1}^{n} \big(\hat{v}_k - \hat{v}_{k+\min\{M,n-k\}}\big)\, x_{k-1} I(|x_{k-1}| \le b_n \sqrt{n}) - \widetilde{\Lambda}_{\xi v} \\
&\qquad = \sum_{j=1}^{M} \sum_{k=1}^{n-j} (\hat{v}_{k+j-1} - \hat{v}_{k+j})\, x_{k-1} I(|x_{k-1}| \le b_n \sqrt{n}) - \widetilde{\Lambda}_{\xi v} \\
&\qquad = -\, \hat{v}_n \sum_{j=1}^{M} x_{n-j} I(|x_{n-j}| \le b_n \sqrt{n}). \qquad (7.44)
\end{aligned}$$

It suffices to show that, for $L_M = o\big[(nb_n)^{1/2}\big]$,

$$C_n \;:=\; \sum_{k=1}^{n} \big(\widetilde{v}_k - \widetilde{v}_{k+\min\{M,n-k\}}\big)\, x_{k-1} I(|x_{k-1}| \le b_n \sqrt{n}) \;=\; o_P(n b_n^{3/2}), \qquad (7.45)$$

$$D_n \;:=\; \sum_{j=1}^{M} x_{n-j} I(|x_{n-j}| \le b_n \sqrt{n}) \;=\; o_P(n b_n^{3/2}) \qquad (7.46)$$

and

$$\begin{aligned}
&\Big(\frac{1}{n b_n^{3/2}} \sum_{k=1}^{n} v_{k+\min\{M,n-k\}}\, x_{k-1} I(|x_{k-1}| \le b_n \sqrt{n}), \quad \frac{1}{n^2 b_n^3} \sum_{k=1}^{n} x_{k-1}^2 I(|x_{k-1}| \le b_n \sqrt{n})\Big) \\
&\to_D \Big((\sigma\Phi)^{-1/2} \Psi_1 \Big[\int_{|x|\le 1} x^2 dx\, L_{J_\tau}(1,0)\Big]^{1/2} N, \;\; (\sigma\,\Phi)^{-1} \int_{|x|\le 1} x^2 dx\, L_{J_\tau}(1,0)\Big). \qquad (7.47)
\end{aligned}$$

Indeed, by noting that the estimation error in $\hat{\alpha}_{bSW}$ is

$$\begin{aligned}
\hat{\alpha}_{bSW} - \alpha \;&=\; \frac{\sum_{k=1}^{n} v_{k+\min\{M,n-k\}}\, x_{k-1} I(|x_{k-1}| \le b_n \sqrt{n})}{\sum_{k=1}^{n} x_{k-1}^2 I(|x_{k-1}| \le b_n \sqrt{n})} \\
&\quad + \frac{\sum_{k=1}^{n} \big(v_k - v_{k+\min\{M,n-k\}}\big)\, x_{k-1} I(|x_{k-1}| \le b_n \sqrt{n}) - \widetilde{\Lambda}_{\xi v}}{\sum_{k=1}^{n} x_{k-1}^2 I(|x_{k-1}| \le b_n \sqrt{n})},
\end{aligned}$$

it follows from (7.44)-(7.47), $\hat{v}_n = O_P(1)$, $\int_{|x|\le 1} x^2 dx = 2/3$ and the continuous mapping theorem that

$$\begin{aligned}
n b_n^{3/2} \big(\hat{\alpha}_{bSW} - \alpha\big) \;&=\; \frac{\frac{1}{n b_n^{3/2}} \sum_{k=1}^{n} v_{k+\min\{M,n-k\}}\, x_{k-1} I(|x_{k-1}| \le b_n \sqrt{n}) + \frac{1}{n b_n^{3/2}} (C_n + \hat{v}_n D_n)}{\frac{1}{n^2 b_n^3} \sum_{k=1}^{n} x_{k-1}^2 I(|x_{k-1}| \le b_n \sqrt{n})} \\
&\to_D\; (3\sigma\,\Phi/2)^{1/2}\, \Psi_1\, L_{J_\tau}^{-1/2}(1,0)\, N.
\end{aligned}$$

This proves (4.9). As $\widehat{\Psi}_{1n}$ is a consistent estimator of $\Psi_1^2 = E v_1^2$, the proof of (4.11) is similar and hence the details are omitted.

We next prove (7.45)-(7.47), starting with (7.45). Note that $\tilde{v}_k = (\hat{\alpha} - \alpha) x_{k-1}$ and $\max_{1\le k\le n} |x_k| = O_P(\sqrt{n})$ due to $x_{[nt]}/\sqrt{n}\sigma \Rightarrow J_\tau(t)$ on $D[0,1]$. It follows from (4.6) that $\max_{1\le k\le n} |\tilde{v}_k| =$

$O_P(n^{-1/2})$ and

$$\begin{aligned} |C_n| &\le 2 \max_{1\le k\le n} |\tilde v_k| \sum_{k=1}^n |x_{k-1}| I(|x_{k-1}|\le b_n\sqrt{n}) \\ &= O_P(nb_n^2)\, \frac{1}{nb_n} \sum_{k=1}^n w(x_{k-1}/\sqrt{n} b_n) = O_P(nb_n^2), \end{aligned}$$

where $w(x) = xI(|x|\le 1)$. This proves (7.45) as $b_n \to 0$.

The proof of (7.46) is simple since $|D_n| \le M\sqrt{n} b_n = o(nb_n^{3/2})$.

To prove (7.47), for $M = M_n$, write

$$\begin{aligned} v_{k+M} &= \sum_{j=0}^\infty \psi_j \zeta_{k+M-j} = \sum_{j=0}^M \psi_{M-j}\zeta_{k+j} + \sum_{j=1}^\infty \psi_{j+M}\zeta_{k-j} \\ &= v_{k,M}^{(1)} + v_{k,M}^{(2)}, \qquad \text{say.} \end{aligned}$$

Consequently, we have

$$\begin{aligned} &\sum_{k=1}^n v_{k+\min\{M,n-k\}}\, x_{k-1} I(|x_{k-1}|\le b_n\sqrt{n}) \\ &= \sum_{k=1}^{n-M} v_{k+M}\, x_{k-1} I(|x_{k-1}|\le b_n\sqrt{n}) + v_n \sum_{k=n-M+1}^n x_{k-1} I(|x_{k-1}|\le b_n\sqrt{n}) \\ &= \sum_{k=1}^n v_{k,M}^{(1)}\, x_{k-1} I(|x_{k-1}|\le b_n\sqrt{n}) + R_n, \end{aligned} \tag{7.48}$$

where

$$|R_n| \le \sqrt{n} b_n \Big[\sum_{k=n-M+1}^n \big(|v_n| + |v_{k,M}^{(1)}|\big) + \sum_{k=1}^{n-M} |v_{k,M}^{(2)}| \Big].$$

It follows from $E\zeta_1^2 < \infty$ and $E|v_{k,M}^{(2)}| \le \big(E|v_{k,M}^{(2)}|^2\big)^{1/2} \le \big(E\zeta_1^2 \sum_{j=M+1}^\infty \psi_j^2\big)^{1/2}$ that

$$E|R_n| \le C\sqrt{n} b_n \Big[M + n\big(\sum_{j=M}^\infty \psi_j^2\big)^{1/2}\Big] = o(nb_n^{3/2}),$$

i.e., $R_n = o_P(nb_n^{3/2})$. Taking this fact into (7.48), result (7.47) will follow if we prove

$$\begin{aligned} &\Big(\frac{1}{nb_n^{3/2}} \sum_{k=1}^n v_{k,M}^{(1)}\, x_{k-1} I(|x_{k-1}|\le b_n\sqrt{n}),\quad \frac{1}{n^2 b_n^3}\sum_{k=1}^n x_{k-1}^2 I(|x_{k-1}|\le b_n\sqrt{n})\Big) \\ &\to_D \Big(\Psi_1\, \Phi^{-1/2}\Big[\int_{|x|\le 1} x^2 dx\, L_{J_\tau}(1,0)\Big]^{1/2} N,\ \ (\sigma\,\Phi)^{-1}\int_{|x|\le 1} x^2 dx\, L_{J_\tau}(1,0)\Big). \end{aligned} \tag{7.49}$$

Let $z_{nk} = v_{k,M}^{(1)}\, x_{k-1} I(|x_{k-1}|\le b_n\sqrt{n})$ and $\mathcal{F}_k = \sigma(\zeta_k, \zeta_{k-1}, ..)$. It is readily seen that $(z_{nk}, \mathcal{F}_k)_{k\ge 1}$ forms a martingale difference with

$$\begin{aligned} \sum_{k=1}^n E(z_{nk}^2|\mathcal{F}_{k-1}) &= \sum_{k=1}^n E\big(v_{k,M}^{(1)}\big)^2 x_{k-1}^2 I(|x_{k-1}|\le b_n\sqrt{n}) = \sigma^2 \sum_{j=0}^M \psi_j^2 \sum_{k=1}^n x_{k-1}^2 I(|x_{k-1}|\le b_n\sqrt{n}) \\ &= \big[1+o(1)\big]\, \Psi_1^2 \sum_{k=1}^n x_{k-1}^2 I(|x_{k-1}|\le b_n\sqrt{n}). \end{aligned}$$

The proof of (7.48) now is the same as that of (7.43) and hence the details are omitted. The proof of Theorem 4.2 is complete. □

8 Tables and Figures

Table 1: The empirical size for the near unit root autoregression model

n	τ \ C_0	1	2	3	4	5	10	20
50	0	0.059	0.058	0.072	0.085	0.091	0.093	0.070
	-1	0.062	0.060	0.070	0.080	0.086	0.079	0.065
	-5	0.058	0.056	0.069	0.070	0.074	0.070	0.069
	-10	0.053	0.057	0.068	0.076	0.079	0.077	0.077
100	0	0.061	0.060	0.071	0.078	0.084	0.091	0.066
	-1	0.060	0.067	0.065	0.079	0.081	0.076	0.064
	-5	0.057	0.051	0.060	0.067	0.073	0.064	0.064
	-10	0.048	0.051	0.067	0.071	0.074	0.071	0.070
200	0	0.057	0.060	0.066	0.074	0.086	0.087	0.069
	-1	0.061	0.058	0.061	0.069	0.079	0.078	0.059
	-5	0.054	0.053	0.059	0.064	0.067	0.060	0.059
	-10	0.046	0.050	0.059	0.066	0.067	0.064	0.064
500	0	0.059	0.063	0.064	0.072	0.081	0.090	0.078
	-1	0.054	0.059	0.061	0.069	0.080	0.080	0.062
	-5	0.049	0.050	0.062	0.064	0.065	0.059	0.057
	-10	0.052	0.053	0.056	0.065	0.065	0.060	0.059

[1] Nominal size 5%.
[2] The number of simulations is $10,000$.

Table 2: Empirical size for the unit root tests when when $\theta = 0$

n	p	Self-Weighted (C_0) 1	2	3	4	5	10	ADF
$\epsilon_k \sim \mathcal{N}(0,1)$								
50	0	0.043	0.043	0.055	0.070	0.086	0.099	0.054
	1	0.061	0.062	0.065	0.075	0.087	0.101	0.056
	2	0.067	0.067	0.070	0.074	0.083	0.094	0.050
100	0	0.042	0.043	0.054	0.066	0.083	0.096	0.051
	1	0.063	0.057	0.061	0.069	0.079	0.093	0.050
	2	0.072	0.065	0.068	0.071	0.080	0.090	0.048
200	0	0.041	0.044	0.049	0.060	0.078	0.096	0.051
	1	0.064	0.056	0.054	0.061	0.076	0.094	0.050
	2	0.072	0.064	0.060	0.063	0.073	0.090	0.046
500	0	0.040	0.045	0.053	0.059	0.070	0.101	0.056
	1	0.058	0.055	0.054	0.059	0.070	0.100	0.053
	2	0.065	0.060	0.057	0.059	0.071	0.098	0.053
$\epsilon_k \sim U(-\sqrt{3}, \sqrt{3})$								
50	0	0.035	0.042	0.052	0.071	0.087	0.100	0.054
	1	0.061	0.062	0.063	0.073	0.086	0.099	0.052
	2	0.059	0.065	0.069	0.078	0.082	0.093	0.049
100	0	0.035	0.041	0.048	0.065	0.086	0.097	0.051
	1	0.064	0.057	0.055	0.066	0.081	0.096	0.050
	2	0.070	0.063	0.062	0.069	0.079	0.093	0.047
200	0	0.037	0.040	0.053	0.062	0.081	0.099	0.054
	1	0.064	0.052	0.057	0.064	0.079	0.098	0.054
	2	0.069	0.060	0.065	0.065	0.079	0.097	0.050
500	0	0.041	0.041	0.046	0.054	0.067	0.094	0.053
	1	0.062	0.047	0.050	0.055	0.067	0.095	0.053
	2	0.069	0.056	0.053	0.055	0.066	0.094	0.052

[1] Nominal size 5%.
[2] The parameter p denotes the number of lag terms in augmented autoregression for the UR test.
[3] The number of simulations is $10,000$.

Table 3: Empirical size for the unit root tests when when $\theta = -0.5$

n	p	Self-Weighted (C_0) 1	2	3	4	5	10	ADF
$\epsilon_k \sim \mathcal{N}(0,1)$								
50	0	0.082	0.139	0.193	0.251	0.286	0.306	0.220
	1	0.057	0.057	0.063	0.070	0.072	0.076	0.056
	2	0.062	0.061	0.061	0.066	0.067	0.070	0.051
100	0	0.098	0.154	0.208	0.264	0.316	0.343	0.252
	1	0.052	0.049	0.051	0.061	0.064	0.064	0.049
	2	0.059	0.054	0.052	0.057	0.060	0.060	0.049
200	0	0.112	0.163	0.215	0.261	0.316	0.361	0.269
	1	0.048	0.039	0.046	0.058	0.061	0.062	0.050
	2	0.054	0.043	0.045	0.054	0.057	0.058	0.047
500	0	0.138	0.180	0.220	0.258	0.307	0.389	0.298
	1	0.042	0.036	0.042	0.056	0.061	0.062	0.054
	2	0.048	0.039	0.042	0.054	0.059	0.060	0.053
$\epsilon_k \sim U(-\sqrt{3}, \sqrt{3})$								
50	0	0.085	0.138	0.193	0.251	0.290	0.302	0.218
	1	0.060	0.055	0.058	0.067	0.070	0.072	0.050
	2	0.064	0.058	0.060	0.066	0.069	0.070	0.049
100	0	0.102	0.154	0.206	0.272	0.322	0.344	0.253
	1	0.053	0.043	0.050	0.062	0.064	0.065	0.050
	2	0.061	0.049	0.049	0.059	0.059	0.061	0.047
200	0	0.120	0.176	0.221	0.269	0.330	0.366	0.280
	1	0.048	0.038	0.050	0.060	0.064	0.065	0.052
	2	0.055	0.043	0.048	0.058	0.063	0.063	0.052
500	0	0.141	0.185	0.216	0.256	0.301	0.379	0.292
	1	0.039	0.035	0.042	0.056	0.061	0.061	0.055
	2	0.044	0.036	0.042	0.054	0.058	0.059	0.052

[1] Nominal size 5%.
[2] The parameter p denotes the number of lag terms in augmented autoregression for the UR test.
[3] The number of simulations is $10,000$.

Table 4: The empirical size for testing predictive ability when $\phi_v = 0$ and $\phi_\xi = 0$

		Self-Weighted (C_0)					IVX (γ)			
n	τ	1	2	3	4	5	0.6	0.7	0.8	0.9
50	0	0.060	0.082	0.102	0.107	0.102	0.071	0.070	0.072	0.070
	-1	0.059	0.081	0.089	0.090	0.090	0.076	0.072	0.071	0.070
	-5	0.051	0.073	0.080	0.077	0.074	0.080	0.080	0.080	0.079
	-10	0.048	0.078	0.085	0.082	0.082	0.086	0.086	0.083	0.085
100	0	0.059	0.074	0.092	0.095	0.092	0.062	0.059	0.060	0.060
	-1	0.052	0.071	0.082	0.086	0.080	0.067	0.064	0.060	0.061
	-5	0.051	0.060	0.070	0.067	0.063	0.074	0.072	0.069	0.068
	-10	0.047	0.068	0.075	0.071	0.071	0.082	0.079	0.076	0.076
200	0	0.058	0.074	0.081	0.085	0.087	0.060	0.057	0.056	0.055
	-1	0.056	0.064	0.071	0.075	0.077	0.063	0.060	0.058	0.056
	-5	0.050	0.060	0.066	0.067	0.061	0.066	0.065	0.065	0.063
	-10	0.049	0.062	0.065	0.063	0.061	0.070	0.068	0.066	0.065
500	0	0.061	0.066	0.074	0.090	0.092	0.057	0.060	0.058	0.060
	-1	0.062	0.061	0.071	0.079	0.080	0.059	0.060	0.060	0.058
	-5	0.053	0.059	0.069	0.068	0.066	0.061	0.060	0.059	0.061
	-10	0.050	0.060	0.067	0.063	0.059	0.062	0.060	0.061	0.061

[1] Nominal size 5%.
[2] The number of simulations is $10,000$.

Table 5: The empirical size for testing predictive ability when $\phi_v = 0$ and $\phi_\xi = 0.5$

		Self-Weighted (C_0)					IVX (γ)			
n	τ	1	2	3	4	5	0.6	0.7	0.8	0.9
50	0	0.050	0.063	0.080	0.091	0.104	0.066	0.066	0.068	0.070
	-1	0.055	0.063	0.075	0.090	0.095	0.069	0.069	0.068	0.066
	-5	0.051	0.060	0.069	0.075	0.077	0.073	0.073	0.072	0.071
	-10	0.051	0.060	0.074	0.079	0.077	0.072	0.075	0.074	0.075
100	0	0.048	0.066	0.069	0.079	0.087	0.056	0.056	0.057	0.059
	-1	0.057	0.061	0.062	0.071	0.080	0.059	0.058	0.057	0.058
	-5	0.052	0.057	0.058	0.064	0.066	0.064	0.062	0.062	0.060
	-10	0.049	0.054	0.056	0.064	0.065	0.067	0.066	0.066	0.064
200	0	0.052	0.061	0.068	0.075	0.074	0.052	0.052	0.052	0.054
	-1	0.055	0.058	0.066	0.068	0.071	0.055	0.053	0.053	0.053
	-5	0.050	0.054	0.055	0.059	0.063	0.058	0.057	0.057	0.054
	-10	0.052	0.054	0.056	0.060	0.065	0.062	0.059	0.058	0.058
500	0	0.055	0.061	0.063	0.066	0.064	0.057	0.057	0.057	0.060
	-1	0.054	0.060	0.063	0.062	0.066	0.057	0.056	0.058	0.058
	-5	0.051	0.054	0.058	0.060	0.065	0.053	0.056	0.056	0.057
	-10	0.047	0.050	0.058	0.062	0.063	0.056	0.053	0.057	0.056

[1] Nominal size 5%.
[2] The number of simulations is $10,000$.

Table 6: The empirical size for testing predictive ability when $\phi_v = 0.5$ and $\phi_\xi = 0.5$

		SW(C_0)			BC-SW (C_0)			IVX (γ)		BC-IVX (γ)	
n	τ	1	3	5	1	3	5	0.7	0.9	0.7	0.9
50	0	0.008	0.022	0.042	0.049	0.033	0.054	0.096	0.082	0.057	0.058
	-1	0.007	0.019	0.042	0.035	0.032	0.056	0.119	0.100	0.062	0.063
	-5	0.011	0.065	0.162	0.011	0.055	0.103	0.297	0.253	0.136	0.124
	-10	0.026	0.309	0.410	0.013	0.169	0.211	0.530	0.477	0.275	0.241
100	0	0.003	0.014	0.027	0.038	0.023	0.047	0.089	0.069	0.044	0.047
	-1	0.003	0.010	0.019	0.026	0.022	0.042	0.123	0.087	0.047	0.048
	-5	0.004	0.020	0.091	0.008	0.030	0.064	0.311	0.238	0.094	0.086
	-10	0.007	0.160	0.364	0.005	0.080	0.140	0.543	0.465	0.194	0.163
200	0	0.002	0.010	0.020	0.027	0.021	0.045	0.103	0.063	0.039	0.041
	-1	0.002	0.006	0.010	0.017	0.019	0.041	0.145	0.083	0.042	0.045
	-5	0.001	0.005	0.045	0.003	0.024	0.049	0.324	0.223	0.074	0.068
	-10	0.002	0.066	0.281	0.003	0.042	0.089	0.553	0.432	0.131	0.110
500	0	0.002	0.009	0.013	0.012	0.023	0.041	0.127	0.051	0.033	0.039
	-1	0.001	0.004	0.006	0.008	0.021	0.038	0.172	0.073	0.038	0.038
	-5	0.000	0.001	0.009	0.002	0.021	0.037	0.354	0.195	0.053	0.052
	-10	0.001	0.015	0.160	0.002	0.029	0.056	0.562	0.396	0.080	0.070

[1] Nominal size 5%.
[2] The number of simulations is $10,000$.

Table 7: Tests for univariate predictive regressions with monthly data

Regressors	OLS	SW	IVX
Book-to-market ratio (b/m)	0.003	-0.007	0.004
Dividend payout ratio (d/e)	0.005	0.003	0.006
Default yield spread (dfy)	0.246	0.267	0.250
Dividend-price ratio (d/p)	0.005	-0.001	-0.005
Dividend yield (d/y)	0.004	0.004	-0.005
Earnings-price ratio (e/p)	0.003	0.013	0.004
Inflation rate (infl)	-0.927^{**}	-1.146^{**}	-0.928^{**}
Long-term yield (lty)	-0.088^{*}	-0.048	-0.095
Net equity expansion (nits)	-0.045	-0.128	-0.023
T-bill rate (tbl)	-0.114^{**}	-0.108	-0.112^{**}

[a] * denotes 10% significance and ** denotes 5% significance.

Table 8: Tests for univariate predictive regressions with quarterly data

Regressors	OLS	SW	IVX
Book-to-market ratio (b/m)	0.012	-0.056	0.012
Dividend payout ratio (d/e)	0.020	0.011	0.024
Default yield spread (dfy)	0.692	-12.290	0.656
Dividend-price ratio (d/p)	0.018	-0.053	-0.027
Dividend yield (d/y)	0.017	-0.083	-0.049
Earnings-price ratio (e/p)	0.007	0.003	0.013
Inflation rate (infl)	-0.998^{*}	-1.145	-1.010^{*}
Long-term yield (lty)	-0.236	-1.105	-0.254
Net equity expansion (nits)	-0.122	-0.330	-0.060
T-bill rate (tbl)	-0.298^{*}	-0.352	-0.290^{*}
Term spread (tms)	0.503	0.849	0.509

[a] * denotes 10% significance and ** denotes 5% significance.

(a)(b)(c)(d)
$\theta = \theta = \theta = \theta =$
0; 0; -0.5; 0.5;
ϵ_k ϵ_k ϵ_k $\epsilon_k \sim$
$\mathcal{N}(0,1)$ $U(-1,1)$ $\mathcal{N}(0,1)$ $U(-1,1)$

Figure 1: The local power for unit root test

[1] The sample size $n = 200$. The number of simulations is $10,000$.
[2] For the SW estimator, we apply the truncation rate $b_n = C_0 \log^{-1}(n)$ for $C_0 \in \{3, 5\}$.
[3] For $y_k \sim$AR(1) with $\theta = 0$ in (a) and (b), the lag parameter $p = 0$.
[4] For $y_k \sim$AR(2) with $\theta = 0.5$ in (c) and (d), the lag parameter $p = 1$.

(a)(b)(c)(d)
$\tau = \tau = \tau = \tau =$
0 -1 -5 -10

Figure 2: The local power for testing predictive ability when $\phi_v = 0$ and $\phi_\xi = 0$

[a] The sample size $n = 200$. The number of simulations is $10,000$.
[b] For the SW estimator, we apply the truncation rate $b_n = C_0 \log^{-1}(n)$ for $C_0 \in \{3, 5\}$.
[c] For the IVX estimator, we apply the instrument with $\gamma \in \{0.7, 0.9\}$.
[d] The autoregressive coefficients of error terms $\phi_v = 0$ and $\phi_\xi = 0$.

(a)(b)(c)(d)
$\tau = \tau = \tau = \tau =$
0 -1 -5 -10

Figure 3: The local power for testing predictive ability when $\phi_v = 0$ and $\phi_\xi = 0.5$

[a] The sample size $n = 200$. The number of simulations is $10,000$.
[b] For the SW estimator, we apply the truncation rate $b_n = C_0 \log^{-1}(n)$ for $C_0 \in \{3, 5\}$.
[c] For the IVX estimator, we apply the instrument with $\gamma \in \{0.7, 0.9\}$.
[d] The autoregressive coefficients of error terms $\phi_v = 0$ and $\phi_\xi = 0.5$.

(a)(b)(c)(d)
$\tau = \tau = \tau = \tau =$
0 -1 -5 -10

Figure 4: The local power for testing predictive ability when $\phi_v = 0.5$ and $\phi_\xi = 0.5$

[a] The sample size $n = 200$. The number of simulations is $10,000$.
[b] For the BC-SW estimator, we apply the truncation rate $b_n = C_0 \log^{-1}(n)$ for $C_0 \in \{3, 5\}$.
[c] For the IVX estimator, we apply the instrument with $\gamma \in \{0.7, 0.9\}$.
[d] The autoregressive coefficients of error terms $\phi_v = 0.5$ and $\phi_\xi = 0.5$.